Masatoshi Nei Lecture

In search of microbe number one

William Martin[1]

[1]University of Dusseldorf (United States)

Life is a chemical reaction. More specifically, life is an exergonic chemical reaction. What was the chemical reaction from which the first cells arose, and what was the chemical reaction that fuelled the first free-living cells? These are questions about chemistry and physiology, but molecular evolution can contribute. The last universal common ancestor (LUCA) is the assemblage of cells from which all life evolved roughly four billion years ago. Genomes and phylogeny have yielded new avenues to understanding early evolution and LUCA. We know LUCA had the universal genetic code shared by all descendant life forms. But how did LUCA harness energy? The chemical reactions that help cells harness energy from their environments today seem almost as diverse as life itself. Which forms of energy harnessing are ancient? We looked at that question using data from sequenced microbial genomes. We found that LUCA lived from gases ? H_2, CO_2, H_2S, CO, N_2 ? in a setting that looked very much like a modern submarine hydrothermal vent. The classical approach to investigate LUCA using genomes is to identify genes that are present in all modern cells hence present present in LUCA. We asked which genes trace to LUCA by phylogenetic criteria. The results indicate that the first forms of life were anaerobic chemoautotrophs that evolved from preexisting geochemical processes involving exergonic reactions of H_2, metals, and CO_2.

Plenary Session 1 (July 10, 9:00–10:00)

Plenary-1

Evolution of sequence-specific anti-silencing systems in Arabidopsis

Tetsuji Kakutani[1, 2]

[1]The University of Tokyo (Japan), [2]National Institute of Genetics (Japan)

The arms race between parasitic sequences and hosts is a major driving force for evolution of gene control systems. Since transposable elements (TEs) are potentially deleterious, eukaryotes silence them by epigenetic mechanisms such as DNA methylation. Little is known about how TEs counteract the silencing to propagate during evolution. Here we report behavior of sequence-specific anti-silencing proteins used by Arabidopsis TEs and coevolution of those proteins and their target sequences. Through this coevolution, these TEs propagate with minimum host damage. Our findings provide insight into the evolutionary dynamics of these apparently "selfish" sequences. They also provide potential tools to edit epigenomes in a sequence-specific manner.

Plenary-2

Tracking a killer: using ancient DNA to understand the evolutionary history of tuberculosis

Anne C. Stone[1]

[1]Arizona State University (United States)

Ancient DNA has become a powerful tool to investigate human population history, plant and animal domestication, as well as the pathogens that have impacted us through time. In this talk, I will discuss how ancient DNA allows us to examine the history of *Mycobacterium tuberculosis* and related strains in the *Mycobacterium tuberculosis*complex (MTBC) which cause tuberculosis (TB). Long a scourge of humans, as well as other animals, TB has now surpassed HIV as the leading cause of death from infectious disease. Here, I will focus on patterns of pathogen exchange before and after the • gAge of Exploration• h in the Americas and discuss ways that TB may have adapted to humans and other animals. Previous research from our group led to the recovery of MTBC genomes from three 1000-year old skeletal TB cases from coastal Peru and showed that these strains are closely related to strains adapted to sea mammals (specifically Southern Hemisphere pinnipeds). However, it is unclear whether these strains spread to inland South America and to North America by human-to-human transmission or whether different strains spread into North America via another route. Also, after contact, European strains were introduced, ultimately replacing pre-contact strains, but the timing and extent of this is unknown. Our present work focuses on skeletal TB cases from pre-contact and historic sites from the Americas and shows evidence for spread of pinniped-derived MTBC strains to people in non-coastal areas and the presence of European *M. tuberculosis* strains post-contact.

This work was supported by funding from the National Science Foundation (BCS-1063939 and BCS- 1515163), the Wenner Gren Foundation, and the Max Planck Institute.

O-02-WF01

Re-thinking a classic clinal trait: Pleiotropic consequences of thermally adaptive dopamine on pigmentation clines in Drosophila

Ana Marija Jaksic[1, 2], Viola Nolte[1], Neda Barghi[1], Francois Mallard[1], Kathrin Anna Otte[1], Lidija Svecnjak[3], Kirsten-Andre Senti[1], Christian Schlotterer[1]

[1]Vetmeduni Vienna (Austria), [2]Vetmeduni Vienna (Austria), [3]University of Zagreb (Austria)

Clinal variation is widely regarded as strong evidence for adaptation in natural populations. One of the best-studied clinal phenotypes is the latitudinal melanism cline in ectotherms. Owing to the repeatability of clinal pigmentation across species and continents, this trait is frequently considered as a classic example for an adaptive trait. Nevertheless, how selection operates on this trait is still debated. The thermal budget hypothesis, one popular explanation, suggests that pigmentation intensity modulates the use of sunlight for thermoregulation. Here, we challenge this hypothesis and show how misleading the evolutionary interpretation of phenotypic variation can be. Using highly replicated experimental evolution setup, we exposed natural Drosophila simulans populations to novel thermal environments. After more than 100 generations of adaptation to hot temperature, the key components of melanin metabolism evolved in the same direction as in natural clines. Because temperature adaptation was not mediated by exposure to sun, we searched for alternative explanations. Melanin, the key ingredient for pigmentation, is synthesized from dopamine, a highly pleiotropic neurotransmitter. Using RNA-seq, metabolomics, and transgenic and pharmacologic modulation of dopamine, we uncovered the adaptive role of dopamine levels. Temperature strongly modulates signaling activity and in our experiment and natural populations, evolution modulates dopaminergic signaling to maintain synaptic homeostasis. We demonstrate that the evolutionary modulation of dopaminergic signaling helps to maintain normal locomotion even at high temperatures. We conclude that the pigmentation cline in Drosophila is a pleiotropic read-out of adaptive changes in dopamine metabolism, rather than an adaptive phenotype. Our results demonstrate that pleiotropy is a major challenge for the interpretation of phenotypic variation in a causative evolutionary framework.

O-02-WF02

Antagonistic pleiotropy is rare among new mutations

Mrudula Sunil Sane[1], Deepa Agashe[1]

[1]National Centre for Biological Sciences, Tata Institute of Fundamental Research (India)

Pleiotropic effects of mutations may lead to tradeoffs, which are thought to constrain adaptation and play an important role in diverse biological phenomena. Specifically, antagonistic pleiotropic effects are thought to be frequent, especially in large-effect mutations. Theory and experiments have typically focused on pleiotropic effects of beneficial mutations arising under strong selection. This is a problem because most mutations are deleterious or neutral, and are rarely observed under strong selection. Additionally, strong selection on a few traits allows deleterious mutations to accumulate in other traits, causing apparent tradeoffs and confounding estimates of true pleiotropy. Here, we characterize the incidence, nature and effect size of pleiotropy for carbon utilization among single evolved mutations in *Escherichia* coli populations from a mutation accumulation experiment. While synergistic pleiotropy was fairly common (SP, ~28% of mutations), very few mutations showed antagonistic pleiotropy (AP, ~10% of mutations). Thus, true tradeoffs, measured as AP, are rare. Comparing with null expectations generated from the independent distributions of effect sizes for each resource, we found that AP is significantly rarer than expected. Similarly, compared to null expectation, we find that large-effect mutations are more likely to show AP. Thus, AP may be more important during the early stages of adaptive evolution when large-effect beneficial mutations dominate. Our unbiased quantification of pleiotropy provides the first empirical support for long-standing evolutionary hypotheses. Contrary to current thought, our results suggest that at the mechanistic level, tradeoffs may only rarely constrain adaptation, largely when it proceeds through large-effect mutations.

O-02-WF03

The mechanistic basis of epistatic variety in a metabolic pathway uncovered by high-resolution fitness interaction mapping

Harry Kemble[1], Catherine Eisenhauer[1], Audrey Chapron[1], Melanie Magnan[1], Herve Le Nagard[1], Philippe Nghe[2], Olivier Tenaillon[1]

[1]INSERM (France), [2]ESPCI (France)

Fitness epistasis, the interdependence of mutations in their fitness impacts, can play a significant role in shaping evolutionary dynamics and determining evolutionarily accessible paths. It results from both direct interactions on the genotype-phenotype scale and non-linearity in the phenotype-fitness map. Empirical measurements of fitness epistasis can therefore provide information on these underlying mappings, knowledge of which is invaluable for predictive models.

Here, we focussed on the epistasis deriving solely from the shape of a phenotype-fitness map, using a metabolic pathway as our model. To this end, we engineered a system enabling precise, high-throughput measurement of the fitness epistasis between mutations affecting the expression of two metabolic genes in *E. coli*. Further, by utilising chemically controllable promoters, our system permitted characterisation of the epistasis between the same sets of mutations from different phenotypic starting points.

We identify a remarkable variety of types and trends of epistasis in this simple system, along with a dependence of this epistasis on phenotypic starting point, and demonstrate that epistasis cannot be reliably predicted from the fitness effects of the single mutations alone. We find that these observations can be explained by an existing two-layer phenotype-fitness landscape model incorporating metabolic control theory (describing pathway flux as a function of enzyme activities) and the assumption of a phenotypic optimum (describing fitness as a function of flux), forming a curved ridge in the activity-fitness map. Such a landscape suggests important evolutionary properties, such as robustness, and may be generalisable to any linear pathway with an optimum output.

O-02-WF04

Long-Term Evolution of Tetrahymena thermopila

Jason Tarkington[1], Rebecca Zufall[1]

[1]University of Houston (United States)

One of the most important questions for evolutionary biologists is how variation builds up over time to create all of the diversity observed around us. Small incremental changes in isolated populations can, given enough time, lead to major differences in the organisms that make up those populations. However, we are only beginning to understand the ways in which genotype and environment dictate this process and contribute to the predictability of evolution. To further our understanding of the role of environment and genotype in determining evolutionary trajectories and their divergence we evolved 24 populations, 12 at 24C and 12 at 37C. These populations were founded from three distinct genotypes, one of which is a hybrid of the others. As expected, growth rate increased with a decelerating rate of return for all combinations of GxE. More interestingly we found significant differences between the patterns of evolution at 24C compared to evolution at 37C. Variation between the three genotypes shrank as evolution proceeded at both temperatures but this happened more quickly at 37C than at 24C. This is true in spite of the fact that we see significantly more divergence in growth rate within genotype for the populations that evolved at 24C when compared to those evolved at 37C. Finally we found that 12/12 of the 37C evolved populations also had an increased growth at 24C while 11/12 of the populations evolved 24C show increased growth rate at 37C. These results demonstrate the roles of environment and genotype in determining evolutionary trajectories.

O-02-WF05

The optimal mating distance resulting from heterosis and genetic incompatibility

Xinzhu Wei[1], Jianzhi Zhang[1]

[1]University of Michigan (United States)

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

O-02-WF06

KRAB-transposase fusion as a source of new regulatory proteins in evolution

Rachel Leigh Cosby[1], Ellen Pritham[2], Cedric Feschotte[1]

[1]Cornell University (United States), [2]University of Utah (United States)

Transcription factors (TFs) are fundamental orchestrators of gene regulation. Yet, the mechanisms underlying the evolution of new TFs are poorly understood. Several TFs are thought to be derived from coopted DNA transposons, suggesting they are potent sources of regulatory innovation. This model remains largely untested because most transposase-derived TFs have deep evolutionary roots, obscuring identification of its originating transposon and binding sites. Here we identify a novel class of proteins, KRAB-transposase fusions (KTFs), which couple a transcriptionally repressive KRAB domain to transposase DNA binding domains. We found that KTFs independently emerged at least 30 times during tetrapod evolution, suggesting a recurrent mechanism for the emergence of new regulatory proteins. To understand the role of KTFs in TF birth, we investigated *KRABINER*, a recently emerged bat KTF derived from a *mariner*transposase. We find that *KRABINER* is expressed and under purifying selection in the bat lineage. Reporter assays indicate that KRABINER has transcriptional repression activity that requires both KRAB and transposase domains as well as sequence-specific binding to its cognate *mariner* elements. Since *mariners* are dispersed throughout the bat genome (>2000 copies), we predict that KRABINER is capable of acting on several loci where it may modulate local gene expression and/or chromatin structure. To test this model, we profiled the transcriptome and epigenome of wild-type and *KRABINER*KO bat cell lines engineered using CRISPR. Our functional analysis of *KRABINER*serves as a powerful model to study a potentially widespread mechanism promoting the birth of new TFs and their associated networks.

O-02-WF07

The impact of Neanderthal ancestry on human phenotypes

Christopher R Robles[1, 2], Andrea Ganna[3, 5, 6], Alexander Gusev[4, 7, 8], David Reich[4, 9, 10], Sriram Sankararaman[1, 2]

[1]University of California Los Angeles (United States), [2]University of California Los Angeles (United States), [3]Massachusetts General Hospital (United States), [4]Broad Institute of MIT and Harvard (United States), [5]Broad Institute of MIT and Harvard (United States), [6]Karolinska Institutet (Sweden), [7]Harvard T.H. Chan School of Public Health (United States), [8]Harvard T.H. Chan School of Public Health (United States), [9]Harvard Medical School (United States), [10]Harvard Medical School (United States)

While a number of instances of admixture between archaic and modern humans have been documented, the specific phenotypic consequences of these admixture events are not fully understood. We analyzed about 80,000 introgressed Neanderthal alleles across 107 distinct phenotypes from European individuals in the UK Biobank. We discovered 1158 independent associations of introgressed Neanderthal alleles with 75 phenotypes that include anthropometric phenotypes, pigmentation, blood and blood pressure, lung and eye-related phenotypes, bone density measures as well as disease phenotypes. The contribution of Neanderthal alleles to phenotypic variation was depleted in a majority of the phenotypes examined, consistent with a model in which Neanderthal alleles entered the modern human gene pool about 50,000 years ago and were subsequently subject to the effects of purifying selection. On the other hand, Neanderthal alleles had an elevated contribution to variation in patterns of male balding, chronotype, wheat intake, lung (Forced Vital Capacity, Forced Expiratory Volume in 1 second) and eye-related phenotypes (corneal hysteresis, corneal resistance factor). Introgressed Neanderthal alleles were also found to have consistent directional effects on several phenotypes: increasing lung capacity and decreasing risk for baldness. Our study reveals that Neanderthal alleles have a broad impact on diverse phenotypes in modern humans.

O-02-WF08

Archaic introgression and gene regulation: a disproportionate degree of Neanderthal ancestry in T-cells enhancers

Martin Silvert[1, 2, 4], Maxime Rotival[1, 2, 3], Lluis Quintana-Murci[1, 2, 3]

[1]Institut Pasteur (France), [2]Institut Pasteur (France), [3]CNRS (France), [4]Universite Pierre et Marie Curie (France)

There is increasing evidence to suggest that admixture with ancient hominins has been a non-negligible source of genetic diversity in non-African populations, with Neanderthal haplotypes covering 1-3% of the genome of modern Eurasians. Recent work has shown that this archaic introgression contributed to phenotypic diversity, mostly through the modulation of gene expression. However, the mechanisms through which Neanderthal variants alter gene expression, and the forces driving the introgression landscape at regulatory regions remain elusive. To answer these questions, we analyzed the Epigenomic Roadmap and 1,000G datasets to study the distribution and frequency of introgressed Neanderthal variants (archaic SNPs, or aSNPs), in Eurasian populations, across functional elements from 127 tissues.

We found that enhancers and promoters are enriched in high frequency aSNPs (>5%). However, while the impact of aSNPs on promoters is stable across tissues, the excess of aSNPs in enhancers is primarily driven by enhancers that are active in T-cells. Interestingly, we show that the enrichment in aSNPs in T-cell enhancers is maintained despite the strong degree of background selection targeting Neanderthal haplotypes in these regions. Furthermore, aSNPs in T-cell enhancers display both high frequencies and high levels of population differentiation (F_{ST}) between European and East-Asian populations, suggesting that introgressed variants in T-cell enhancers were adaptive.
Overall, our study sheds new light onto how admixture with Neanderthals introduced advantageous regulatory variation related to immune responses into Eurasian genomes, targeting in particular phenotypes related to acquired immunity.

Neutral Theory Symposium-01

Becoming Motoo Kimura

Daniel L. Hartl[1]

[1]Harvard University (United States)

Motoo Kimura was one of the most famous geneticists of his generation, not only because of his celebrated and much debated neutral theory of molecular evolution, but also because of his many other important contributions too numerous to detail here. My presentation will focus on what features of his personality, education, and collaborations helped bring him to this level of prominence. He was, first of all, a genius?and I use that term in the literal sense to describe his "exceptional intellectual and creative power." But genius alone is not enough. His genius was also accompanied by an exceptional work ethic that I will tell you about. He also needed recognition and nurturing, which was provided mainly by his mentor and long-term collaborator James F. Crow at the University of Wisconsin and in Crow's occasional visits to Kimura at Mishima. And finally, Kimura was persistent and determined in defending his ideas against critics and aggressively promoting of his own views. The story starts in the early 1950's when one of Crow's favorite students traveled to Japan and discovered an exceptionally bright and creative Japanese student whose name he passed along to Crow. Crow read Kimura's (at that time few) papers, and was impressed. Some years later at a genetics meeting in Madison, Crow befriended a young Japanese man (Kimura) who looked lost, and took him under his wing. I will tell the rest of this story in Crow's own words with pictures of the major players as we go along.

Neutral Theory Symposium-02

Epigenetics, Chromatin, Gene Activity and Near-Neutrality in Evolution

Tomoko Ohta[1]

[1]National Institute of Genetics (Japan)

Consider the evolution of the whole human genome. It is known that only 1.5% of the genome encodes proteins. The ENCODE project reported that the majority of the human genome is transcribed and natural selection may be working on the transcribed region. It is now known that non-coding regions are mainly involved in gene regulation. Genome DNA and histones form chromatin in a cell, and genome activity is controlled by chromatin accessibility. Histones contribute to chromatin structure and function, providing versatile system for accessibility. RNA and non-histone proteins also participate. Epigenetics results from all processes of chromatin structure and function in developing individuals. DNA and histone modifications are important for signaling of pathways of epigenetics. The modification depends both on genetics and environment, and may be or may not be transmitted to the offspring. Thus the connection between genotype and phenotype becomes highly complicated. Numerous molecular machineries are involved and resulting complex network systems are robust and at the same time responsive to environmental cues. At the other level of environment, diet and gut microbiome are now thought to be important for epigenetics, particularly for brain development. Molecular evolution is mainly measured by individual amino acid or nucleotide substitutions. In terms of complicated interactions of numerous molecular machineries, many individual amino acid or nucleotide substitutions are thought to evolve under the nearly neutral process.

Neutral Theory Symposium-03

My memories of Kimura and the neutral theory.

Warren Ewens[1]

[1]University of Pennsylvania (United States)

The modern synthesis, the fusion of Darwinism and Mendelism, had been forged by 1950 largely by the great population geneticists Fisher, Haldane and Wright. Kimura was the greatest population geneticist of the modern (post 1950) era, fully the equal of these three. His initial work built on, and sometimes corrected, their work, both in the stochastic and the deterministic (essentially selective) theory. Given Kimura's work in the selective theory, the neutral theory proposal came as a shock, and some population geneticists said it was "clearly wrong". I never agreed with that opinion, but I did think that the reasons for the theory, in particular the genetic load argument, were misguided. In this talk I elaborate on this view, showing that load calculations derive from an inappropriate reductionist argument. I discuss an alternative way of testing the neutral theory using current data which sometimes supports, and sometimes does not support, the theory. What is the broad impact of the neutral theory? It relates to the past behavior of the population: is the currently observed genetic variation due to selective or non-selective processes in the past? Thus it fits within the currently active retrospective theory, dominated by Kingman's coalescent process, leading to questions about the past, like "When did mitochondrial Eve live?" Kimura never took a leading part in developing the retrospective theory, although the neutral assumption arises in parts of that theory, especially coalescent theory.

Neutral Theory Symposium-04

New mathematical insights into the regulation and optimization of translation dynamics

Yun S. Song[1]

[1]University of California,
Berkeley (United States)

Translation of mRNA to protein is a fundamental biological process underlying all forms of life. Many different biochemical and biophysical mechanisms are involved in this central process, and explaining the major determinants of translation efficiency has been a subject of intense research. In evolutionary biology, many studies have examined the potential effect of codon usage bias on translation, and the role of mutation and selection in shaping this pattern, entailing heated debate between selectionists and neutralists. In this talk, I will discuss new insights into the regulation and optimization of translation dynamics that emerge from a detailed mathematical analysis of a probabilistic model. (Joint work with Dan Erdmann-Pham and Khanh Dao Duc.)

O-03-AP01

Northern Asian Genome Project

JEONG-SUN SEO[1]

[1]Gong-Wu Genomic Medicine Institute (G2MI), Seoul National University Bundang Hospital (Republic of Korea)

We have been working on Northern Asian genome Project. As part of this project, we produced whole-genome sequencing (WGS) data, on average sequencing depth of 15X or 30X, for 1908 samples which composed of 852 Korean, 396 Japanese, 386 Mongolian (the Buryats and the Khalkha Mongols), and 274 Chinese individuals. We discovered millions of novel single nucleotide polymorphism (SNP) and it turned out that large portion of rare SNPs (minor allele frequency < 0.5%) in the 1000 genome project phase 3 were low frequency or even common in our dataset. We were able to construct robust imputation panel for Northern Asian populations which produces great imputation accuracy for rare and low frequency SNPs.

Series of analyses demonstrated distinctive population structure of the Mongolians from the other East Asians. We identified significant gene flow from the Buryats to the Finnish and we found that Buryat admixture existed in the Finnish genome. Since part of these samples was included in the GA100K (Genome Asia 100K) pilot project, we firstly analyzed the mitochondrial haplogroups, which show a clear differentiation between Northern and Southern part of Asian haplogroups. These observations suggest that at least two waves of maternal lineage migration from North Asia to South Asia.

O-03-AP02

High-coverage sequencing of diverse human populations in the HGDP-CEPH panel

Anders Bergstrom[1], Shane A McCarthy[1, 2], Ruoyun Hui[2], Qasim Ayub[1, 3, 4], Petr Danecek[1], Mohamed A Almarri[1], Javier Prado-Martinez[1], Yuan Chen[1], John A Kamm[1, 2], Joshua Randall[1], Swapan Mallick[5, 6, 8], Pontus Skoglund[7], Manjinder S Sandhu[1], David Reich[5, 6, 8], Scally Aylwyn[2], Yali Xue[1], Richard Durbin[1, 2], Chris Tyler-Smith[1]

[1]Wellcome Sanger Institute (United Kingdom), [2]University of Cambridge (United Kingdom), [3]Monash University Malaysia (Malaysia), [4]Monash University Malaysia (Malaysia), [5]Harvard Medical School (United States), [6]Broad Institute of Harvard and MIT (United States), [7]The Francis Crick Institute (United Kingdom), [8]Harvard Medical School (United States)

The Human-Genome Diversity Project (HGDP-CEPH) panel consists of samples from 53 diverse human populations from six continents, including many of particular evolutionary, anthropological, and linguistic interest. We have generated a dataset consisting of 929 high-coverage genome sequences from this panel, including 120 samples previously sequenced as part of the Simons Genome Diversity Project. We have also sequenced 26 of the individuals using 10x Genomics linked-read technology, enabling physical haplotype phasing. We perform analyses of detailed effective population size histories, including in the last 10 thousand years, genetic split times, archaic admixture and patterns of rare variant sharing, to further our understanding of human population history and genetic variation. The data is open access with no analysis restrictions, and we hope it will be constitute a useful resource for the human genetics community.

O-03-AP03

Back migrations of Southeast Asian ancestors to South Asia during the Last Glacial Maximum

Namrata Kalsi[1], Stephan Schuster[1], Hie Lim Kim[1, 2], The GenomeAsia 100K consortium

[1]Nanyang Technological University (Singapore), [2]Nanyang Technological University (Singapore)

The ancestor of the South Asian tribal people, the oldest groups in the region, has not been clearly known, due to lack of genome data. Our study used more than 800 high coverage and whole genome data, including 470 South Asians and 280 Southeast Asians, generated by the GenomeAsia 100K project. We found admixture signals between tribal South Asians and indigenous people living in the Malay Peninsula, Kensiu and Kintak, who are the earliest settlers in the region and also speak Austroasiatic languages. According to the MSMC analysis, the Malay groups have once increased their population size during the Last Glacial Maximum. During this period, the sea level in the Southeast Asia region was dramatically increased. In the same time period, the South Asian tribes show admixture signal and higher cross-coalescence rate with Kensiu than other South Asian groups. These results support the possible contacts between the ancestors of South and Southeast Asian groups, during the Last Glacial Maximum. Based on analyses using sex chromosomes and mitochondrial genomes, we found abundant male Southeast Asian ancestries in the South Asian tribal groups, which is consistent with previous studies. Overall, our study suggests the possibility of male-driven migrations from Southeast Asia to South Asia and noticeable contacts between them, with the sea level rise in Southeast Asia.

O-03-AP04

Human population history in the southwestern coastal region of Sea of Okhotsk, inferred from ancient genome analysis

Takehiro Sato[1], Noboru Adachi[2], Ryosuke Kimura[3], Minoru Yoneda[4], Hiroki Oota[5], Atsushi Tajima[1], Atsushi Toyoda[6], Hiromi Matsumae[7, 8], Kae Koganebuchi[3], Kentaro K Shimizu[7, 8], Tsunehiko Hanihara[4], Andrzej Weber[9], Hirofumi Kato[10], Hajime Ishida[3]

[1]Kanazawa University (Japan), [2]University of Yamanashi (Japan), [3]University of the Ryukyus (Japan), [4]The University of Tokyo (Japan), [5]Kitasato University (Japan), [6]National Institute of Genetics (Japan), [7]University of Zurich (Switzerland), [8]Yokohama City University (Japan), [9]University of Alberta (Canada), [10]Hokkaido University (Japan)

In 2013, an ancient human skeleton (NAT002) of the prehistoric Okhotsk culture was excavated from Hamanaka 2 site, Rebun Island, northern Japan. Radiocarbon age of NAT002 was 1060-1155 (68.2%) calAD. Some bones of this individual were affected by severe hyperostosis, suggesting SAPHO syndrome. To investigate the genetic features of the Okhotsk people, we extracted DNA from 3rd molars of NAT002 and performed whole genome sequencing. As a result of sequencing for 18 NGS libraries, the sequence data with 35-fold coverage was obtained. The typical deamination pattern and sufficiently low contamination rate were observed, ensuring the DNA authenticity. The mtDNA haplotype of NAT002 was assigned to haplogroup G1b, which is commonly observed among the modern North Asian populations. Results of outgroup f_3 test, PCA, and ADMIXTURE analysis indicated that NAT002 was genetically close to the modern Nivkh and Ulch, who are living around the northern Sakhalin and the Lower Amur Basin. In addition, TREEMIX analysis indicated gene flow from the Ainu to NAT002. These findings suggest the past human migration from the Lower Amur region to the northern part of the Japanese archipelago and the admixture between the Okhotsk and Ainu lineages, corresponding to the archaeological evidences. HLA typing indicated that NAT002 possessed HLA-B40 allele, which has been reported as one of the risk factors of ankylosing spondylitis, reactive arthritis, and undifferentiated spondyloarthritis. These diseases are classified into seronegative arthritis as well as SAPHO syndrome. Therefore, HLA type of NAT002 might be one of the cause of the hyperostosis.

O-03-AP05

Physiological and genetic adaptations to diving in Sea Nomads

Melissa Ilardo[1], Ida Moltke[2], Thorfinn Korneliussen[1], Jade Cheng[3], Aaron Stern[3], Fernando Racimo[1], Peter de Barros Damgaard[1], Martin Sikora[1], Andaine Seguin-Orlando[1, 4], Simon Rasmussen[8], Inge van den Munckhof[5], Rob ter Horst[5], Rasmus Nielsen[3], Eske Willerlsev[1, 6, 7]

[1]University of Copenhagen (Denmark), [2]University of Copenhagen (Denmark), [3]University of California at Berkeley (United States), [4]Technical University of Denmark (Denmark), [5]Radboud University (Netherlands), [6]University of Cambridge (United Kingdom), [7]Sanger Institute (United Kingdom), [8]Danish Technical University (Denmark)

Understanding the physiology and genetics of hypoxia tolerance in humans has important medical implications, however this phenomenon has thus far only been investigated in populations living at high altitudes. Another system, yet to be explored, is humans who engage in breath-hold diving. The indigenous Bajau people (sometimes referred to as Sea Nomads) of Southeast Asia live a subsistence lifestyle based on breath-hold diving. They are renowned for their extraordinary breath holding abilities, diving over 70 meters with nothing more than a set of weights and a pair of wooden goggles and spending 60 percent of their daily working time underwater. However, it is unknown whether these observed abilities have a genetic basis. Using a comparative genomic study, we show that natural selection on genetic variants in the PDE10A gene have increased the spleen size in the Bajau, providing them with a larger reservoir of oxygenated red blood cells. We also find evidence of strong natural selection specific to the Bajau on BDKRB2, a gene affecting the human diving reflex. Thus, the Bajau, and possibly other natural diving populations, provide a new opportunity to study human adaptation to hypoxia tolerance.

O-03-AP06

Evolutionary history and adaptation from whole-genome sequences of a pygmy population of Flores Island, Indonesia

Serena Tucci[1, 2, 3], Sam H. Vohr[4], Rajiv C. McCoy[1, 2], Benjamin Vernot[5], Matthew R. Robinson[6, 7], Chiara Barbieri[8], Wenqing Fu[9], Gludhug A. Purnomo[10], Herawati Sudoyo[10, 11], Guido Barbujani[3], Peter M. Visscher[12, 13], Joshua M. Akey[1, 2], Richard E. Green[4]

[1]Princeton University (United States), [2]Princeton University (United States), [3]University of Ferrara (Italy), [4]University of California, Santa Cruz (United States), [5]Max Planck Institute for Evolutionary Anthropology (Germany), [6]University of Lausanne (Switzerland), [7]Quatier Sorge, Lausanne (Switzerland), [8]Max Planck Institute for the Science of Human History (Germany), [9]University of Washington, Seattle (United States), [10]Eijkman Institute for Molecular Biology, Jakarta (Indonesia), [11]University of Indonesia, Jakarta (Indonesia), [12]The University of Queensland, Brisbane (Australia), [13]The University of Queensland, Brisbane (Australia)

Despite the pivotal role that Flores Island (Eastern Indonesia) plays in understanding human evolutionary history, genomic data from this region is lacking. Here we generate genome-scale data for over 2 million SNPs and the first whole-genome sequences from a pygmy population living on Flores, near the Liang Bua cave where remains of the enigmatic small-bodied hominin species, *Homo floresiensis*, were recently found. The genomes of the Flores pygmies reveal a complex history of hominin admixture east of the Wallace's line and a striking signature of recent positive selection encompassing the FADS gene cluster on chromosome 11, encoding for fatty acid desaturases that regulate the metabolism of long-chain polyunsaturated fatty acids (LC-PUFA). Our results add to emerging evidence that the FADS region has been a recurrent target of selection in diverse human populations, possibly in response to changing diets. Further, we find that polygenic selection acting on standing variation contributed to the short-stature phenotype of the Flores pygmies, which suggests that multiple, independent instances of hominin insular dwarfism occurred on Flores Island.

O-03-AP07

Allele frequency of pathogenic variants in a Japanese population based on the whole-genome reference panel of ToMMo and inter-population differences

Yumi Yamaguchi-Kabata[1], Jun Yasuda[1], Yoichi Suzuki[1], Nobuo Fuse[1], Hiroshi Kawame[1], Masao Nagasaki[1], Fumiki Katsuoka[1], Yosuke Kawai[1], Kaname Kojima[1], Takahiro Mimori[1], Nobuo Yaegashi[1], Kengo Kinoshita[1], Shigeo Kure[1], Masayuki Yamamoto[1]

[1]Tohoku University (Japan)

Clarifying population frequencies of pathological variants is essential to construct infrastructure for genomic medicine for different populations, because incidence rates for inherited diseases vary among population. Here, by analyzing a whole-genome reference panel of a Japanese population based on 2,049 individuals, we characterized the genomes of Japanese individuals in terms of allele frequency of functional and pathological variants. Among the 28 million autosomal single nucleotide variants (SNVs) in 2KJPN, 6,862 SNVs were identified as pathologically annotated variants that overlap with reported pathogenic variants. In particular, we focus on estimating carrier frequency of autosomal recessive disorders and identifying genome-wide pathogenic variants showing higher frequencies in Japanese compared with other ethnic groups. By focusing on 32 genes for 17 congenital metabolic diseases for newborn screening in Japan, we identified reported pathogenic variants and estimated their carrier frequencies by variant filtering based on variant annotations and allele frequencies. Our results showed that variant frequencies were relatively higher in *PCCA* and *SLC25A13* among the genes examined. In addition, we identified genome-wide pathogenic SNVs showing higher allele frequencies in 2KJPN than in other ethnic groups, and these SNVs included clinically important variants (e.g. *CDH1* for hereditary diffuse gastric cancer, *APRT* deficiency and *CETP* for cholesterol ester transfer protein deficiency) for personalized medicine and prevention for the Japanese. Our results and ongoing activities on variant curation would lay the foundation for i) evaluating the relationships of genomic variants and disease prevalence, and ii) improving diagnostic strategies in Japan and East Asia.

O-01-AE01

Insights into mutational pathways of biochemical adaptation using ancestral protein resurrection

Jay F Storz[1], Xiaojia Zhu[2], Chandrasekhar Natarajan[1], Anthony Signore[1], Federico Hoffmann[3], Jeremy Tame[4], Fumin Lei[2], Angela Fago[5]

[1]University of Nebraska (United States), [2]Chinese Academy of Sciences (China), [3]Mississippi State University (United States), [4]Yokohama City University (Japan), [5]Aarhus University (Denmark)

During the evolution of novel protein functions, some selectively fixed mutations may be directly causative and others may be purely compensatory. The relative contribution of these two classes of mutation to adaptive evolution depends on the form and prevalence of mutational pleiotropy. To investigate the nature of adaptive substitutions and their pleiotropic effects, we are using a protein engineering approach to characterize the molecular basis of evolved changes in avian and crocodilian hemoglobins. To test the effects of observed substitutions on evolutionarily relevant genetic backgrounds, we synthesized all possible genotypic intermediates in specific lines of descent that connect genotypes with derived vs. ancestral functions. Site-directed mutagenesis experiments have enabled us to identify specific cases where genetic compensation of deleterious pleiotropic effects plays an important role in the evolution of novel protein functions.

O-01-AE02

An animal without aerobic cellular respiration

Dorothee Huchon[1, 2], Dayana Yahalomi[1], Sally Chang[3], Herve Philippe[4], Paulyn Cartwright[3], Jerri Bartholomew[5], Stephen Atkinson[5]

[1]Tel-Aviv University (Israel), [2]Tel-Aviv University (Israel), [3]University of Kansas (United States), [4]UMR CNRS 5321 (France), [5]Oregon State University (United States)

Aerobic cellular respiration is the hallmark of mitochondria within eukaryotes, and only a few protists have lost this ability. Heretofore, all animal mitochondria were thought to possess their own genomes, which encode the proteins essential for cellular respiration. We have discovered that the myxozoan species *Henneguya salminicola*, has lost its mitochondrial genome, and thus the ability to perform aerobic cellular respiration. The Myxozoa are a large group of parasitic Cnidaria (corals, jellyfish) with a very simple body organization, reduced to a few cells.

For this species, we obtained genomic and transcriptomic assemblies with good completeness parameters and high base coverage. In both genome and transcriptome assemblies we could not identify any mitochondrial genes. We also showed that in parallel with the absence of mitochondrial genomes, this myxozoan has lost nuclear genes involved in transcription and replication of the mitochondrial genome. Similarly, key nuclear mitochondrial metabolic enzymes such as pyruvate dehydrogenase are detectable only as pseudogenes. The identification of these pseudogene sequences supports the view that the absence of a mitochondrial genome is not an assembly artifact.

The absence of a mitochondrial genome is surprising since other myxozoans have been shown to possess mitochondrial genes. Evolutionary drivers for the loss of the mitochondrial genomes in *Henneguya salminicola* might be related to low oxygen availability in parts of the life cycle: for example, some annelid hosts can live in anoxic environments, and certain fish tissues (where parasite cysts can form) are oxygen-poor.

O-01-AE03

Why panmictic bacterial species are rare

Daniel Falush[1], Yujun Cui[2], Yang Chao[2]

[1]University of Bath (United Kingdom), [2] Beijing Institute of Microbiology and Epidemiology (China)

Many bacteria recombine frequently. Nevertheless, despite enormous population sizes and effective dispersal, bacterial populations which are panmictic, with well mixed gene pools, are rare. Specifically, in a survey of well-sampled bacterial species, only two have an estimate of a scaled effective size NeR of substantially over 100, namely the Asian population of Vibrio parahaemolyticus and Helicobacter pylori. The site frequency spectrum of Asian V. parahaemolyticus is out of equilibrium, showing that its high population size is a relatively recent phenomenon, probably reflecting a larger habitat since the last interglacial. If this population size was stable on a macro evolutionary timescale, nucleotide diversity would be predicted to increase 20-fold or more. However, this would result in higher nucleotide diversity than is observed in any well characterized bacterial species. A more likely outcome is that barriers to gene flow would start to develop. Other recombinogenic species, such as V. cholera and Klebsiella pneumoniae show diagnostic features of early-stage speciation. This model has many implications for interpreting evolutionary and functional diversity of recombinogenic bacterial species.

O-01-AE04

Epigenetic conservation of human duplicated genes associated to their transposable element neighborhood

Romain Lannes[3], Carene Rizzon[2], Emmanuelle Lerat[1]

[1]CNRS - Universite Lyon 1 (France), [2]University of Evry (France), [3]Universite Pierre et Marie Curie (France)

Epigenetic modifications correspond to heritable changes caused by modifications in the chromatin structure rather than in the DNA molecule itself. Understanding the mechanisms involved in initiation, maintenance and heritability of epigenetic states is important to better understand the cell functioning and to evaluate their implication in the genome evolution. At an intraspecific level, epigenetic modifications may be implicated in functional divergence by facilitating tissue-specific regulation.

The human genome contains <2% of protein-coding genes, whereas transposable elements (TEs) represent more than the half. Because of their presence, TEs have a significant impact on genome evolution by promoting various mutations, which are expected to be mostly deleterious. To counteract these effects, TEs are regulated via epigenetic mechanisms to suppress or silence the TE activity. A change in the local epigenetic landscape associated with the presence of TEs is expected to affect the expression of neighboring genes since these modifications occurring at TE sequences can spread to neighboring sequences.

The question we want to address is how the epigenetic modifications of genes are conserved and what is the role of TEs in this conservation. For that, we have studied the conservation of the epigenome at an intraspecific level in human. By measuring, in a given environmental condition, the divergence of epigenetic modifications associated to duplicated genes and linked to the differential presence of TEs near the genes, we have determined the impact of TEs on epigenetic changes associated with the time since duplication, and the function divergence of these genes.

O-01-AE05

Germline and somatic mutation rates in a single cell

Rebecca Zufall[1]

[1]University of Houston (United States)

Since mutation is the ultimate source of all genetic variation, it is important to understand the forces that drive the evolution of mutation rate. Previously, we found that the rate of mutation in the germline genome of the ciliate *Tetrahymena thermophila*is unusually low. Ciliates are microbial eukaryotes that maintain two types of nuclei in each cell, a transcriptionally-active somatic macronucleus and a quiescent germline micronucleus. Mutations in the germline genome are only expressed after sexual reproduction, but not during asexual reproduction, which makes up most of the life cycle. We hypothesize that the low rate of germline mutation evolved in response to selection to decrease the high mutational load that would be generated after sexual reproduction following these many rounds of asexual division. To test this hypothesis, we measured the mutation rate in the somatic genome, which expresses mutations every generation. Consistent with our prediction, we find that the somatic genome has a higher genome-wide mutation rate than the germline genome.

O-01-AE06

Polarization of ancestor relations reveals the order of traits in the evolution of cyanobacterial multicellularity

Katrin Hammerscmidt[1], Giddy Landan[1], Fernando Domingues Kuemmel Tria[1], Tal Dagan[1]

[1]Kiel University (Germany)

Understanding the events during the evolution from unicellular organisms to multicellularity is a long-standing challenge. The first known transition to multicellularity occurred more than 3 billion years ago in cyanobacteria. This is a monophyletic phylum that nowadays still includes unicellular and filamentous genera, some of which represent the peak of prokaryotic complexity. Previous research on the evolution of multicellularity in this phylum has been hampered by the lack of knowledge on the phylogenetic relations among unicellular and filamentous genera. We present a novel approach that does not depend on previous knowledge of phylogenetic relationships but infers ancestor-descendent relations of ancestral nodes in phylogenetic trees. Combining the method with the assignment of phenotypic traits to ancestral nodes enable the inference of pairwise relative priority, without the need to postulate a species tree. Applying our approach to 199 cyanobacterial genomes yields a temporal sorting of traits. Our results reveal, for example, that the formation of filaments co-occurred with the ability to fix nitrogen before higher complexity evolved. This finding is in line with theory that predicts the trade-off between oxygen-producing photosynthesis and oxygen-sensitive nitrogen fixation to be the main driving force for multicellularity in this phylum.

O-01-AE07

Evolution of mRNA editing and linear multipartite genome in mitochondria of calcaronean sponges

Dennis Lavrov[1]

[1]Iowa State University (United States)

Calcaronea is a group of common marine sponges that display two remarkable features in their mitochondrial biology: i) a widespread messenger RNA (mRNA) editing and ii) mitochondrial (mt-) genomes comprised of multiple small linear chromosomes. Both features are highly unusual in animals: only a few isolated examples of editing have been reported so far in animal species, while linear chromosomes have been found only in calcareous sponges and medusozoan cnidarians. Although, phenotypically, the editing found in calcaronean sponges -- insertions of single or double uridylate (U) in pre-existing short poly(U) tracts -- is similar to that in some other eukaryotes, it is unique by the fact that both the occurrence of editing and its pattern (the number of inserted Us) can be predicted from the primary DNA sequence. In a sense, the 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 I will explore relationship between the presence of editing, high rate of sequence evolution, and unusual (linear and multipartite) mitochondrial genome organization in calcaronean sponges. More broadly, I will look at the co-evolution between DNA- and RNA-level processes.

O-01-AE08

Impact of feminizing Wolbachia endosymbionts on the evolution of a male heterogametic system of sex chromosomes (XY-XX)

Thomas Becking[1], Mohamed Amine Chebbi[1], Isabelle Giraud[1], Bouziane Moumen[1], Jean Peccoud[1], Clement Gilbert[2], Richard Cordaux[1]

[1]CNRS / University of Poitiers (France), [2]Universite Paris-Sud (France)

In animals, male/female development is generally determined by genetic factors carried by sex chromosomes. Sex determination can also be influenced by maternally-inherited, feminizing microbial endosymbionts. The isopod *Armadillidium vulgare* has a female heterogametic system of sex chromosomes (ZZ males/ZW females), and many populations harbor *Wolbachia* bacterial endosymbionts which can convert genetic males into phenotypic females. As predicted by models, the W (female) sex chromosome has been lost in *Wolbachia*-infected lines and all individuals are ZZ genetic males. Individuals inheriting *Wolbachia* develop as females while those lacking *Wolbachia* develop as males. By analogy, feminizing *Wolbachia* infection in male heterogametic species (XY males/XX females) is expected to lead to X chromosome loss and all individuals should end up carrying YY sex chromosomes. This prediction requires that YY genotypes are viable, which may be possible only if X and Y chromosomes both carry all vital loci and are genetically very similar. To empirically test these predictions, we sequenced the genome and analyzed pedigrees of *Armadillidium nasatum*, an XY-XX isopod infected by feminizing *Wolbachia*. We found that the X and Y chromosomes are genetically very similar and that YY females are viable and fertile. Nevertheless, YY females carrying *Wolbachia* produce only males (lacking *Wolbachia*), suggesting that the YY genotype somehow prevents *Wolbachia* vertical transmission. Thus, the X chromosome cannot be lost in *A. nasatum*lines infected by feminizing *Wolbachia*. We conclude that bacterial endosymbionts can have a profound impact on the evolution of fundamental biological processes in animals, such as sex determination.

O-04-EG01

Biotic and abiotic influences on balancing selection in nature

Thomas Mitchell-Olds[1], Julius Mojica[1], Lauren Carley[1], Baosheng Wang[1], Emily Chan[1], Nadeesha Perera[1], Carrie Olson-Manning[2]

[1]Duke University (United States), [2]Augustana University (United States)

How genetic variation is maintained in nature is a central question in evolutionary biology. Balancing selection is frequently invoked as an explanation for persistent genetic and phenotypic variation, but few studies have demonstrated its role in maintaining in complex trait variation in natural populations. We combined manipulative field experiments, population genetics, and genome-wide association mapping to investigate selection on variation in defensive chemistry in a wild relative of Arabidopsis that is native to the Rocky Mountains.

We find that antagonistic pleiotropy of a gene controlling defensive chemistry influences selection on complex trait variation in natural populations. Ecological studies show that these tradeoffs appear to be driven by variation in two selective agents - herbivore damage and drought stress, which often favor contrasting physiological strategies in this species. Our results reveal complex variation in pleiotropy itself, with a single locus affecting variation in functional traits (e.g., water use, herbivore defense) and fitness components (e.g., survival, fecundity) to various extents, depending on environment. By integrating ecological field experiments with population genomics, we are able to test for balancing selection, and to gain a mechanistic understanding of the genetic and molecular underpinnings of antagonistic pleiotropy for these traits.

O-04-EG02

From Population Genomics to Mechanistic understanding: A possible role in adaptation to low calcium serpentine soils by Two Pore Channel 1

Sian M Bray[1], Levi J Yant[1]

[1]John Innes Centre (United Kingdom)

Serpentine soils are enormously hostile to plant life. In addition to extremely low Ca:Mg ratios, they present a perfect storm of high porosity, high toxic metal concentrations and low macronutrient levels. In spite of this, some plants have repeatedly adapted to these harsh conditions offering fascinating models of extremophile adaptation. A recent population genomic study by our group identified TPC1 as exhibiting strong signals of adaptive evolution associated with serpentine colonisation in outcrossing *Arabidopsis arenosa*. TPC1 is the primary calcium transporter in the vacuole and has functional roles in stress signalling.

Close examination of the TPC1 locus reveals a single missense mutation, Leu634, that is exclusive to the serpentine population: of 946 alleles sequenced, all the plants in the serpentine populations carry Leu634. Indeed, the site is conserved as either Valine or Isolucine (but never Leucine) across plant, animal and archaeal diversity. Strikingly, this unique polymorphism is directly adjacent to a residue shown to mediate selectivity. Together, this suggests that Leu634 mediates a functional effect to modulate the selectivity of the channel.

We are currently investigating these alleles using electrophysiology and structural biology, as well as broader population genomic scans for selection across plant diversity. We aim to reveal the mechanism underlying this striking evolved change that is entirely specific to its calcium-starved extreme environment.

O-04-EG03

Molecular mechanisms and evolutionary processes underlying genetic assimilation in the digestive tract of medaka

Takafumi Katsumura[1, 2], Suguru Sato[3], Kana Yamashita[3], Shoji Oda[4], Takashi Gakuhari[5], Shodai Tanaka[3], Tadashi Imai[6], Yasutoshi Yoshiura[6], Hirohiko Takeshima[7], Yasuyuki Hashiguchi[8], Hiroshi Mitani[4], Motoyuki Ogawa[2], Hideaki Takeuchi[1], Hiroki Oota[2]

[1]Okayama University (Japan), [2]Kitasato University School of Medicine (Japan), [3]Kitasato University (Japan), [4]The University of Tokyo (Japan), [5]Kanazawa University (Japan), [6]Japan Fisheries Research and Education Agency (Japan), [7]Tokai University (Japan), [8]Osaka Medical College (Japan)

Many organisms exhibit phenotypic plasticity in their traits according to environmental changes. However, it has long been discussed whether the phenotypic plasticity will contribute to adaptive evolution of natural populations. To address the question, we should investigate the molecular mechanisms which caused phenotypic plasticity and fixed the phenotype, and then examine the order in which these occurred. In this study, we focus on digestive tract showing seasonal phenotype plasticity, using a small freshwater fish model organism, medaka. We investigated the DNA methylation region correlated to the seasonal difference in the length of digestive tract by MBD-seq and the DNA mutation associated to the difference in the length between the local populations by RAD-seq. We found a seasonally-varying DNA methylation region in the CpG island (CGI) upstream of a gene involved in a signal transduction that suppressed axon guidance of neurons. This gene expression was up- and down-regulated in winter and summer, respectively, suggesting that the methylation is related to plasticity of digestive tract. We also found a mutation on a gene involved in axon outgrowth of neurons, which were associated with the differences in digestive tract length between the populations. Furthermore, the CGI had been lost in the DNA methylation region described above in the population in which this mutation was fixed. In this presentation, based on the natural history of medaka, we discuss the two steps: the appearance/disappearance of CGI related to phenotypic plasticity and the fixation of DNA mutation regulating plasticity in one direction.

O-04-EG04

Whole Genome Sequencing Reveals Metabolic Adaptation to High-Altitude Hypoxia in a Tibetan Locust

Ding Ding[1, 2], Guang-Jian Liu[3], Bing Chen[1], Le Kang[1]

[1]Institute of Zoology, Chinese Academy of Sciences (China), [2]University of Chinese Academy of Sciences (China), [3]Novogene Bioinformatics Institute (China)

Animals indigenous to very high altitudes have evolved a distinctive suite of physiological traits to cope with the environmental hypoxia. Different from vertebrates, insects employ a tracheal system for efficient oxygen delivery. However, the genetic mechanisms for an adaptation to high-altitude hypoxia by insects remain unexplored. Here we report a whole-genome resequencing of the Tibetan and lowland populations of the migratory locust (Locusta migratoria). The positively selected genes were highly enriched in carbon metabolism pathways, notably the gene PTPN1 that encodes PTP1B, a phosphatase of insulin receptor. Transcriptome analysis demonstrated that the Tibetan locusts are distinctive from the lowland locusts with a well maintained carbon and energy metabolism processes in response to hypoxia. Functional analysis showed that a coding mutation of PTPN1 in Tibetan locusts regulates the hypoxia-induced metabolic adaptation through mediating the activity of insulin signaling pathway. This study firstly uncovers the genetic mechanism of wild insects for high-altitude hypoxia adaptation and facilitates our understanding of an organism's metabolic adaptation under hypoxic environment.

O-04-EG05

Severe loss of genetic diversity due to cold-temperature adaptation in a progressively warming climate: the Alpine marmot genome

Toni Gossmann[1]

[1]University of Sheffield (United Kingdom)

The Alpine Marmot (*Marmota marmota*) is a large, ground-dwelling rodent species, known for its social behaviour, long hibernation, and dominance within the high-altitude Alpine habitat. We constructed a *M. marmota* reference genome, which revealed selective cold-climate adaptation affecting metabolic genes required for fatty acid metabolism, fat storage, and saturation of white adipose tissue. Unexpectedly, we report that despite the rodent is highly social and possesses a large, alpine-spanning population, it has one of the least heterozygous genomes among mammalia. The marmot is less genetically diverse as extreme cases of isolated mammals like Mountain Gorilla, Iberian Lynx or Island Fox, so that a lower degree of heterozygosity was only detected in artificially inbred laboratory mice. Reconstructing the genetic past of three independent marmot populations, we find evidence that the marmot did however once possess a high degree of genetic diversity. This diversity has been lost over time and is explained by a cease in migration between adjacent population. In all 12 sequence whole-genomes, inbreeding had progressed to the extent of genomic flatlining, with purifying selection dramatically reduced. Our results hence suggest that a continued pressure to adapt to cold-temperatures during hibernation, in an otherwise warming climate, has insularized individual marmot-subpopulation to the extent that an overall large population size was not sufficient to maintain genetic diversity. Our results show that progressive changes in climate can have dramatic effects on genetic diversity for specialized mammals, revealing why Alpine species are among those most affected by climate change.

O-04-EG06

The importance and evolutionary dynamics of chromosomal inversions in sympatric Neotropical cichlid radiations

Alexander Nater[1], Andreas F. Kautt[1], Axel Meyer[1]

[1]University of Konstanz (Germany)

Despite growing empirical evidence for the occurrence of sympatric speciation, the processes involved in genomic divergence in the face of gene flow, and particularly the transition from localized to genome-wide divergence, remain poorly understood. Chromosomal inversions might play a prominent role in sympatric speciation, since they can capture multiple beneficial alleles and prevent recombination from breaking up such locally adapted complexes. However, due to the lack of suitable population genomic data sets in non-model organisms, the evolutionary significance of this effect remains largely unexplored. Capitalizing on multiple long-read reference genomes and whole-genome resequencing of over 400 individuals of neotropical Midas cichlids, we investigate the role of inversions in promoting speciation with gene flow in multiple rapid crater lake radiations. In a first step, we ran extensive forward simulations to assess the population genetic parameters under which inversion can get differentially fixed under strong gene flow between incipient species. These simulations revealed characteristic genomic footprints of inversions involved in ecological divergence, allowing for targeted searches in resequenced genomes. After stringent filtering, we found dozens of inversions segregating within the Lake Apoyo and Lake Xiloa radiations. Interestingly, we identified multiple highly differentiated inversions with characteristic signatures of divergent selection between morphologically divergent sympatric species. Contrasting the prevalence of differentially fixed inversions in sympatric and allopatric species pairs across a broad range of divergence times, effective population sizes, and ecological settings sheds light on the genomic processes driving rapid adaptive divergence, providing novel insights into the processes generating global biodiversity.

O-04-EG07

Evolution of mating systems by a dominant mutation conferred epigenetically by a small RNA in the allopolyploid Arabidopsis

Kentaro K. Shimizu[1], Chow Lih Yew[1], Takashi Tsuchimatsu[2], Masaomi Hatakeyama[1], Rie Shimizu-Inatsugi[1], Shinsuke Yasuda[3], Seiji Takayama[4]

[1]University of Zurich (Switzerland), [2]Chiba University (Japan), [3]NAIST (Japan), [4]University of Tokyo (Japan)

Polyploid speciation and the evolution of self-compatibility typically co-occur, and both of them are often associated with environmental changes affecting mating systems. Paradoxically, polyploidization may prevent adaptation such as the evolution of self-compatibility due to the extra gene copies that can mask the effect of recessive mutations as Stebbins pointed out long time ago. Here we showed that epigenetic regulation by a small RNA conferred a dominant self-compatible mutation. We focus on a self-compatible allotetraploid Arabidopsis kamchatica, which is estimated to have originated during recent glacial cycles (about 20,000-250,000 years ago) from two parental species, A. halleri and A. lyrata. A. kamchatica is self-compatible, in contrast to predominantly self-incompatible two parental species. Among the duplicated two S-haplogroups, the haplogroup B harbors a loss-of-function mutation in the male self-incompatibility gene S-locus cysteine-rich protein (SCR) / SP11 at the S-locus, and more importantly, a small RNA that epigenetically suppress the expression of the SCR of the recessive haplogroup D. As a result, this haplogroup confers a self-compatible dominantly. We validated the importance of the small RNA by transgenic restoration of ancestral self-incompatibility. Our study provides an empirical support for Haldane's sieve, in which dominant or partially dominant beneficial mutations but not recessive ones are more likely to be fixed in adaptive evolution both in diploid and polyploid species. We suggest that epigenetic dominance regulation by a small RNA contributed to rapid evolution in response to environmental changes.

O-04-EG08

Single generation selection experiments reveal adaptive loci in highly fecund, long-lived species

Melissa H Pespeni[1], Reid Brennan[1], April Garrett[1], Kaitlin Huber[1], Heidi Hargarten[1]

[1]University of Vermont (United States)

Standing genetic variation is critical for adaptation to rapidly changing conditions, but to identify the types of variation important for survival and to learn how this variation is maintained in natural populations remain important challenges, particularly for long-lived organisms. Here we develop a pooled capture sequencing approach using genomic DNA and apply the approach to a single generation acidification experiment using larvae generated from all crosses of 25 wild-caught, outbred adult purple urchins. The millions of offspring were pooled, sampled to estimate initial allele frequencies, then split and reared in replicate control and low pH conditions to induce selective mortality. We found consistent, directional changes in allele frequency in response to low pH treatments, particularly in regulatory regions of the genome. Responsive loci were enriched for biological functions related to low pH tolerance, indicating an adaptive response to the selection regime. Patterns of variation across the genome and across populations suggest that spatially varying selection, rather than balancing selection, maintains this adaptive variation. These results demonstrate the utility of a pooled capture approach to identify the mechanisms underlying tolerance to a selective pressure and provide insight into the biological and evolutionary processes enabling adaptive responses to these conditions. We explore how the genetic bases of these complex adaptive phenotypes can be functionally validated with network analyses and protein modeling and how measuring the genomic impact of selective mortality can be used to model population persistence in future conditions.

O-04-EG09

Genomic footprints of past selection at a local scale associate with present phenotypic variation in teosintes

Natalia Martinez[1], Margaux-Alison Fustier[1], Jonas A Aguirre-Liguori[3], Helene Corti[1], Venon Antony[1], Fabrice Dumas[1], Falque Matthieu[1], Agnes Rousselet[1], Hannes Dittberner[2], Salvador Montes-Hernandez[4], Luis E Eguiarte[3], Yves Vigouroux[5], Domenica Manicacci[1], Maud Tenaillon[1]

[1]INRA - Universite Paris-Sud - CNRS - AgroParisTech (France), [2]University of Cologne Biocenter (Germany), [3]Universidad Nacional Autonoma de Mexico (Mexico), [4]INIFAP (Mexico), [5]IRD (France)

We combined reverse ecology and association mapping to mine the determinants of local adaptation in teosintes, the closest wild relatives of maize. First, we applied methods based on allelic differentiation and correlation with environmental variables to detect signals of past selection in teosinte populations growing along two altitudinal gradients in Mexico. Second, we selected a subset of 218 SNPs among our best candidates to test the association between genotypic and phenotypic variation measured at 18 traits in 11 of the teosinte populations. We determined genotypes of >1,500 individuals, for which traits were measured in two common gardens for two years at mid-altitude. We controlled for underlying population structure using 35 microsatellites. Over 50% of candidate SNPs displayed an association with at least one of the traits. This enrichment was particularly pronounced for male and female flowering time and the number of lateral branches, for which we also found evidence for spatially varying selection. We further analyzed correlations between traits that may be adaptive or, on the contrary, hamper the response to natural selection. We will discuss the power of reverse ecology methods to identify SNPs affecting present phenotypic variation.

O-04-EG10

Detecting and interpreting the genomic basis of convergent local adaptation

Sam Yeaman[1], Michael Whitlock[2], Aleeza Gerstein[3], Kathryn Hodgins[4]

[1]University of Calgary (Canada), [2]University of British Columbia (Canada), [3]University of Minnesota (United States), [4]Monash University (Australia)

Local adaptation is common in many species but we still know little about its genetic basis: how many genes are involved and how repeatable is evolution at the genomic level? Here, I discuss two methods we have developed to study the extent of convergence at the genome scale when different species adapt to similar environmental gradients. Our first method relies upon the detection of candidate loci in a first species to serve as a reduced set of test loci in a second species. We quantify whether signatures of adaptation in this test set are enriched relative to a null distribution for the background signatures in the genome. This approach alleviates two problems inherent in genotype-environment association scans: it reduces the rate of false negatives by reducing the number of tests and therefore the false discovery rate correction, and avoids correction for population structure, which can reduce the statistical signal for true positives. Our second method aims to interpret whether observed signatures of convergence are more repeatable than expected under a range of null models. This provides a unitless metric of the constraints driving repeated adaptation that can be readily compared among different species, traits, and study designs. I present an example of how these methods can be applied in a comparative study of local adaptation to climate in conifers. These methods will contribute to a growing body of analytical tools to study patterns of adaptation at the genome scale, and provide insights about what drives variation in these patterns.

O-04-EG11

Genome-wide RAD-seq revels adaptive divergence among seven stream stoneflies along a nationwide latitudinal gradient in Japan

Maribet Gamboa[1], Kozo Watanabe[1]

[1]Ehime University (Japan)

Genomic variation in locally adapted populations is influenced by complex dynamics of environmental conditions; understanding this relationship remains a challenge in evolutionary biology. We evaluated the degree of genome-environmental association of seven stonefly species across a wide geographic area in Japan as well as the genome-wide variation exhibited by local adaptation of co-existing multiple stonefly species. Double-digest restriction-associated DNA (ddRAD) libraries were independently sequenced for 219 individuals. A total of 1,181 candidate SNPs strongly associated with local adaptation were discovered using Latent mixed models, of which 58 SNPs showed strong correlation with environmental variables, specifically precipitation and altitude, using distance-based redundancy analysis. Adaptive genome regions of seven stonefly species revealed a possible parallel genomic evolution or co-evolution process based on high genetic similarity, while low intra-species variation revealed strong signatures of local adaptation. Our results revealed genomic signatures of environmental adaptation and their influence on multiple species. These data can potentially be applied to future research on river management and impacts of climatic change.

O-04-EG12

Molecular data support an early shift to an intermediate-light niche in the evolution of mammals

Yang Liu[1], Hai Chi[1], Longfei Li[1], Stephen J Rossiter[2], Shuyi Zhang[1]

[1]Shenyang Agricultural University (China), [2]Queen Mary, University of London (United Kingdom)

The visual ability and associated photic niche of early mammals is debated. The theory that ancestral mammals became nocturnal to avoid diurnal archosaurs (nocturnal bottleneck) is suggested from observed sensory, physiological and behavioural adaptations. Findings from studies of visual opsin gene sequences, and their photopigment products, have indicated that the middle or long wavelength sensitive (M/LWS) pigment in the ancestral mammal was either red- (~560nm) or green-sensitive (~530nm). However, the phylogenies from which these values were inferred are now known to be incorrect, and also did not contain sequences from prototherians. Moreover, others have argued that the combined spectral tuning of photopigments in ancestral mammals is more likely to represent an adaptation to twilight conditions at dusk and dawn (mesopic bottleneck). To address this controversy, here we analyze newly available sequence data to perform the most detailed study to date of middle/long-wave mammalian colour vision in ancestral mammals. By performing functional assays on resurrected proteins obtained from our ancestral reconstructions, we show that substitutions in the M/LWS pigment of amniotes led to a 9nm spectral shift towards shorter wavelengths in the Mammalia ancestor. This spectral shift, which is greater than predictions based on known critical sites, implies that the earliest mammals became yellow-sensitive (551nm). We hypothesize that these early changes arose as a consequence of an adaptive trade-off for exploiting both mesopic (twilight) and nocturnal (moonlight) niches.

O-04-EG13

Elucidating the genetic basis of inbreeding depression by contrasting the California Channel Island fox with Isle Royale gray wolf

Jacqueline Robinson[1, 2], Robert Wayne[2], Kirk Lohmueller[2]

[1]UCSF (United States), [2]UCLA (United States)

Isle Royale gray wolves have a small population size and suffer from severe inbreeding depression. However, other populations, such as California's Channel Island foxes, appear healthy and even persist for thousands of generations despite their small population size and isolation. The mechanistic explanation as to why one small population shows inbreeding depression and the other does not remains elusive. To address this question, we combine morphological assessment, analyses of complete genomes, and population genetic simulations to investigate these two systems. We find that although island foxes show no canonical signs of inbreeding depression, their genomes harbor more deleterious alleles and low levels of diversity due to long-term small population size. Additionally, our analyses show that Isle Royale wolf genomes contain large runs of homozygosity megabases in length, occasionally spanning entire chromosomes, interspersed with regions of high variation, consistent with recent and intense inbreeding since their founding less than a century ago. Simulations under different demographic models help us understand these contrasting systems. Isle Royale wolves recently descended from a historically large population of gray wolves. Simulations suggest that Individuals recently descended from historically large populations will carry more heterozygous deleterious variants than individuals descended from smaller populations. As such, recent inbreeding is predicted to create more homozygous recessive deleterious genotypes in the Isle Royale wolves than in the Island foxes. More generally, our findings argue that inbreeding depression is caused by strongly deleterious recessive alleles, having implications for the management of populations increasingly threatened by habitat loss and fragmentation.

O-04-EG14

Genomics of Bwindi mountain gorillas and conservation in eastern gorillas

Javier Prado-Martinez[1], Qasim Ayub[2, 3], Michael Andrew Quail[1], Steven Leonard[1], Yong Gu[1], Ruben Bautista[1], Irene Lobon[4], Sojung Han[4], Tomas Marques-Bonet[4], Michael Cranfield[5], Chris Tyler-Smith[1], Yali Xue[1]

[1]Wellcome Trust Sanger Institute (United Kingdom), [2]Monash University Malaysia (Malaysia), [3]Monash University Malaysia (Malaysia), [4]Universitat Pompeu Fabra/CSIC, ICREA (Spain), [5]University of California, Davis (United States)

Mountain gorillas (Gorilla beringei beringei) are one of the most endangered great apes, totalling less than 1,000 individuals, although their numbers have in recent decades recovered from less than 300 thanks to intensive conservation efforts. Previously, we studied the genomics of mountain gorillas from Virunga and here we add new whole genome sequences from six Bwindi mountain gorillas, the only other remaining population. We observe that the two populations have very similar demographic histories, but have accumulated subtle genetic differences. Both carry very low genetic diversity and high rates of inbreeding due to extremely low census population sizes. We are investigating the effects of this extreme inbreeding and exploring possible genetic adaptations of these populations to high altitude.

We have also sequenced an individual from Mount Tshiaberimu, where a very isolated population of less than 10 individual gorillas live. We show that, despite living at high altitude close to the mountain gorilla range, this population should be classified genetically as eastern lowland gorillas (Gorilla beringei graueri) and this information is being used by field researchers to reintroduce appropriate rescued gorillas to rescue the population from extinction.

Finally, we have generated 10X Genomics linked-read sequences for nine western lowland, eastern lowland and mountain gorillas to physically phase the variants and refine demographic inferences between and within the different species and subspecies. This provides a great genomic resource to study population history and structural variation in gorillas, which can contribute to the genomics and conservation of great apes.

O-04-EG15

Widespread adaptive lateral gene transfer in grasses

Luke Dunning[1]

[1]University of Sheffield (United Kingdom)

Descent with modification is the current paradigm for evolution in eukaryotes, but lateral gene transfers (LGT) among eukaryotes are increasingly reported. LGT represent a mechanism to bypass many of the limitations of evolution by descent, potentially allowing organisms to adapt beyond their inherent capabilities. However, little is known about the extent of LGT in eukaryotes, the mechanisms behind the process, and their adaptive significance. To investigate this further we generated a chromosomal level genome assembly for the grass *Alloteropsis semialata*, a species that has previously been shown to express key photosynthetic genes that were laterally acquired from distantly related grass species. Using the genome assembly and high-coverage genome data for a number of other grass species we identify 46 genes laterally acquired from at least six donors. These genes are clustered in 16 fragments distributed across the genomes, each smaller than 200kb in length. Using resequencing data for 24 other accessions of *Alloteropsis*, we establish the distribution of these LGT among populations, and show that LGT happened recurrently during the diversification of the group. While some of the LGT were later pseudogeneized, others remain functional and play a role in photosynthesis, adaptation to low nutrient soils and disease resistance. The acquired genes differ from the native copies in terms of expression and/or kinetics, so that their acquisition added novelty to the genetic apparatus of grasses. Our results demonstrate that LGT among eukaryotes can be significant, representing a recurrent source of material for adaptive diversification.

O-04-EG16

Harnessing natural variation to study the evolution of social behavior

Sarah Kocher[1]

[1]Princeton University (United States)

Natural variation can help us understand how ecological and evolutionary dynamics shape complex traits. Halictid bees harbor extraordinary variation in social behavior within and between species. This variation encompasses the full spectrum of sociality, from solitary to eusocial, making halictids a model system for studying the evolution of social behavior. Building on the wealth of natural history data for this group, we have developed genomic resources for 20 halictid species. This enables an integrative examination of the link between the proximate mechanisms underlying variation in social behavior and the ecological processes driving their evolution. From these studies, we have identified: (1) genetic factors associated with social behavior in a single, behaviorally polymorphic species, (2) a core set of genes that are rapidly evolving in social versus solitary halictids, and (3) several environmental factors that are strongly correlated with social behavior in this group of bees. This work integrates intraspecific studies with comparative molecular and ecological approaches to improve our understanding of the evolution social behavior across a range of ecological and evolutionary timescales.

O-04-EG17

Global natural variation of DNA methylation and its genetic architecture in Arabidopsis thaliana

Eriko Sasaki[1], Magnus Nordborg[1]

[1]Gregor Mendel Institute of Molecular Plant Biology (Austria)

Cytosine-methylation is an epigenetic mark that involved in diverse molecular mechanisms, such as silencing of transposable elements (TEs) and chromatin modifications. Although natural variation of DNA methylation is largely regulated by genetic factors, the genetic architecture shaping this variation remains unclear. To address this, we carried out genome-wide association studies for DNA methylation levels using a large population of *Arabidopsis thaliana* from the 1001 methylome project (Kawakatsu et al., 2016). We identified a large number of genetic variants that are significantly associated with CHH, CHG, and CG methylation levels (H = C, A or T), and also observed strong enrichment of a priori DNA methylation regulators in our candidate gene list, including *ARGONAUTE 1*, *9*, *NRPE1*, *CHROMOMETHYLASE2*(*CMT2*), *CHROMOMETHYLASE3* (*CMT3*), and *METHYLTRANSFERASE I* (*MET1*). Some of these alleles showed a characteristic geographical distribution, and DNA methylation levels of the targeted-sites were correlated with different climate variables. Together, our findings suggest that natural selection could have shaped global variation of DNA methylation levels in A. thaliana. In this presentation, I will discuss how genetic variation affects natural variation of molecular and organismal phenotypes through DNA methylation.

O-04-EG18

The genomics of behavioral adaptation to photoperiodism in an Asian burying beetle

Yuki Haba[1, 2], Hsiang-Yu Tsai[3], Sheng-Feng Shen[3], Dustin Rubenstein[2]

[1]Princeton University (United States), [2]Columbia University (United States), [3]Academia Sinica (Taiwan)

Scientists have long been fascinated by the stunning variation of innate behavior in animals, but little is known about the genetic architecture of behavioral variation in nature. Here we examined the genomic architecture underlying seasonal variation of burying behavior in Asian burying beetle, *Nicrophorus nepalensis*. We first showed the regional differentiation of burying behavior has fitness effect. We then showed that the behavioral differentiation among *N. nepalensis* populations in Taiwan and Japan was genetically-determined by a common-garden experiment. Furthermore, by genome-wide investigation we found that Nrxn1 was putatively responsible for the behavioral adaptation. In contrast, the circadian core genes, which have been suggested to be involved in photoperiodic behavior from lab experiments in insects and other taxa, did not show signatures of selection. Together, the results highlight the importance of studying the genetics of naturally occurring behavior to fully understand how innate behavior evolves in nature.

O-04-EA01

Evolution of adaptive immunity

Masanori Kasahara[1]

[1]Hokkaido University Faculty of Medicine (Japan)

I discuss the evolution of adaptive immunity focusing on two topics. First, the last decade has witnessed major progress in our understanding of the adaptive immune system (AIS) of jawless vertebrates. Surprisingly, unlike jawed vertebrates which use T-cell receptors and B-cell receptors, jawless vertebrates use variable lymphocyte receptors for antigen recognition. They also lack major histocompatibility complex molecules. However, emerging evidence also indicates that, despite these differences, the AISs of jawed and jawless vertebrates have much in common. On the whole, it appears that once established, the basic architecture of the AIS has not changed during vertebrate evolution. Second, many genes coding for key molecules of adaptive immunity are located on paralogous regions thought to have emerged as a result of whole-genome duplication (WGD) events. This observation suggests that the precursors of such genes formed a gene cluster before WGD. As genome information became available from many vertebrate and invertebrate species, we now have a better understanding of this gene cluster.

O-04-EA02

Generalists Versus Specialists: A New View Of How MHC Molecules Respond To Infectious Pathogens

Jim Kaufman[1, 2]

[1]University of Cambridge (United Kingdom), [2]University of Cambridge (United Kingdom)

The major histocompatibility complex (MHC) is a large and complex genetic region that, because of biomedical and theoretical importance, is one of the most intensively-studied regions of DNA on our planet. Of all known genes, the classical MHC genes have the highest level of allelic polymorphism as well as inter-allelic sequence diversity, due to their selection for resistance and susceptibility to a wide range of diseases (and in some cases, for mate choice and reproductive fitness). As such, the classical MHC genes in humans were the most important early examples of selection against which the impact of neutral evolution could be measured. Selection for combination of genes can also be important. This talk will use the work on the chicken MHC to discuss the role of gene co-evolution in the origin and selection of classical MHC alleles over at least 450 million years, highlighting the changes in biological function that occurred due to changes in genomic organization in the lineage leading to placental mammals. More recently, examination of chicken and human MHC class I molecules has led to the new concept of promiscuous generalists and fastidious specialists, in which relatively poorly-expressed molecules that bind a wide variety of peptides confer resistance to a range of common pathogens, while relatively well-expressed molecules that have stringent requirements and thus bind a narrow range of peptides can provide protection from new and/or virulent pathogens. This idea might also extend to other immune gene families, including MHC class II molecules.

O-04-EA03

Evolutionary trade-offs shape genomic diversity in the MHC

Tobias L Lenz[1]

[1]Max Planck Institute for Evolutionary Biology (Germany)

An optimal immune response requires a delicate balance of maximizing recognition of pathogens while minimizing damage to self tissue by the activated immune machinery. In vertebrates, this process is mediated by the major histocompatibility complex (MHC), whose highly polymorphic molecules present both self and non-self antigens for recognition by immune effector cells. Depending on which antigens are presented by an individual's MHC variants, this can trigger either pathogen resistance or autoimmunity. This trade-off is reflected at the genomic level; each individual genome carries only a limited subset of the MHC variants available in the population. Here I present the underlying conceptual framework and several layers of novel empirical evidence highlighting the different and partly antagonistic selective factors that contribute to the evolution of an optimal level of individual MHC diversity. This includes on the one hand quantitative and qualitative associations between MHC and HIV as well as evidence for pathogen-mediated selection in historical and modern human populations. On the other hand, I will show evidence for a positive association between individual MHC diversity and risk for autoimmunity as well as experimental results elucidating MHC-dependent T cell repertoire depletion. This mechanism evolved to prevent autoimmunity, but is assumed to impair overall T cell-based antigen-recognition. Combined, these antagonistic forces would select for an intermediate level of individual MHC diversity, as observed in natural populations. Understanding these trade-offs will shed light on the evolutionary and genetic basis of inter-individual differences in immunocompetence within and among populations.

O-04-EA04

Differences in peptide-binding affinities among alleles: a key to understand the complex patterns of natural selection on HLA genes

Da Di[1, 2], Thomas Goeury[1, 2], Jose Manuel Nunes[1, 2], Alicia Sanchez-Mazas[1, 2]

[1]University of Geneva (Switzerland), [2]University of Geneva (Switzerland)

The last 50 years of population genetics research are marked by impressive technological progress in high-throughput DNA sequencing and bioinformatics, leading to the exploration of complex molecular evolutionary models on both real and simulated biological data. In this context, many genes have been shown to be targets of specific selective pressures, one well known example being the positive selection of markers related to lactase persistence in milk-consuming human populations. The MHC genomic region is particularly fascinating in this respect, as its huge polymorphism is clearly the result of many different and sometimes antagonistic evolutionary forces.

Based on analyses of large sets of HLA-typed population data and peptide-binding predictions, we show that the heterogeneous distribution of HLA (class I) alleles worldwide can be explained by a combination of negative, positive and diversifying selective pressures in addition to joint asymmetric selection on different loci. The probability of observing several frequent alleles with similar peptide-binding affinities, either at the same locus or on identical haplotypes across two loci, in the same population is very low, which supports a model of purifying selection. In contrast, alleles with unique peptide-binding affinities are often common and sometimes very frequent, as expected from divergent allele advantage and selective sweep, respectively. While these mechanisms are evidenced in all populations, they do not seem to affect distinct HLA loci with the same intensity. Further evidence is revealed at the DNA sequence level, where different HLA loci show contrasted amounts of molecular variation across their coding and non-coding regions.

O-04-EA05

Host-parasite evolution and speciation in Neotropical cichlids

Marta Barluenga[1]

[1]Museo Nacional de Ciencias Naturales (Spain)

A thrilling topic in evolutionary biology is understanding the mechanisms generating biodiversity. Adaptive radiations are particularly good models for the study of speciation, since phenotypic divergence leading to speciation happens rapidly and repeatedly. The Neotropical Midas cichlid adaptive radiations in Nicaragua are driven by natural selection. Adaptation to alternative habitats and diets linked to morphological shifts has shaped species distribution, and has contributed to the development of reproductive barriers. However, whether this mechanism alone is sufficient to cause and maintain divergence in this system remains to be elucidated. Following on the strong evidence that ecological preference (habitat choice) is the main force shaping populations and driving reproductive isolation, we introduce a new environmental factor that might drive speciation in this system: parasite mediated selection. We evaluate the role of host-parasite interactions as a potential driving force for divergence and speciation. We have investigated associations between the community of parasites and variation at the Major Histocompatibility Complex (MHC) genes in several Midas cichlid radiations.

O-01-EN01

The PIWI-piRNA Pathway Targets Transposons in Hydra Somatic Stem Cells

Celina Juliano[1], Bryan Teefy[1], Stefan Siebert[1], Jack Cazet[1]

[1]University of California, Davis (United States)

Transposable elements (TEs) are hypothesized to be a driver of aging in animals; increased transposition in aging somatic cells increases genome destabilization. The PIWI-piRNA pathway, a small RNA pathway that represses TE mobilization, is expressed in non-aging tissues including the germline, some cancers, and the somatic cells of long-lived animals. Our previous work demonstrates that the PIWI-piRNA pathway is present in all somatic stem cells of Hydra, which has an estimated lifespan of at least a few hundred years. Thus, we hypothesize that PIWI-mediated TE repression in the somatic stem cells of Hydra is necessary for maintaining genomic stability thereby promoting longevity. However, TE repression by the PIWI-piRNA pathway in somatic cells outside of the gonad is not well understood. Our previous study found that somatic expression of the Hydra PIWI protein Hywi is necessary for animal survival. Here we provide evidence that TEs are targets of the PIWI-piRNA pathway in the somatic stem cells of Hydra, supporting our hypothesis that TE repression mediated by the PIWI-piRNA pathway is required for somatic longevity in Hydra.

O-01-EN02

Small RNA Pathways In Arthropods

Eric Miska[1, 2]

[1]University Of Cambridge (United Kingdom), [2]Wellcome Trust Sanger Institute (United Kingdom)

Some organisms are able to take up environmental RNA through ingestion, affecting their gene expression and phenotype. Nevertheless, the mechanisms and roles of such transmissible RNA are largely unknown. We have recently demonstrated horizontal RNA transfer among honeybees mediated by ingestion of worker and royal jelly. The presence of RNAs in the jellies indicates a transmission route, but also the RNAs ability to persist in the unsterile hive environment. Here, we show that MRJP3 is a secreted non-sequence specific RNA binding jelly protein. RNA mediates higher order assembly of the protein, leading to the formation of large ribonucleoprotein complexes that protect RNA from degradation and enhance ingested RNA uptake. Remarkably, specific viral RNA fragments represent high proportion among the natural partners of MRJP3. Our findings suggest that bees have evolved an environmental RNA-coating factor to concentrate, stabilize and share RNA among individuals.

O-01-EN03

microRNAs in the Same Clusters Evolve to Coordinately Regulate Functionally Related Genes

Yirong Wang[1, 2], Junjie Luo[1], Hong Zhang[1], Jian Lu[1]

[1]Peking University (China), [2]Peking University (China)

MicroRNAs (miRNAs) are endogenously expressed small noncoding RNAs. The genomic locations of animal miRNAs are significantly clustered in discrete loci. We found duplication and de novo formation were important mechanisms to create miRNA clusters and the clustered miRNAs tend to be evolutionarily conserved. We proposed a "functional co-adaptation" model to explain how clustering helps newly emerged miRNAs survive and develop functions. We presented evidence that abundance of miRNAs in the same clusters were highly correlated and those miRNAs exerted cooperative repressive effects on target genes in human tissues. By transfecting miRNAs into human and fly cells and extensively profiling the transcriptome alteration with deep-sequencing, we further demonstrated the functional co-adaptation between new and old miRNAs in the miR-17~92 clusters. Our population genomic analyses suggest that positive Darwinian selection might be the driving force underlying the formation and evolution of miRNA clustering. Our model provided novel insights into mechanisms and evolutionary significance of miRNA clustering.

O-01-EN04

Noncoding-RNA mediated epigenetics in the ciliate Oxytricha

Laura Landweber[1]

[1]Columbia University (United States)

The ciliate *Oxytricha trifallax* possesses a dynamic pair of genomes, and massive DNA rearrangements produce a highly fragmented but functional somatic macronucleus from a complex germline micronucleus. This process eliminates nearly all noncoding DNA, including transposons, and rearranges over 225,000 short DNA segments to produce a second genome containing thousands of gene-sized "nanochromosomes." In the precursor, germline genome, the shattered segments of different genes often interweave with each other, frequently overlap and sometimes combinatorially assemble (Chen et al. 2014). The mature, somatic genome contains over 16,000 nanochromosomes (Swart et al. 2013). Noncoding RNAs program the entire process of genome rearrangement. Millions of 27nt piRNAs provide the critical information to mark and protect the retained DNA segments of the genome (Fang et al. 2012) and a distinct set of piRNAs mark a subset of deleted regions to assist with their elimination. Maternally-inherited, long, non-coding (lnc) RNAs provide three additional layers of continuity across generations, including serving as templates for genome remodeling and RNA-guided DNA repair (Nowacki et al., 2008) while also regulating gene dosage and chromosome copy number (Nowacki et al., 2010 PNAS). This illustrates the ability of noncoding RNAs to transmit heritable changes to the next generation. Together, *Oxytricha*'s elaborate epigenome, assembled through complex interacting networks of both long and small non-coding RNAs, encapsulates an RNA-driven world, packaged in a modern cell. I will present progress on understanding how this RNA-guided system of epigenetic inheritance evolved.

O-01-EN05

First evidence for HEN1 mediated methylation of miRNAs in animals, and the functional importance of small RNA stabilization in cnidarians

Arie Fridrich[1], Vengamanaidu Modepalli[1], Yehu Moran[1]

[1]The Hebrew University of Jerusalem (Israel)

Small non-coding RNAs (sRNAs) such as microRNAs (miRNAs), small interfering RNAs (siRNAs) and Piwi interacting RNAs (piRNAs) regulate the levels of endogenous, viral and transposable element's RNA levels in plants (excluding piRNAs) and animals. These pathways are explored mainly in bilaterian animals, such as vertebrates, arthopods and nematodes, where siRNAs and piRNAs, but not miRNAs bind their targets with a perfect match, and mediate the cleavage of the target RNA. Methylation of the 3' ends of piRNAs and siRNAs by the methyl-transferase HEN1 protects these sRNAs from degradation. There is a noticeable selection in bilaterian animals against miRNA-mRNA perfect matching, as it leads to the degradation of miRNAs. Cnidarians (Sea anemones, corals, hydroids and jellyfish), are separated from bilaterians by more than 600 million years. As opposed to bilaterians, cnidarian miRNAs frequently bind their targets with a nearly perfect match. Knowing that an ortholog of HEN1 is widely expressed in the sea anemone *Nematostella vectensis*, we tested in this work whether it mediates the stabilization of its sRNAs. We show that the knockdown of HEN1 in *Nematostella* results in a developmental arrest. Small RNA sequencing revealed that the levels of both miRNAs and piRNAs drop dramatically in the morphant animals. Thus, we provide the first evidence of a methylation mechanism that stabilizes miRNAs in animals, and highlight the importance of post-transcriptional regulation in non-bilaterian animals.

O-01-EN06

Prevalent selection against microRNA target sites in human populations

Andrea Hatlen[1], Antonio Marco[1]

[1]University of Essex (United Kingdom)

MicroRNAs are powerful gene regulators that play an important role in the evolution of regulatory networks. At the population level, the study of polymorphisms allows the identification of microRNA target sites under positive or purifying selection. Mutations generate novel microRNA target sites, some of which may affect the expression of genes. Thus, some transcripts will 'avoid' target sites for specific microRNAs. In a previous work we have shown that, in *Drosophila*, maternal genes are under selection against maternal microRNA target sites. Here we present our results in human populations, in which we investigated the allele frequencies at untranslated regions. We have found that selection against microRNA target sites is frequent in human populations. Interestingly, the selective pressure for specific microRNAs is different across human populations. This effect is stronger in the X chromosome, suggesting a hemizygotic effect of otherwise recessive alleles. Specific analyses on oncogenic microRNAs indicate that their targets are strongly affected by this type of selection. We also show how this pattern may be used to identify oncogenes and tumour-suppressor genes based on the allele frequency distribution at potential microRNA target sites. In conclusion, selection against microRNA target sites is prevalent in human populations. We hypothesise that similar evolutionary patterns may also affect other regulatory sequences such as RNA-binding protein motifs and transcription factor binding sites.

O-04-EE01

The evolution of gene body methylation in plant species.

Takuno Shohei[1]

[1]SOKENDAI (Japan)

DNA methylation is an important epigenetic modification that affects both chromatin packing and transcription. DNA methylation occurs in three sequence contexts, that is, CG, CHG, and CHH (where H is A, C, or T) in plants. All three contexts are methylated within repetitive elements. The major role of DNA methylation within repetitive elements is to silence transcription and functions as a host defense against transposable elements. On the other hand, only the CG context is predominantly methylated within active coding regions in *Arabidopsis thaliana* that is called gene body methylation. Gene body methylation is considered a byproduct of the process of removing heterochromatic marks within active genes; gene body methylation might not be functional. However, I found that gene body methylation was observed in a biased subset of genes, tended to be conserved between plant orthologs, and the drastic changes of gene body methylation significantly affected expression levels. I will discuss the evolutionary consequences of gene body methylation based on these observations.

O-04-EE02

Epigenetic variation and regulation of imprinted gene expression

Mary Gehring[1]

[1]Whitehead Institute for Biomedical Research (United States)

DNA methylation is a conserved epigenetic modification found in plants, animals, and fungi. In both plants and animals, differential methylation of maternal and paternal alleles is associated with parent-of-origin specific, or imprinted, gene expression. In plants, imprinting occurs in the endosperm, an essential seed tissue formed from one of two fertilization events that characterize flowering plant reproduction. The endosperm is triploid, with two maternally and one paternally inherited genomes. Most genes are expressed in a maternal: paternal ratio of 2:1, but 100+ genes are imprinted, meaning they are expressed preferentially either from the maternally or the paternally inherited alleles. This talk will focus on our latest understanding of the role of DNA methylation dynamics, small RNAs, transposable elements, and various epigenetic pathways in establishing and maintaining imprinted expression, and will discuss insights derived from single cell and comparative evolutionary studies of imprinting.

O-04-EE03

Reconstructing Denisovan Anatomy Using DNA Methylation Maps

David Gokhman[1], Nadav Mishol[1], Marc de Manuel Montero[2], David de Juan[2], Jonathan Shuqrun[1], Tomas Marques-Bonet[2, 3], Yoel Rak[4], Liran Carmel[1]

[1]The Hebrew University of Jerusalem (Israel), [2]UPF-CSIC (Spain), [3]ICREA (Spain), [4]Tel Aviv University (Israel)

The Denisovan is a human group unique for having its DNA sequence and methylation mapped, but whose morphology remains almost completely unknown. Here, we present a method to reconstruct anatomical profiles from DNA methylation patterns, based on linking unidirectional promoter methylation changes to loss-of-function phenotypes. We tested the performance of our method by assembling anatomical profiles of the Neanderthal and the chimpanzee, and comparing them to their known morphology. We demonstrate that this method reaches ~87% precision in identifying divergent traits, and ~89% in predicting their direction of change. We then reconstruct a putative anatomical profile of the Denisovan, offering 56 traits in which the Denisovan is expected to be different from modern humans or Neanderthals. We suggest that Denisovans likely shared many traits with Neanderthals, including a projecting face, robust jaws, low forehead and wide pelvis. We also identify additional changes along the Denisovan lineage, such as increased length of the dental arch, and expanded biparietal width. We find that the vast majority of morphologies identified in the late Pleistocene Xuchang crania from China, which were hypothesized to belong to Denisovans, are included in our reconstruction of the Denisovan anatomical profile, providing first genetic support to the classification of these individuals as Denisovans. We conclude that DNA methylation maps provide means to predict morphology, and can be used to uncover anatomical features that do not survive in the paleontological record.

O-04-EE04

Molecular evolution of a recombination suppressed avian autosome linked to alternative reproductive phenotype

Dan Sun[1], Iksoo Huh[1], Paramita Chatterjee[1], Wendy M. Zinzow-Kramer[2], Donna L. Maney[2], Soojin V. Yi[1]

[1]Georgia Institute of Technology (United States), [2]Emory University (United States)

A large rearrangement on the second chromosome of the white-throated sparrow (Zonotrichia albicollis) determines two color morphs that differ dramatically in social behaviors. Birds of the white-striped (or white; ZAL2/2^{m}) morph exhibits more territorial aggression and less parental behaviors, relative to the tan-striped (or tan; ZAL2/ZAL2) counterparts. Although white birds almost exclusively mate with tan birds, occasional same-morph matings are known to occur. We compared the genome of a super-white individual (homozygous for ZAL2^{m}) produced from a rare mating between white birds with the genome of a tan individual and identified a ~1.2% of nucleotide difference between the two chromosomes and limited positive selection. Consistent with suppressed recombination, ZAL2^{m} displays a significant excess of substitutions on functional protein-coding sites relative to ZAL2, although the degeneration is minimal evidenced by few signs of pseudogenization. On the other hand, brain transcriptome analysis reveals ~40% of ZAL2/2^{m}-linked genes that are differentially expressed between the two chromosomes. Interestingly, a large proportion of these genes do not show morph-biased expression, signifying an operating compensatory mechanism. Indeed, ZAL2 alleles in the white birds tend to adjust their expression to compensate for disrupted expression of ZAL2^{m} alleles. These findings indicate rapid regulatory evolution in the early stage of chromosomal differentiation. In addition, new data on whole-genome DNA methylation of the same brain regions reveal intriguing chromosome-wide differences between ZAL2 and ZAL2^{m}. The chromosomal polymorphism in the white-throated sparrow promises to be a powerful system to investigate molecular mechanisms underlying phenotypic divergence.

O-04-EE05

CTCF in early vertebrate evolution: lessons from jawless and cartilaginous fishes about its phylogeny and establishment of epigenomic functions

Mitsutaka Kadota[1], Yuichiro Hara[1], Kazuaki Yamaguchi[1], Osamu Nishimura[1], Shigehiro Kuraku[1]

[1]RIKEN (Japan)

The nuclear protein CCCTC-binding factor (CTCF) contributes as an insulator to chromatin organization in the genomes of diverse animals. Despite its pivotal role in epigenomic regulation, our knowledge of its binding property is confined mainly to mammals, and the molecular phylogeny of CTCF and its relatives has not been thoroughly investigated. Our laboratory has conducted whole genome sequencing and chromosome-scale scaffolding with Hi-C data, followed by superimposition of transcriptomic and epigenomic data onto produced *de novo* genome assemblies for multiple jawless and cartilaginous fishes. These novel resources allowed us to identify CTCF homologs in those non-osteichthyan fishes and facilitated genome-wide detection of binding sites of the CTCF protein in the Arctic lamprey and elasmobranch sharks. Our ChIP-seq analysis suggested that the mode of CTCF binding known from studies on mammals was established in the last common ancestor of extant vertebrates (more than 500 million years ago). Moreover, the long-standing chromatin regulator CTCF was shown to have undergone a complex evolutionary history in non-osteichthyan lineages, giving rise to CTCFL (BORIS) and CTCF2 whose duplication timings have been investigated in depth with the latest, enriched sequence data set in our study.

Reference: Kadota et al., 2017. CTCF binding landscape in jawless fish with reference to Hox cluster evolution. *Scientific Reports* 2017. 7, 4957.

O-04-EE06

Robustness of Transposable Element regulation but no genomic shock observed in an interspecific Arabidopsis hybrid

Juliette de Meaux[1], Ulrike Goebel[1]

[1]University of Cologne (Germany)

The merging of two divergent genomes in a hybrid is believed to trigger a genomic shock, disrupting gene and transposable element (TE) silencing. Here, we tested this expectation by comparing the pattern of expression of transposable elements in their native and hybrid genomic context. For this, we sequenced the transcriptome of the Arabidopsis thaliana genotype Col-0, the A. lyrata genotype MN47 and their F1 hybrid. Contrary to expectations, we observe that the level of TE expression in the hybrid is strongly correlated to levels measured in the parental species. We detect that at most 1.1% of expressed transposable elements belonging to three specific subfamilies changing their expression level upon hybridization. Most of these changes, however, are of small magnitude. We observe that the few hybrid-specific modifications in TE expression are more likely to occur when TE insertions are close to genes. In addition, changes in epigenetic histone marks H3K9me2 and H3K27me3 following hybridization do not coincide with TEs with changed expression. Finally, we further examined TE expression in parents and hybrids exposed to severe dehydration stress. Despite the major reorganization of gene and TE expression by stress, we observe that hybridization does not lead to increased disorganization of TE expression in the hybrid. We conclude that TE expression is globally robust to hybridization and that the term genomic shock is no longer appropriate to describe the anticipated consequences of merging divergent genomes in a hybrid.

O-01-EG01

Metabolites and lipids of the human brain: evolution and function.

Philipp Khaitovich[1, 2]

[1]Skolkovo Institute of Science and Technology (Russian Federation), [2]Shanghai Institutes for Biological Sciences (China)

Small molecules with molecular weight lower than 1500Da are commonly referred as metabolites and their hydrophobic fraction, as lipids. In the human brain, metabolites and lipids represent main building blocks, signaling molecules and energy sources. Still currently little is known about roles of these molecules in the human brain development and evolution.

We investigate changes in the concentration of thousands of metabolites and lipids composing the human brain tissue using high precision untargeted mass spectrometry techniques. In this presentation, I will address the following questions: (i) are there metabolic or lipid composition features unique to humans? (ii) does lipid composition of the human brain change in development and aging?

O-01-EG02

Evolution of human-specific gene co-expression networks

Genevieve Konopka[1]

[1]UT Southwestern Medical Center (United States)

It has been hypothesized that one of the consequences of the highly evolved cognitive capacity of the human brain is the development of increased vulnerability to cognitive disorders. Recent technical breakthroughs have allowed for the identification of genetic and molecular signatures in the central nervous system that distinguish humans from non-human primates. We have identified novel human-specific patterns of gene expression and regulation in the neocortex. These data suggest that the human brain has undergone rapid modifications of gene expression patterns to support our enhanced cognitive abilities. In addition, we identified an enrichment of cognitive disease related genes that demonstrate unique gene expression changes in the human brain. We have carried out functional follow up of a number of these genes with human-specific coexpression patterns. We have focused on genes that regulate transcription and/or mRNA splicing. These follow up studies have manipulated these genes in primary human neurons and rodent models followed by further genome-wide expression analyses using RNA-sequencing. These new data have uncovered additional coexpression patterns and molecular pathways that might be involved in human disorders of cognition

O-01-EG03

The evolutionary trajectory of spatial transcriptome and epigenome in primate brains

Chuan Xu[1, 2], Qian Li[1, 2], Olga Efimova[3], Liu He[1], Shoji Tatsumoto[4], Philipp Khaitovich[1, 3], Yasuhiro Go[4, 5]

[1]CAS-MPG Partner Institute for Computational Biology (China), [2]University of Chinese Academy of Sciences (China), [3]Skolkovo Institute of Science and Technology (Russian Federation), [4]National Institutes of Natural Sciences (Japan), [5]National Institute for Physiological Sciences (Japan)

Despite the substantial efforts made to compile the molecular maps of the human brain, they do not inform us of the features unique to humans. Yet, identification of these features might help us to understand both the evolution and nature of the human cognitive and physiological traits. Here, we approached this question by analyzing gene expression and H3K27ac chromatin modification data collected in eight brain regions of humans, chimpanzees, gorillas, gibbon and macaques. Integrated analysis of these data revealed the existence of 1,851 human-specific transcriptome differences affecting one or multiple brain regions, in contrast to 240 chimpanzee-specific ones. More than a half of these differences represented elevated expression in neurons and astrocytes of human hippocampus, while the rest affected microglial functions in frontal cortex and cerebellum. Analysis of predicted regulatory interactions driving these differences revealed the role of transcription factors in species-specific transcriptome changes, while epigenetic modifications were linked to spatial expression differences conserved across species.

O-01-EG04

Functional studies in 2D and 3D stem cell systems of candidate genes underlying human-specific features of cerebral cortex development

Alessio Strano[1], Tomoki Otani[1], Jocelyn Tang[1], Irene Gallego Romero[2], Rick Livesey[1]

[1]University of Cambridge (United Kingdom), [2]University of Melbourne (Australia)

Sequencing of primate genomes and comparative studies of primate cerebral cortex have generated numerous hypotheses on the genetic bases for evolutionary differences in cortical development; however, the contribution of selected genes has seldom been tested functionally. We have previously shown that cortical cultures and organoids derived from pluripotent stem cells can recapitulate *in vitro* key features of human and non-human primate cortical development, including progenitor proliferation dynamics and developmental timing. Therefore, we decided to use these *in vitro* models to test how cortical development might be affected by evolutionary differences in gene expression patterns. Genes that were differentially expressed between humans and crab-eating macaque *in vitro* significantly overlapped with genes previously shown to have different levels of chromatin modifications associated with active transcription *in vivo*, indicating at least partial recapitulation of species-specific gene expression patterns. Among the genes expressed at different levels in humans and macaque, we found several known to cause microcephaly or megalencephaly when mutated in humans. For a few of these (importantly, not all), modification of their expression levels in macaque cortical cultures to resemble human-like expression resulted in more human-like progenitor proliferative behaviour, including increased progenitor population expansion and decreased early neuronal differentiation.

Our results highlight the importance of regulation of gene expression as a mechanism for evolutionary change in cortical development. Moreover, we demonstrate how *in vitro* differentiation models can aid testing of hypotheses regarding gene function, a critical step towards elucidating the molecular mechanisms underlying the evolution of the human cerebral cortex.

O-01-EG05

Big Brains: What High-Throughput Enhancer Knockouts Reveal about Human Cortical Evolution

Evan Geller[1, 2], James Noonan[1, 2, 3]

[1]Yale University (United States), [2]Yale University (United States), [3]Yale University (United States)

The majority of genetic variation between humans and other primates resides within regions of gene regulation. It is hypothesized that genetic changes occurring within gene regulatory elements modified conserved developmental processes and contributed to human-specific biological phenotypes, such as the expansion and elaboration of the cerebral cortex. Multiple studies over the last decade have identified human-specific genetic changes, such as Human Accelerated Regions (HARs), that have been linked to changes in developmental gene regulation. However, the biological functions of these loci remain almost entirely unknown. To address this question, we used a high-throughput CRISPR-Cas9 knockout strategy in human neural stem cells to disrupt >55,000 potential transcription factor binding sites in >2,300 enhancers active in human cortical development. We targeted two classes of enhancers relevant to human evolution: human-gain enhancers showing increased epigenetic activity during corticogenesis and HARs that encoded enhancers active in the developing human cortex. Our assay quantified the effect of enhancer knockout on a critical phenotype during corticogenesis-- neural stem cell proliferation. A quantitative proliferation phenotype was obtained by measuring the abundance of each enhancer knockout at initial, intermediate, and final experimental time points by high-throughput sequencing. We found evidence of enhancer knockouts, including human-gain enhancers and HARs, with very strong effects on neural stem cell proliferation. Our genome-scale survey reveals the quantitative landscape of gene regulatory enhancer control of proliferation and utilizes cutting-edge machine learning methods to interpret the biological impact of distinct enhancer classes on human cortical development and evolution.

O-01-EG06

Single-cell transcriptional signatures of the aging nonhuman primate brain

Kenneth L Chiou[1], Alex R DeCasien[2, 3], Michael J Montague[4], Chet C Sherwood[5], Michael L Platt[4, 6, 7], Noah Snyder-Mackler[1, 8]

[1]University of Washington (United States), [2]New York University (United States), [3]New York Consortium in Evolutionary Primatology (United States), [4]University of Pennsylvania (United States), [5]The George Washington University (United States), [6]University of Pennsylvania (United States), [7]University of Pennsylvania (United States), [8]University of Washington (United States)

The human brain is capable of rapidly and reliably processing complex stimuli. These abilities decline with age, sometimes manifesting through neurodegenerative diseases such as Alzheimer's. Yet little is understood about how aging influences the distribution and function of individual neurons, data critical for understanding the heterogeneity of the aging process across cells and individuals. Here, we characterize gene expression in single brain cells from rhesus macaques, a model primate organism closely related to humans that shows similar age-related declines in sensory, motor, and cognitive function. We sampled cells from the prefrontal cortex due to its role in higher cognitive functions and known changes during healthy or neurodegenerative aging. We characterized gene expression using sci-RNA-seq, which combines flow cytometry with combinatorial indexing to profile gene expression at single-cell resolution. First, we used sci-RNA-seq to recover transcriptional signatures from 1,473 macaque brain cells sequenced to an average depth of ~10,000 reads per cell. We then used unsupervised clustering to classify cell types based on transcriptional patterns. Cells formed four distinct clusters sufficient for distinguishing neurons and several glial types. We then expanded our sci-RNA-seq experiment to characterize gene expression in 30 free-ranging macaques of varying ages. These data allowed us to identify changes in cell-type composition and single-cell gene expression associated with aging. Future studies will determine the impact of social and environmental factors on changes in gene expression related to aging and will serve as a resource for the development of medical interventions for age-associated diseases in humans.

O-02-50Y01

The evolutionary significance of polyploidy

Yves Van de Peer[1]

[1]Ghent University - VIB (Belgium)

Thousands of species are currently polyploid and contain multiple copies of their genome. On the other hand, the long-term establishment of organisms that have undergone ancient whole genome duplications (WGDs) has been exceedingly rare. The apparent paucity of ancient genome duplications and the existence of many species that are currently polyploid provides a fascinating paradox. Interestingly, many ancient WGDs seem to have been established at very specific times in evolution, for instance during major ecological upheavals and periods of extinction. Our work has shown that WGDs observed for many different plant lineages seem to have coincided with the most recent major mass extinction, i.e. the K/Pg extinction, 66 million years ago. I will put forward different hypotheses of why polyploids, compared to their diploid progenitors, might have had some selective advantage that might explain their survival at times of extinction. Also, I will discuss how WGD events might lead to an increase in biological complexity. WGDs copy entire pathways or networks, and as such create the unique situation in which such duplicated pathways or networks could evolve novel functionality through the coordinated sub- or neofunctionalization of its constituent genes.

O-02-50Y02

The evolution of meiosis in autotetraploid Arabidopsis arenosa

Kirsten Bomblies[1], Clare Henry[1]

[1]John Innes Centre (United Kingdom)

Meiosis is essential for fertility of sexual eukaryotes and its core structures and progression are conserved across kingdoms. However, this system has been fine-tuned by evolution over the eons to separate pairs of homologs. Polyploids, however, have more than two copies of every chromosomes. In neopolyploids this presents an enormous challenge, with multiple chromosomes often recombining and forming so-called multivalents that can cause chromosome segregation problems. But polyploids can evolve to solve these problems, but how meiotic stability evolves in polyploids has remained largely mysterious. In a genome scan for adaptation to whole genome duplication in Arabidopsis arenosa, we previously found evidence that eight interacting meiotic proteins were under strong selection in the polyploid lineage. The proteins encoded by these genes are critical for axis formation, recombination and synapsis, in other words, some of the most central structural processes in meiosis. We hypothesize that modifications of these proteins stabilize polyploid meiosis by directly altering crossover number and/or the strength of crossover interference, with the outcome that there are fewer multivalent associations among the available chromosome copies. Our work provides insights not only into polyploid stabilization, but also more generally how modified recombination rates can evolve and what pleiotropic effects this might have.

O-02-50Y03

Genomic hotspots of adaptation to whole genome duplication

Levi Yant[1]

[1]The John Innes Centre (United Kingdom)

Whole genome duplication (WGD) occurs in all eukaryotic kingdoms and is implicated in organismal complexity, adaptation and speciation. Despite the evolutionary potential of WGD, a sudden duplication of all chromosomes poses challenges to key processes, especially the reliable segregation of chromosomes at meiosis. Nonetheless, nature reveals solutions: the many polyploid species with diploid-like meiosis show that difficulties can be overcome. The molecular basis of this is still a largely a mystery. Our work in autotetraploid Arabidopsis arenosa discovered obvious WGD-associated selective sweeps on meiosis genes with roles in crossover regulation(1). Natural variation in at least one of these genes has a dramatic effect on meiotic chromosome pairing. Here we assess whether species that independently adapted to the challenges attending WGD evolved similar solutions, whether crossover regulation is a common target of WGD-associated adaptation, and whether standing variation in diploid populations contributes to adaptation to WGD. We assay for repeatability and convergence in adaptation to WGD in A. arenosa, Arabidopsis lyrata, Cardamine amara, Cochlearia officinalis, Mimulus guttatus, Arabis pumila, and the frog Neobatrachus, all of which harbor extant intraspecific ploidy variation. We do this by applying genome scanning and cytological approaches, as well as broader population genomics. We seek to provide a kingdom-spanning view of repeatability and constraint in the context of intense selection on a crucial, conserved process. This will additionally provide insight into how organisms adapt to the altered intracellular environment following WGD, an important ongoing force in evolution.

1. Yant et al. Current Biology (2013) 23, 2151

O-02-50Y04

Genome restructuring during early vertebrate evolution

Yoichiro Nakatani[1], Aoife McLysaght[1]

[1]Trinity College Dublin, University of Dublin (Ireland)

Whole-genome duplications (WGD) at the origin of vertebrate has been considered as key evolutionary events providing raw genetic materials for new gene functions, leading to innovations in developmental mechanisms, and facilitating evolutionary success and diversification of vertebrate lineages. The impacts of WGD persist over a long evolutionary time, even after 500 million years since the WGD events: for example, ohnologs (i.e. WGD paralogs) shape the landscape of copy number variations in the present-day human genomes, and duplications/deletions of ohnologs are frequently associated with genetic diseases including Down syndrome. In this way, influences of the ancient vertebrate WGDs on modern genomes have been studied extensively. On the other hand, the evolution of ohnologs soon after the vertebrate WGD events, especially at an early stage of rediploidization/fractionation process, has remained elusive.

In order to study the evolution of genome structure and ohnologs at the onset of rediploidization/fractionation after the WGD events in vertebrate, we reconstructed ancestral genome structures before and after the two WGD events, using a recently developed probabilistic macrosynteny model. Our reconstruction showed large-scale genome restructuring during early vertebrate evolution, and confirmed two hypotheses proposed by Ohno regarding WGDs at the origin of vertebrate and conservation of microchromosomes. We also discuss remaining problems regarding early vertebrate genome evolution, rediploidization after WGDs, and a possible link between large-scale genome restructuring and neo-/sub-functionalization of ohnologs at the origin of vertebrate and jawed-vertebrate.

O-02-50Y05

Genome stabilization mechanisms in early post-polyploidization evolution

Sira Sriswasdi[1, 2], Wataru Iwasaki[1, 3, 4]

[1]The University Of Tokyo (Japan), [2]Chulalongkorn University (Thailand), [3]The University Of Tokyo (Japan), [4]The University Of Tokyo (Japan)

Many studies have highlighted the multi-stage nature of polyploidization and shown that differing evolutionary forces became dominant depending on the time scale. However, few studies were able to investigate the early stages of polyploidization that are best characterized by abrupt and catastrophic disruptions of genome stability and gene expression equilibrium. Our recent discovery of two recently and independently hybridized Basidiomycota fungi of genus *Trichosporon*, *T. coremiiforme* and *T. ovoides*, offered an opportunity for comparatively interrogating early evolutionary mechanisms that resolved detrimental effects of polyploidization. We previously identified widespread evolutionary rate reduction, which would limit functional divergence between homeologs, and frequent losses of genes involved in transcription and translation, which would dampen the impact of increased gene dosage on global expression level, as signs that genome re-stabilization was a major evolutionary consideration in young polyploids (Sriswasdi *et al*. Genome Research 2016). In the current analyses, we further investigated the inter-play between evolutionary forces operating on the genomic sequence, 3-dimensional chromosome organization, and transcriptional activity in these hybrids and their non-polyploid relatives. For the most part, *T. coremiiforme* and *T. ovoides* followed analogous evolutionary trajectories that clearly prioritize reconciliation and re-stabilization of the involved genomes. Nonetheless, there were also markedly differences between the two hybrids possibly driven by the difference in divergence levels between their parental species (7% and 16% amino acid sequence divergence among *T. coremiiforme*'s and *T. ovoides*'s subgenomes, respectively). Our efforts would shed more lights on the complex evolutionary processes that shaped the critical, early stages of polyploidization.

O-02-50Y06

Adaptation to the whole genome duplications in Australian burrowing frogs Neobatrachus

Polina Yu. Novikova[1, 2], Stephen C. Donnellan[3], Yves Van de Peer[1, 2], Levi Yant[4]

[1]VIB (Belgium), [2]Ghent University (Belgium), [3]University of Adelaide (Australia), [4]John Innes Centre (United Kingdom)

Polyploidy plays an important role in evolution, providing a backup genetic material and increasing genetic novelty. However, polyploids have to adapt their cellular machinery to the instantaneous doubling of chromosomes, in particular the segregation of chromosomes during meiosis, since random crossovers between each copy may compromise regular chromosomal segregation. The majority of natural stable sexual polyploids are plants, where genes regulating crossover interference, chromosomal pairing and segregation have been shown to have undergone selection. Polyploidy in animals is rare and most polyploid animals reproduce asexually. The *Amphibia* is the only exception among vertebrates with multiple independent sexualy reproducing polyploid lineages, as in the Australian frog genus *Neobatrachus*. Our preliminary results comparing nucleotide diversity and differentiation between the diploid *N. pictus*and the tetraploid *N. sudellae* show that potentially selected genes in the tetraploid are enriched for microtubule motor activity function. This suggests modifications of the homologous pairing process during meiosis. To achieve further clarity on the mechanism of adaptation to WGD in *Neobatrachus*, we work on the genome assembly of *N. pictus*. This reference genome will be only a third assembled genome for the Order *Anura* and the first one for the Superfamily *Hyloidea*. By extending low-coverage genome data for the rest of the genus we will reconstruct its evolutionary history, and identify potentially selected regions and adaptive changes in tetraploids compared to diploids. This will provide the first description of adaptation mechanism(s) to autotetraploidy in animals.

O-02-50Y07

Comparative Analysis Between Allopolyploid *Arabidopsis kamchatica* And Its Diploid Progenitors Reveals Effects Of Polyploidy On Genetic Diversity And Selection

Gwyneth Halstead-Nussloch[1], Timothy Paape[1], Kentaro Shimizu[1]

[1]University of Zurich (Switzerland)

Whole genome duplication (WGD) has been shown to be a common mechanism of plant evolution, speciation and diversification. Comparison studies between diploid parents and their derived subgenomes in polyploids are essential for understanding the genomic consequences of WGD. Here, we provide empirical evidence for theoretical long-term effects of polyploidy on the evolution of genomes. *Arabidopsis kamchatica*is a self-compatible, natural allotetraploid derived from the hybridization of self-incompatible *A. halleri* subsp. *gemmifera* (found in East Asia) and *A. lyrata* subsp. *petraea* (from Far East Russia). Genome wide diversity of the two subgenomes was less than half of either diploid parent, which is an expected consequence of a transition from self-incompatibility to self-compatibility. Demographic modeling and population structure showed multiple hybrid origins. About 50% of SNPs in the subgenomes were shared with their corresponding parents. Weaker purifying selection and thereby an increase in relaxed constraint, is a predicted result of whole genome duplication due to masking of deleterious mutations. The distribution of fitness effects showed the subgenomes of A. kamchatica contained higher proportions of neutral mutations and lower proportions of severely deleterious mutations than in the respective diploid parental genomes. Using genome scans, we found evidence of conserved patterns of selection between the polyploid and diploid parents suggesting some adaptive mutations have been inherited from the diploid parents. Characterizing the conserved and divergent genomic patterns will help identify novel adaptations that have arisen after polyploidization and describe the impact of redundant gene copies on the evolution of plant genomes.

O-01-GF01

Methods to characterize geographic structure in genetic variation

John Novembre[1, 3], Hussein Al-asadi[2], Ben Peter[4]

[1]University of Chicago (United States), [2]University of Chicago (United States), [3]University of Chicago (United States), [4]Max Planck Institute for Evolutionary Anthropology (Germany)

Population structure is a fundamental feature of genetic variation that has importance for addressing questions in evolutionary biology, conservation genetics, and trait mapping. However, most analytical methods to represent population structure do not incorporate geography directly, and it must be considered post hoc alongside a visual summary. In this talk, I will present results on human population structure at differing spatial scales and the computational tools that are making new insights possible. In particular, I will show results from a spatially explicit method that estimates "effective migration" surfaces to visualize how human genetic diversity is geographically structured (the EEMS method). The resulting surfaces are "rugged", which indicates the relationship between genetic and geographic distance is heterogeneous and subtly distorted as a rule. I will also show a complementary technique for making clear the underlying allele frequency patterns that are common among large collections of human single nucleotide variants (SNVs). These results provide visualizations of human genetic diversity that reveal local patterns of differentiation in detail. The fine-scale population structure depicted here is relevant for understanding complex processes of human population history and the tools have application to a diverse range of species.

O-01-GF02

Efficient representations of local trees in Coalescent Hidden Markov models for demographic inference

Matthias Steinruecken[1], Alexey Miroshnikov[2], Gautam Upadhya[3]

[1]University of Chicago (United States), [2]University of California, Los Angeles (United States), [3]University of Chicago (United States)

Inference of historical population sizes from contemporary genomic sequence data has received a lot of attention in recent years. A particular focus has been on recent exponential growth in humans. This recent growth strongly impacted the distribution of rare genetic variants, which are of importance when studying disease related genetic variation. The popular method PSMC (Li and Durbin, 2011) is used to infer population sizes from a sample of two chromosomes. However, the small sample size severely limits the power of this method in the recent past.

To improve inference in the recent past, we extend the Coalescent Hidden Markov model approach of PSMC to larger sample sizes. Since using the full genealogical trees relating the sample at each locus is computationally prohibitive, we introduce a flexible framework to employ different representations of these local trees. We implement this framework using the height of the local trees, corresponding to PSMC for sample size 2. Moreover, we use the coalescence time of two distinguished lineages, as has been successfully employed in the method SMC++ (Terhorst, Kamm, and Song, 2017).

We demonstrate the strength and weaknesses of different representations in simulation studies and applications to the 1000 genomes dataset. Moreover, we present the theoretical foundations to extend this framework to the total branch length of the local trees. We discuss potential extension of the framework to infer divergence times and migration rates in structured populations, and employing the posterior distribution of the local trees to detect regions under selection.

O-01-GF03

Detecting unknown introgressed archaic haplotypes in modern and ancient human genome sequences

Laurits Skov[1], Ruoyun Hui[2], Asger Hobolth[1], Aylwyn Scally[2], Mikkel Heide Schierup[1], Richard Durbin[2]

[1]Aarhus University (Denmark), [2]University of Cambridge (United Kingdom)

Introgression of archaic haplotypes into human populations is an already known phenomenon, with some haplotypes even providing a selective advantage such as adaptation to living in high altitudes or haplotypes carrying alleles of genes involved in the immune-system.

Most published methods for identification of archaic haplotypes rely on ancient DNA samples from the archaic population, to compare the modern samples with. Here we use a novel method for identifying regions of archaic introgression in ancient and modern human genomes. The method is a Hidden Markov Model that classify the genome into human any number of archaic states.

Not relying on a reference genome allows us to identify the introgressed haplotypes much more accurate. This allows us to calculate the length distribution and then estimate the admixture times with archaic hominins for different present day and ancient human populations.

Because our method does not rely on an archaic reference genome we are able to reconstruct the several MB of the unknown introgressing hominin genomes. These include unknown archaic populations from Africa, East and South East Asia along with a highly divergent archaic population into Melanesian populations.

O-01-GF04

Assessing population structure through time using ancient DNA

Lindsey Fenderson[1], Bastien Llamas[1], Julien Soubrier[1], Alan Cooper[1]

[1]University of Adelaide (Australia)

Recognizing changes in population structure through time is vital to understanding a species evolutionary history, from identifying admixture, to detecting range shifts, and ascertaining founding or extirpation events. Moreover, unaddressed population structure causes significant biases for many evolutionary analyses, and a firm grasp of temporal population structure should be a critical first step in any study of demography or adaptation. Yet, few currently available analysis methods are explicitly designed to address temporal population structure and little is known about the potential biases or pitfalls of population structure analyses when applied to heterochronous samples. To address this knowledge gap, we simulated SNP data for 12 known static and dynamic population histories and analyzed the temporal genetic data with 7 population structure analysis methods: Admixture, principal components analysis, network theory clustering, discriminant analysis of principal components, spatiotemporal principal components analysis, spatial factor analysis, and Tess. The accuracy and interpretability of each method was assessed to clarify potential biases and highlight the pros and cons of each approach under different evolutionary scenarios. We further evaluated the usefulness of each analysis approach after subsampling and modifying the simulated data to mimic irregular, opportunistic sampling of historical specimens as well as ancient DNA damage. Our results illuminate the temporal utility and limitations of common population structure analysis approaches and are highly relevant to any research that incorporates samples from disparate time periods.

O-01-GF05

Native American Genetic History Through Admixed Brazilians

Alex Mas-Sandoval[1, 2], Lara R Arauna[2], Eduardo Tarazona-Santos[3], Maria Catira Bortolini[1], David Comas[2]

[1] Universidade Federal do Rio Grande do Sul (Brazil), [2]Universitat Pompeu Fabra (Spain), [3]Universidade Federal de Minas Gerais (Brazil)

Brazilian urban admixed populations are composed of three main genetic components: Native American, European and African. Although their proportions between and within populations are different, Native American component is usually the one found in lowest proportions (6-8%). We take advantage of 6487 urban admixed Brazilians genotyped at 2,5M SNPs in three locations of eastern Brazil (North-East, South-East and South regions) to extract the Native American haplotypes and rearrange them in Assembled Individuals with full Native American ancestry, emulating a historical landscape where Native Americans had not admixed with Europeans and Africans. This method allows us to analyze the Native Americans population structure in a region where most of Native American populations disappeared after European colonization. We find that population structure of Native American Assembled Individuals is driven by the geography of the sampling location while the populations structure of urban admixed Brazilians is modulated by the European or the African ancestry. Finally, in order to explore in depth the origin and the demographic history of the Brazilian Native American component we analyze how admixture with non-admixed Native-American populations have shaped these populations.

O-01-GF06

Reconstructing and dating gene flow using efficient haplotype-based techniques

Garrett Hellenthal[1]

[1]University College London (United Kingdom)

Analyses of genetic variation data have shown that populations sampled world-wide descend from multiple, genetically different groups that intermixed at different time periods, including the recent past. In individuals carrying mixed ancestry, statistical models can study the lengths of contiguous segments of DNA inherited from each mixing group to date precisely when these different groups intermixed. This talk will describe the advantages of such methods over other approaches to detect intermixing (i.e. admixture) events, illustrating how they can (1) identify which groups mixed in the absense of any prior information, (2) determine precisely when these groups mixed, and (3) potentially increase power by exploiting haplotype information from correlations among tightly linked Single-Nucleotide-Polymorphisms (SNPs). I describe a new, computationally efficient haplotype-based approach that can infer admixture in genetically homogeneous populations consisting of hundreds to thousands of individuals that share a similar ancestral history, an important problem as increasingly more data are collected from narrow geographic regions. I also describe how DNA from ancient human remains (aDNA) can be readily incorporated into these models despite their currently small sample sizes, e.g. even if there is only one sample per culture, with additional extensions to extract haplotype information from low-coverage aDNA samples. I highlight applications to data from Africans, Latin Americans and other world-wide populations, showcasing the ability of genome-wide variation data to provide new insights into human history.

SY10: Genetic conflict and speciation (July 9, 13:30–15:30)

O-01-GC01

Conflict and speciation: do empirical data support a role for conflict in the isolation of plant species?

Leonie Moyle[1]

[1]Indiana University (United States)

Because evolutionary conflicts, including genetic conflict, can impose strong antagonistic selection on the loci involved, conflict has been proposed as a potentially important driver of lineage divergence and speciation. Theories connecting the expression of reproductive isolation with divergence in selfish genetic elements, and/or with differences in the expected magnitude of antagonistic selection among lineages, could in principle apply broadly. Nonetheless, beyond several elegant cases in model systems, the general importance of conflicts in the evolution of reproductive isolation is unclear. I will discuss empirical evidence that conflict could play a role in genetic divergence, and the expression of reproductive barriers, among plant lineages. This includes data on whether loci expected to experience stronger antagonistic selection also show faster evolutionary rates, and evidence for an association between the expected magnitude of conflict among lineages and the strength of their species barriers. Assessing evidence for these broad associations is essential for evaluating a general role for conflict as a driver of speciation.

O-01-GC02

Genetic conflict and speciation in Drosophila

Nitin Phadnis[1]

[1]University of Uth (United States)

Speciation, the process by which one species splits into two, involves the evolution of reproductive barriers such as the sterility or death of hybrids between previously interbreeding populations. Even in his masterpiece On The Origin of Species, Darwin could find no satisfactory solution to the apparent paradox of why natural selection would tolerate the onset of genetic barriers such as hybrid sterility and inviability that diminish the prospect of successful reproduction and, therefore, termed this problem the mystery of mysteries. Here, I describe the key developments with new genomics and cell biological approaches that are rapidly changing our understanding of the molecular basis of speciation. Our studies of the cellular and developmental anomalies in inter-species hybrids also provide surprising insights into the otherwise hidden evolutionary conflicts that ultimately shape the architecture of our genomes, cells and species.

O-01-GC03

Rapid evolution leads to rapid onset of centromeric histone incompatibility in Drosophila

Emily Coelho*[1, 2], Aida Flor de la Cruz*[1, 2], Benjamin D. Ross[1, 3], Harmit Singh Malik[1, 2]

[1]Fred Hutchinson Cancer Research Center (United States), [2]Fred Hutchinson Cancer Research Center (United States), [3]University of Washington (United States)

Chromosome segregation is essential for faithful transmission of genetic information to daughter cells during cell division. In eukaryotes, accurate chromosome segregation relies on specific chromosomal regions called centromeres, which recruit centromeric histones (CenH3) and other components of the kinetochore. First identified as Cenp-A in mammals, CenH3 replaces canonical H3 in centromeric nucleosomes. Despite its essential role in eukaryotic chromosome segregation, CenH3 in plants and animals evolves rapidly. It remains unclear why an essential gene required for chromosome segregation is subject to adaptive evolution. One hypothesis to explain this rapid evolution is that CenH3 and other centromeric proteins evolve rapidly to suppress the deleterious consequences associated with the rapid evolution of centromeric DNA satellites, resulting from meiotic drive during female meiosis.

We reasoned that if CenH3 positive selection occurred to suppress deleterious effects of centromeric DNA satellite changes, then replacing D. melanogaster CenH3 with 'ancestral' mal-adapted CenH3 in vivo might reveal these deleterious consequences. To test this hypothesis, we generated allelic swaps of FLAG-tagged CenH3 alleles in D. melanogaster in the native location and assessed viability and fertility of the resulting flies. Our studies reveal incompatibilities of the Dmel-Dsim ancestor and D. simulans CenH3 alleles, but not D. melanogaster alleles in the D. melanogaster genome. Furthermore, our findings support both an epigenetic as well as genetic basis of CenH3 centromere identity inheritance. Studying consequences of adaptively evolving essential genes in vivo is a powerful strategy to reveal the causes and consequences of genetic conflicts.

O-01-GC04

Comparative genomics reveals rampant gene duplication and reorganization of the Drosophila melanogaster and the simulans clade Y chromosomes

Ching-Ho Chang[1], Colin Meiklejohn[2], Taylar Mouton[1], Amanda Larracuente[1]

[1]U of Rochester (United States), [2]University of Nebraska, Lincoln (United States)

Non-recombining Y chromosomes in Drosophila species are gene-poor but contribute to phenotypic variation and hybrid incompatibilities. We know little about the genetic sources of this variation because Y chromosomes are difficult to sequence and assemble. We used long-read sequencing from Pacific Biosciences in *D. melanogaster*and three simulans clade species to discover the sequence and organization of their Y chromosomes. Our assemblies are the first for the simulans clade Y chromosomes and exceed the *D. melanogaster* reference Y chromosome 4-fold in sequence and contiguity. We discovered the rapid evolution of Y-linked gene copy number, intron size, repeat content, and gene order across these 4 species. We identified 56 loci that independently duplicated from other chromosomes to the Y chromosome, including the gene *SR Protein Kinase (SRPK)*. The Y chromosome duplicates (*SRPK-Ys*) arose in the ancestor of the simulans clade and amplified in each lineage to over 40 copies. Interestingly, *SRPK-Ys* retained exons from a testis-specific isoform deleted in the autosomal copies. *SRPK-Ys* evolve three-fold faster and are overexpressed 20-fold in the testes of *D. simulans* compared to their parental copy in *D. melanogaster*. These data suggest that *SRPK-Ys* may be under sexually-antagonistic selection and upregulated via gene duplication. Interestingly, we also find evidence for upregulation of conserved Y-linked genes, e.g. ARY via duplication. Our results suggest that gene duplications may accelerate Y chromosome evolution in Drosophila and subfunctionalization of *SRPK* and *SRPK-Ys* may contribute to hybrid incompatibilities between *D. melanogaster* and species of the simulans clade.

O-01-GC05

Allorecognition loci act as speciation genes in Podospora anserina

S. Lorena Ament-Velasquez[1], Aaron Vogan[1], Corinne Clave[2], Eric Bastiaans[3], Jesper Svedberg[1], Suzette de Groot[3], Alfons Debets[3], Sven J. Saupe[2], Hanna Johannesson[1]

[1]Uppsala University (Sweden), [2]Universite de Bordeaux (France), [3]Wageningen University & Research (Netherlands)

Understanding the origins of reproductive isolation is key for speciation research, but a tractable study system is often lacking. The fungus *Podospora anserina* has long been considered a single species with abundant intraspecific sexual incompatibilities. This taxon is a model organism for the study of vegetative self/nonself discrimination (allorecognition) governed by *het* genes, as well as for meiotic drive (Mendelian segregation distortion). Previous observations suggest a connection between vegetative and sexual recognition systems, but information on the evolutionary history is needed in order to confirm the link. Here, we used Illumina and PacBio technologies to sequence a large collection of *P. anserina* strains, along with related taxa. We showed that *P. anserina* maintains balanced polymorphisms at most known *het* genes, as predicted for allorecognition loci. Importantly, we discovered that mating success correlates nearly perfectly to the identity of the *het-v* gene. The epistatic interaction of *het-v* alleles with the unlinked *het-r* gene defines the boundaries of two mating groups. We confirmed a significant deficit of recombinant genotypes in the wild, demonstrating the lack of current mixing. Moreover, the genomic regions around both loci are strongly differentiated between groups, but divergence elsewhere in the genome is minor. We conclude that the interaction of *het-v* and *het-r* acts as an intrinsic reproductive barrier, and suggest that they drive the process of incipient speciation in *P. anserina*. Lastly, we discovered that *het-v* is physically linked to meiotic drivers, suggesting an interaction between segregation distorters and a gene directly responsible for reproductive isolation.

O-01-GC06

Evolution-guided mutagenesis to understand antiviral protein function

Rossana Colon-Thillet[1, 3], Emily Hsieh[1], Maria Gutierrez [4], Richard McLaughlin[1], Harmit Singh Malik[1, 5], Michael Emerman[1, 2]

[1]Fred Hutchison Cancer Research Center (United States), [2]Fred Hutchison Cancer Research Center (United States), [3]University of Washington, Seattle (United States), [4]University of Texas, El Paso (United States), [5]Search Results The Howard Hughes Medical Institute (United States)

The dynamin-like GTPase MxA is a rapidly evolving antiviral protein in primates. In particular, the Loop 4 (L4) domain of MxA shows an enrichment of positively selected sites. Evolution-guided approaches performed in our laboratory identified the MxA L4 as a determinant for antiviral specificity against two orthomyxoviruses: Thogoto virus (THOV) and influenza virus (IAV). We found that single amino acid changes at position 561 underlie MxA's antiviral specificity against THOV and IAV. However, these experiments do not evaluate the contribution of the other rapidly evolving sites in the L4. In this work, we take an unbiased approach to evaluate the contribution of all five rapidly evolving sites in L4 to antiviral activity. We built a library of MxA variants that encode all amino acid combination at the five rapidly evolving sites. We assessed the ability of nearly 1000 randomly selected variants to restrict THOV. We found that only 5% of the variants retain wildtype human MxA-like activity; all active variants encode F, W or Y at residue 561. Interestingly, we noticed several variants that contained these aromatic amino acids at 561 but were inactive against THOV, an indication of negative epistasis. However, these 'inactive' variants retained activity against IAV. In parallel, we discovered a handful of L4 combinations that had better specificity against THOV compared to wt human MxA, but these 'super-restrictor' variants were poorer against IAV. Together, our results indicate that the positively selected sites in the L4 of MxA modulate functional tradeoffs in target recognition and specificity.

O-03-GU01

Molding the genome with LAVA: Exploring functional roles of a gibbon-specific retrotransposon

Lucia Carbone[1, 3], Pryce Michener[1], Kimberly A Nevonen[1], Brett Davis[1], Jeff D Wall[2], Mariam Okhovat[1]

[1]Oregon Health & Science University (United States), [2]University of California, San Francisco (United States), [3]Oregon National Primate Research Center (United States)

Understanding the molecular mechanisms that drive evolution of novel, species-specific traits is a long-standing problem in biology. We have been studying gibbons as these species have evolved many distinct traits and are the first hominoids to split from the common ancestor. Specifically, gibbons shows unusually rearranged karyotypes compared to other primates with chromosome numbers ranging from 38 to 52. Moreover, they have acquired several adaptations to become the ultimate brachiators, including modified wrists, shoulders, and tendons. We have been surveying the gibbon genome and epigenome to identify possible contributors to evolution of novelty. In particular, we have been examining the role of the gibbon specific retrotransposon, the LAVA element, in shaping the gibbon genome and transcriptome by various transcriptional and post-transcriptional mechanisms. To this end, we generated 23 gibbon whole-genome sequencing data (10-30X) from the 4 extant gibbon genera and identified polymorphic (4,585) and fixed across genera (905) LAVAs. We find that these insertions mainly (57%) occur in introns or nearby genes. We additionally collected RNA-seq data from EBV-transformed cell-lines of a subset of 12 individuals to assess the effect of LAVA on gene transcription and termination. Finally, we characterized the epigenetic landscape of LAVA insertions and their consequences by various ChIP-seq and DNA methylation datasets. Through this wide array of genetic and epigenetic data, we aim to shed light on the role of the LAVA elements in gene regulation and evolution of novel traits in gibbons.

SY11: Genomic underpinnings of primate phenotypic evolution and diversity (July 11, 9:30–11:30)

O-03-GU02

Variant discovery and consequence in the genomes of a bottlenecked vervet population

Vasily Ramensky[1, 2], Anna J Jasinska[1, 3], Susan K Service[1], Richard K Wilson[4], Mark Cline[5], Jay R Kaplan[5], Matthew J Jorgensen[5], Nelson B Freimer[1], Wesley C Warren[6]

[1]Center for Neurobehavioral Genetics, University of California Los Angeles (United States), [2]Moscow Institute of Physics and Technology (Russian Federation), [3]Polish Academy of Sciences (Poland), [4]Institute for Genomic Medicine (United States), [5]Section on Comparative Medicine, Wake Forest School of Medicine (United States), [6]Washington University School of Medicine (United States)

Studies on the genetic history of human populations has shown some experienced founder effects, such as Finland, which result in an observed higher frequency of damaged genes, classified as loss of function (LOF) variants. Presumably, this is due to the reduced level of recombination as a means of purifying these deleterious alleles. Unlike other non-human primates adopted for the study of human disease, vervet research populations experienced two key genetic bottlenecks that like human founder effect populations also make them particularly well-suited for the study of putatively deleterious sequence variation. In this study we utilize 719 pedigreed vervets sequenced at various coverages, align all to the Chlorocebus_sabaeus 1.1 reference genome then call SNPs and indels using the Genome Analysis Toolkit (GATK) v3.4. From all variants, we derive a set of LOF variants using the PolyPhen-2 software with prioritized categories: stop-gain, stop-loss, and splice-site (donor, acceptor). We observed a total of 1,051,886 SNPs and 241,648 indels that when assessed for LOF state among autosomal genes with human orthologs yielded an average of 166 per monkey. Upon associating homozygous LOF gene variants to recurrent phenotypes observed in the pedigree, we find genes unique to mothers that experience repeated early infant mortality. Several genes that are linked to this phenotype reveal overlap with canonical pathways that are reported in the occurrence of human miscarriages. Our findings illustrate the opportunities for the discovery of damaged alleles among bottlenecked populations of non-human primates and their potential health consequence.

O-03-GU03

Population genomics of white faced capuchin monkeys (Cebus capucinus imitator) with unbiased fecal genomes

Joseph Orkin[1, 2], Mike Montague[3], Aoifie Doherty[4], J. Pedro Magalhaes[4], Wes Warren[5], Shoji Kawamura[6], Tomas Marques-Bonet[7], Amanda Melin[1, 2]

[1]University of Calgary (Canada), [2]University of Calgary (Canada), [3]University of Pennsylvania (United States), [4]University of Liverpool (United Kingdom), [5]Washington University in St. Louis (United States), [6]The University of Tokyo (Japan), [7]Universitat Pompeu Fabra/CSIC (Spain)

To better understand the genetic basis of capuchin monkey biology and adaptation, we report the first annotated, high-coverage genome assembly of a white-faced capuchin monkey, along with 24 additional genomes from two populations of capuchins inhabiting drastically different habitats. The *Cebus capucinus imitator* 1.0 genome was generated from 81x (50x fragments, 26x 3kbs, and 5x 8kbs) coverage on an Illumina HiSeq 2500 instrument. The combined sequence reads were assembled using ALLPATHS-LG software, resulting in a 2.72 Gb genome with a contig N50 of 41.2 kb and scaffold N50 of 5.27 Mb. We annotated the genome with a *de novo* assembled transcriptome (built using Trinity software) that resulted in 37,471 genes and pseudogenes, 20,740 of which are protein-coding.

We analyze the genomes of 6 capuchins inhabiting the highly seasonal dry forests of the Guanacaste province, and 4 genomes of capuchins residing near Manuel Antonio National Park, a lowland rainforest. Importantly, we describe a genome assembly for a non-model organism, and demonstrate the successful application of a new method of obtaining whole genome data from non-invasively collected field samples of low quality (feces). We focus on genes under positive selection in capuchins, and on genes that underlie sensation, cognition, and immune function due to the tendency of these genes to evolve quickly in primates, including our lineage, and other animals, and also due to their relevance to capuchin-specific biology and adaptation.

Acknowledgements: ACHRI, University of Calgary, Washington University in St. Louis, Kids Saving the Rainforest

O-03-GU04

Genomic signatures of high altitude adaptation in gelada monkeys (Theropithecus gelada)

Noah Snyder-Mackler[1, 2, 3], Kenneth Chiou[1, 2]

[1]University of Washington (United States), [2]University of Washington (United States), [3]University of Washington (United States)

Life at high altitude is associated with myriad physiological challenges, including exposure to conspicuous stressors such as hypoxia and extreme cold. Consequently, most animals living at high altitude have been under strong selection to develop adaptations to these challenges. Unveiling adaptations in other high-altitude-living animals, including nonhuman primates, could therefore help illuminate the mechanisms underlying adaptive evolution of myriad traits. Here, we investigated the genetic adaptations to high altitude in a novel nonhuman primate model, the gelada monkey. To do so we generated a de novo genome assembly and regulatory map of the gelada. We generated a 90x coverage Illumina library and linked it into megascaffolds using a 10x coverage Hi-C library, which resulted in a total assembled genome of 2.7Gb. The transcriptome was annotated with RNA-seq libraries generated from wild gelada skin and blood cells. We then investigated which gene families have undergone significant expansion in geladas compared to their close phylogenetic relatives, macaques, humans, and baboons. To expand this comparison, we then deeply sequenced (10x coverage) blood samples from 25 high altitude geladas and 25 low altitude hamadryas baboons. These data allowed us identify candidate loci that show signatures of positive selection in high-altitude geladas relative to their close phylogenetic relatives. Follow up studies will determine the functional impacts of variants underlying high-altitude adaptation on gene regulation using a massively parallel functional approach (STARR-seq). Together, these results significantly advance our knowledge of genotype-phenotype relationships underlying high-altitude adaptations in a comparative primate model.

O-03-GU05

Duplication and Convergent Evolution of the Pancreatic Ribonuclease Gene (RNASE1) in a Non-Colobine Primate, the Mantled Howler Monkey (Alouatta palliata)

Mareike Cora Janiak[1, 2], Andrew S Burrell[3], Todd R Disotell[3]

[1]Rutgers University (United States), [2]University of Calgary (Canada), [3]New York University (United States)

Pancreatic ribonuclease (*RNASE1*) and lysozyme C (*LYZ*) are enzymes, originally involved in immune defense, that have evolved new digestive functions in foregut fermenting animals, like colobine monkeys and artiodactyl ruminants. In colobines, *RNASE1* was duplicated, in some species twice, with the daughter genes (*RNASE1B, RNASE1C*) evolving new digestive roles. Howler monkeys are the most folivorous of the New World monkeys but, lacking the sacculated stomachs of colobines, digest foliage using caeco-colic fermentation. We report on the *RNASE1* and *LYZ* genes in the mantled howler monkey (*Alouatta palliata*). Results indicate that the *RNASE1* gene was duplicated twice in *A. palliata*, leading to two daughter genes, *RNASE1B* and *RNASE1C*. While the parent gene (*RNASE1*) is conserved, *RNASE1B* and *RNASE1C*have multiple amino acid substitutions that are convergent with those found in the duplicated *RNASE1* genes of colobines. As in colobines, the duplicated RNases in *A. palliata* have lower isoelectric points, a lower charge, and changes that are indicative of a reduced efficiency against double-stranded RNA, suggesting a novel, and possibly digestive function. Howler monkey *LYZ* is conserved and does not share the substitutions found in the colobine and bovine sequences. These findings suggest that in both foregut and caeco-colic fermenting primates pancreatic ribonuclease has convergently evolved a new role for digesting the products of microbial fermentation. Energy gains from the digestion of these products can be substantial, therefore, these duplicated proteins may be crucial digestive enzyme adaptations allowing howler monkeys to survive on a folivorous diet during times of fruit scarcity.

SY11: Genomic underpinnings of primate phenotypic evolution and diversity (July 11, 9:30–11:30)

O-03-GU06

Ancient proteins and the thrifty gene hypothesis: Uric acid's contribution to primate evolution

Jennifer E. Farrar[1], Lily Tran[2], Eric A. Gaucher[2]

[1]Georgia Institute of Technology (United States), [2]Georgia State University (United States)

The protein uricase, which breaks down the insoluble molecule uric acid, is found ubiquitously throughout each domain of life. However, certain animal lineages, including birds, reptiles, and apes, have experienced distinct evolutionary events leading to the pseudogenization of uricase. In order to understand the evolutionary process behind the inactivation of the uricase gene in the primate lineage, we have synthesized ancestral uricases by employing evolutionary models to predict their ancestral sequences. We discovered that uricase gradually lost activity prior to its inactivation. Ancestral sequence reconstruction on two uric acid transporters, URAT1 and ABCG2, also showed that the function of the ancient transporters changed concomitantly with the ancient uricases. Further, phylogenetic analysis of the enzyme that converts xanthine to uric acid (xanthine oxidoreductase, XOR) suggests that functional constraints acting on this enzyme changed at the same time that uricase and its transporters experienced a change in function. It is unknown why these events have occurred, particularly because the buildup of uric acid in the body causes hypertension, renal disease, liver damage and gout. However, with our collaborators, we have shown that increased uric acid levels are important for the conversion of fructose into triglycerides. These results support a recent hypothesis that frugivory is responsible for the evolution of large brain size in primates. Higher levels of uric acid would have facilitated the digestion of the fructose-rich diets of these primates, allowing more energy to be used for an increased encephalization quotient in primates.

SY12: Genomics and evolution of symbiotic interactions (July 9, 10:30–12:30)

O-01-GE01

The genome of *Paulinella* reveals pathways of plastid integration

Debashish Bhattacharya[1, 3], Hwan Su Yoon[2], Udi Zelzion[3], Dana C Price[4], Eva Nowack[5], Duckhyun Lhee[2]

[1]Rutgers University, New Brunswick (United States), [2]Sungkyunkwan University (Republic of Korea), [3]Rutgers University, New Brunswick (United States), [4]Rutgers University, New Brunswick (United States), [5]Heinrich-Heine-Universitaet, Duesseldorf (Germany)

The endosymbiotic origin of oxygenic photosynthesis in the ancestor of algae and plants was a turning point for our planet, ultimately laying the foundation for the rise of humans and other multicellular life. The widespread photosynthetic organelle in eukaryotes, the plastid, originated >1 billion years ago in a heterotrophic protist via primary endosymbiosis. In my talk, I will summarize some key aspects of genome evolution in these taxa to set the stage for our recent work with the only other known case of primary plastid origin, in the clade of photosynthetic amoebae, *Paulinella*whose plastid originated about 100 million years ago. Members of this genus contain a permanent photosynthetic organelle, referred to as a chromatophore that is derived from an alpha-cyanobacterial endosymbiont. I will present the results of our collaborative research on the behavior of *Paulinella* cultures, its physiology, bacterial symbionts, and of genome and transcriptome data from two species, *P. chromatophora*and *P. micropora*. This information provides the first direct insights into what was once considered an unreachable goal: to reconstruct with relative confidence, early events that mark the transition from symbiont to organelle.

SY12: Genomics and evolution of symbiotic interactions (July 9, 10:30–12:30)

O-01-GE02

Experimental evolution of an insect-bacterium symbiotic association

Takema Fukatsu[1]

[1]National Institute of Advanced Industrial Science and Technology (AIST) (Japan)

Many insects are obligatorily associated with symbiotic microorganisms in their gut, body cavity or cells. Such obligate symbionts are essential for survival and reproduction of their hosts by playing important biological roles like provisioning of essential nutrients, and therefore tend to be conserved within each insect taxon.

Recently, we discovered an unprecedented case wherein an insect species consists of local populations associated with distinct lineages of bacterial symbionts. The symbiont polymorphisms are found not only between the insect populations but also within each of the insect populations. Experimental studies showed that all the symbionts are essential for their hosts: symbiont elimination consistently resulted in host mortality. Experimental symbiont transfers revealed that the host-symbiont combinations are functionally exchangeable: mortality of an insect strain due to curing of its original symbiont can be consistently rescued by transfection with one of the different symbiont lineages. Notably, some symbiont lineages are with reduced genomes and uncultivable, while other symbiont lineages are with larger genomes and easily cultivable in vitro. These cultivable symbiont lineages are ubiquitously found in surrounding soil environments, and, although the host insects normally inherit the symbiotic bacteria vertically, the host nymphs are capable of acquiring the free-living symbionts environmentally.

Using the insect's symbiotic system, we have established an experimental association between the insect and an environmental free-living bacterium, thereby monitoring the process of symbiotic evolution in the laboratory.

SY12: Genomics and evolution of symbiotic interactions (July 9, 10:30–12:30)

O-01-GE03

Genomic hijacking - how parasitic worms manipulate their hosts

Eddy Dowle[1], Robert Poulin[1], Robert Day[1], Neil Gemmell[1]

[1]University of Otago (New Zealand)

Parasites routinely manipulate the behaviour of their hosts to enhance their survival and transmission. One of the most extraordinary of these host manipulations is the water-seeking behaviour that some nematodes and hairworms induce in their hosts so that the worms might exit the host in a suitable environment and reproduce. The worm hijacks the host's central nervous system forcing the normally terrestrial host to seek water. Once water is found the adult worm, often equal to or bigger than the host, erupts in an explosive frenzy, sacrificing the host, so that the parasite might complete its lifecycle. This amazing alteration in behaviour is induced by worms spanning two phyla (Nematoda and Nematomorpha) and is observed in a variety of arthropod hosts, notably crickets, weta and earwigs. Host manipulations are the consequence of genes in the parasite genome modifying the hosts' phenotypic traits. But the development and genetic control of these behavioural modifications are not well understood as experimentally tractable systems are rare. Using total RNA sequencing we reveal the transcriptomic changes in both the parasite and host brain, which is informing how host manipulations develop and the molecular mechanisms behind them. Using two evolutionarily divergent phyla (Nematoda and Nematomorpha) that induce the same behavioural manipulation in two different insects (earwigs and weta) allows us to determine the similarity of the genetic mechanism(s) across divergent lineages.

SY12: Genomics and evolution of symbiotic interactions (July 9, 10:30−12:30)

O-01-GE04

Comparative genomics reveals the distinct evolutionary trajectories of the robust and complex coral lineages

Hua Ying[1], Ira Cooke[2], Susanne Sprungala[2], Weiwen Wang[1], David Charles Hayward[1], Yurong Tang[1], Gavin Huttley[1], Eldon Edward Ball[1, 3], David John Miller[2, 3]

[1]Australian National University (Australia), [2]James Cook University (Australia), [3]James Cook University (Australia)

Although the timing of the origins and major divergences remain equivocal, all the available molecular data imply that most extant corals fall into two major clades known as the Complexa (complex corals) and Robusta (robust corals). This dichotomy is well supported in molecular analyses - the split is recognised as real, despite the fact that few morphological or biological criteria resolve the two groups. One reason for the lack of features distinguishing the two clades is the paucity of large molecular datasets for a representative range of corals. To provide a platform for investigation of both differences between individual species and general differences between complex and robust corals, genome sequencing and assembly was carried out on number of corals selected to reflect phenotypic and physiological diversity. The most significant implication of these comparative analyses was that, uniquely amongst animals, robust corals are capable of de novo histidine biosynthesis. Previously, the only known difference between corals with respect to biosynthetic capacity was the lack of the enzyme cystationine β-synthase (suggesting a requirement for cysteine) in *Acropora*spp. but not other corals. Whilst these metabolic differences may play roles in selection of compatible *Symbiodinium* strains, experimental support for this idea is presently lacking. Indeed, the robust corals studied here host strains of clade C and clade D *Symbiodinium*, as do many complex corals. Note, however, that enormous variation exists within the clades, and few genome data are available, so the possibility of metabolic influences on strain selection cannot be dismissed.

SY12: Genomics and evolution of symbiotic interactions (July 9, 10:30–12:30)

O-01-GE05

Rapid evolution of host dependence on environmentally acquired microbes

Aparna Agarwal[1], Deepa Agashe[1]

[1]National Centre for Biological Sciences (NCBS), Bangalore (India)

Eukaryotic hosts are often associated with microbes that enhance their fitness. Although faithful transmission of the microbes across generations is important to establish the relationship, in some cases the microbes are environmentally acquired. Thus, when hosts disperse to new habitats, they may not have access to the appropriate microbes, leading to reduced fitness. In such cases, can the host establish new microbial partnerships? Are such partnerships specific, and how long does the established process require? We address these questions with the red flour beetle *Tribolium castaneum*, a generalist insect pest. We find that beetle fitness is enhanced by flour-acquired microbes in their ancestral habitat (wheat flour), but not in novel environments carrying a different microbial community (e.g. millet flour). Thus, when introduced to novel habitats, the lack of appropriate flour-derived partner microbes reduces beetle fitness. During subsequent laboratory adaptation to novel habitats, we found that beetles again established a beneficial relationship with specific flour-derived microbes within 10-12 generations. Thus, microbes and their hosts can rapidly evolve to form new partnerships in novel environments, perhaps allowing generalist species to succeed despite habitat fluctuations and missing microbial partners.

SY12: Genomics and evolution of symbiotic interactions (July 9, 10:30–12:30)

O-01-GE06

The genome study of Apophlaea lyallii (Florideophyceae, Rhodophyta) provides new insights for the algal-fungal symbiotic relationship

Hwan Su Yoon[1], Chung Hyun Cho[1], Wendy Nelson[2, 3]

[1]Sungkyunkwan University (Republic of Korea), [2]National Instituted of Water & Atmospheric Research (New Zealand), [3]University of Auckland (New Zealand)

Over a billion years of evolutionary history, red algae have adapted in a wide range of environments, and diversified into more than 7,000 species. During the evolution, some red algal species have established parasitic or symbiotic relationships with other organisms to help better survival in harsh environments. One well-known example is the genus *Apophlaea*, which forms an obligate symbiosis with one marine fungal species *Mycosphaerella* sp. (Capnodiales). *Apophlaea* is a member of the Hildenbrandiophycidae (Florideophyceae) and distributed only in intertidal zones of New Zealand. *Apophlaea* species are well known for high resistance to various stresses including ultraviolet radiation and desiccation. To investigate the genomic view of the symbiotic relationship, we have sequenced the nuclear genome of *Apophlaea lyallii*using next generation sequencing methods. Based on a comparative genomic analysis, we report on the general features of the genome and address the algal-fungal symbiotic relationship related to the environmental tolerance.

O-03-HE01

ANCESTRAL AND ADMIXTURE HISTORIES IN THE INDIAN SUBCONTINENT

Partha P Majumder[1]

[1]National Institute of Biomedical Genomics (India)

Based on the analyses of genome-scale SNP data and whole-genome sequence data from individuals drawn from a large number of extant ethnic groups, we have estimated archaic admixture contributions of Neanderthal and Denisovan genomes to extant ethnic groups in the Indian sub-continent. We have found a high level of Denisovan admixture among the Negrito tribals of Andaman and Nicobar Islands. Comparison with similar data from other Asian populations has provided clues to areas of interaction of modern humans and Denisovans. Some populations from South and Southeast Asia harbour a small proportion of ancestry from an unknown extinct hominin; this ancestry is absent from Europeans and East Asians.

Our results show that that all Asian and Pacific populations share a single origin and expansion out of Africa, contradicting an earlier proposal of two independent waves of migration.

Systematic analysis of genome-wide data, using multiple robust statistical methods, from individuals drawn from populations selected to represent geographic, linguistic, and ethnic diversities have revealed four major ancestries in mainland India. This contrasts with an earlier inference of two ancestries based on limited population sampling. A distinct ancestry of the populations of Andaman archipelago was identified and found to be co-ancestral to Oceanic populations. Prior to the establishment of the socially-stratified caste system in India that imposed strict rules of endogamy, there was wide admixture across tribal and caste groups which came to an abrupt end 1,900 to 4,200 years before present.

O-03-HE02

Reconstructing the human population history of Africa

Brenna M Henn[1]

[1]UC Davis (United States)

Over twenty-five years ago, geneticists sequenced mitochondrial DNA from a diverse sample of human populations and hypothesized that all humans have a common origin in Africa 200,000 years ago.The broad outlines of this hypothesis remain remarkably unaltered, but many details of our African origin continue to be elusive. After decades of advances in human genetics, we are no longer data limited (either in terms of samples or genomic loci) but there is little consensus on most key issues. I will outline the models underlying the origin of modern humans. For example, was there a single ancestral population or multiple ancestral populations? Additionally, is there a discordance between anatomically modern humans and behaviorally modern humans? I will explore patterns of genetic diversity across Africa, and specifically focus on the complex history of southern African KhoeSan groups and adaptations to African environments. I will highlight our recent research on the evolution of skin pigmentation, malarial resistance, and infectious disease. We demonstrate that Africa continues to be an evolutionarily dynamic, heterogenous continent for human populations.

O-03-HE03

Genome wide analysis of negrito groups in Southeast Asia

Timothy Jinam[1], Naruya Saitou[1]

[1]National Institute of Genetics (Japan)

The term "negrito" refers to indigenous groups in the Philippines, Malaysia and Andaman Islands who share some physical characteristics such as short stature, frizzy hair and darker skin. Genome-wide SNP analysis showed the negrito groups are phylogenetically basal to other Southeast and East Asians, but negritos also experienced substantial gene flow from these populations. SNP loci which have low Fst among all negrito groups but highly differentiated with East Asians may be associated with height, skin pigmentation and malarial resistance. We also found relatively high traces of Denisovan introgression in the Philippine Negritos, particularly in the Aeta. We will present preliminary results of whole genome sequencing analysis of 10 Philippine negritos.

O-03-HE04

Human prehistoric demography revealed by polymorphic pattern of CpG transitions

Xiaoming Liu[1]

[1]School of Public Health, The University of Texas Health Science Center at Houston (United States)

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

O-03-HE05

Mother' tongues? A global study of sex-biased genetic and linguistic transmission after Out-of-Africa

Nicole Creanza[1], Alexandra Surowiec[1]

[1]Vanderbilt University (United States)

Worldwide patterns of genetic variation are driven by human demographic history during and after Out-of-Africa. To test whether this demographic history has left similar signatures on languages to those it has left on genes, we analyzed phonemes from 2,082 languages and microsatellite polymorphisms from 246 populations. Globally, populations that were geographically closer tended to be more similar genetically and linguistically. Our analysis suggests that two processes influence this pattern: vertical transmission of both genes and languages after Out-of-Africa, and linguistic borrowing (often coupled with genetic admixture) when neighboring populations speak very different languages. To understand whether sex-biased patterns in human history affect genetic and cultural evolution, we merged genetic, linguistic, and ethnographic data to perform two new studies. First, we address the hypothesis that children preferentially learn language from their mothers by comparing linguistic variation to mitochondrial and Y-chromosome variation separately. We show that long-range patterns of correlation between linguistic and genetic distance are present for both maternally inherited genes (mtDNA) and paternally inherited genes (Y-chromosome). However, mitochondrial genetic variation also significantly correlates with short-range patterns of linguistic variation (calculated with a different metric), whereas Y-chromosome variation does not. In addition, we gathered ethnographic data to categorize 182 genotyped populations by kinship system and postmarital-residence pattern. With this annotated database, we observe measurable evolutionary effects of matrilineal kinship and matrilocal residence; our results suggest that matriliny and matrilocality strengthen the association between languages and geography. These analyses shed new light on genetic and cultural population histories after Out-of-Africa.

O-03-HE06

Population Genomic Inference from Palaeo-Neutralomes of Mediaeval Germans

Joachim Burger[1], Michaela Harbeck[2], Krishna R. Veeramah[3]

[1]Institute of Organismic and Molecular Evolution, Mainz University (Germany), [2]SNSB Munich (Germany), [3]Stony Brook University (United States)

Many modern European states trace their roots back to the Early Medieval period. To genetically better characterize this formative phase of Central European population history we have conducted a population-level analysis of people from this era, generating genomic data from 41 graves from archaeological sites in present-day Germany mostly dating to around 500 AD. We developed a novel capture array that generated high coverage (mean 72x) sequence data in all individuals from neutral regions spanning a total of 5Mb, as well as 486 functional polymorphic sites. Examination of the panel of functional loci revealed that many alleles associated with recent positive selection were already at modern-like frequencies in European populations ~1,500 years ago. Analysis of the neutralome indicates that while men generally had ancestry that closely resembles modern northern and central Europeans, women exhibit a very high genetic heterogeneity; this includes signals of genetic ancestry ranging from West Europe to East Asia. The inferred female-biased migration indicates complex demographic processes during the Early Medieval Period may have contributed in an unexpected way to the making of the modern European genetic landscape. Lastly, we used the Medieval neutralome data to generate the allele frequency spectrum (AFS) and estimate the start and rate of recent European population growth.

SY14: Improving inference frameworks by accounting for population structure (July 9, 16:00–18:00)

O-01-II01

Modeling the interaction between population structure and selection

Sohini Ramachandran[1, 2], Lauren Alpert Sugden[1, 2], Michael C Turchin[1, 2]

[1]Brown University (United States), [2]Brown University (United States)

While the genomic era has yielded detailed knowledge about the strong footprint of population histories and gene flow on human genomes, the interaction of population structure and selection is discussed often but not well understood. Over the last decade, high-throughput studies identifying disease-associated mutations have largely focused attention on individuals with homogeneous and European genomic ancestry, greatly limiting the promise of personalized medicine. Only 414 published GWA studies to date have jointly analyzed multiple ethnic cohorts in the process of identifying human disease mutations; these studies often detect heterogeneous effect sizes for the same variant across ancestries, and usually identify unique associations within each cohort studied. Similarly, population structure is known to confound the identification of adaptive mutations, but few machine learning frameworks for detecting selection draw on the power of population differentiation to identify adaptive mutations.

I will discuss new statistical approaches - ranging from gene network analysis to machine learning approaches - developed by my group and others, that gain traction when inferring selection by explicitly accounting for population structure. These offer insight into how to construct inferential frameworks that leverage population structure to characterize signatures of selection.

SY14: Improving inference frameworks by accounting for population structure (July 9, 16:00–18:00)

O-01-II02

Mechanistic models of social processes impacting admixture

Amy Goldberg[1]

[1]UC Berkeley (United States)

Admixture is ubiquitous throughout the animal kingdom and human history. The admixture process is often complex, governed by social processes such as inbreeding avoidance, mate choice, and phenotypic preferences. We devise a flexible mechanistic model to describe the admixture history of a population, allowing for contributions from the source populations to vary over time and between sexes. Under the model, we study how sex-biased contributions and mate preferences affect the distribution of genetic ancestry and a simple quantitative trait in an admixed population. We demonstrate that cryptic sex bias or assortative mating can lead to biased inference of admixture histories, particularly for estimates of the timing of admixture based on the variance of ancestry. Admixture scenarios with continuing contributions from the parental populations over time can lead to qualitatively different inferences than admixture that occurs in a single generation. Our model provides a framework to quantitatively study admixture under flexible scenarios of mating and hybridization over time.

SY14: Improving inference frameworks by accounting for population structure (July 9, 16:00–18:00)

O-01-II03

Whole-genome hierarchical population structure analysis using network-based clustering

Gili Greenbaum[1], Amir Rubin[2], Ellie Armstrong[1], Alan R. Templeton[3, 4], Noah Rosenberg[1]

[1]Stanford University (United States), [2]Ben-Gurion University (Israel), [3]Washington University (United States), [4]Haifa University (United States)

Population structure is ubiquitous in natural populations, and the ability to detect it based on genetic information has been instrumental to our understanding of evolution and ecology. With whole-genome datasets of populations now becoming available, new computational challenges for detecting population structure arise, but so does the potential to study structure at many spatial scales and with finer detail than ever before. We develop a network-based model-free method that detects hierarchical population structure, is applicable to whole-genome datasets, and requires few prior assumptions. We first construct a network that describes genetic similarity between all pairs of individuals. We then apply network clustering techniques to detect dense substructures within the network, interpreted as population structure at a single scale. In order to detect structure at different scales, we iteratively prune weak connections from the network and re-detect clusters, focusing on the strong genetic similarities that are expected within fine-scale clusters. The results in a population structure tree that is amenable to ecological and evolutionary interpretation. We demonstrate the applicability of our method using two whole-genome datasets: tigers, an endangered species, and the model plant species *Arabidopsis thaliana*. We show that we are able to meaningfully interpret hierarchical population structure at many levels, including very fine scales: population fragments and familial groups in the case of tigers, and clusters delineating small-scale structure as small as several kilometers for *A. thaliana*. These results support previous studies regarding the coarse-scale delineation of subpopulations, but reveal many meaningful fine-scale clusters that were previously undetected.

SY14: Improving inference frameworks by accounting for population structure (July 9, 16:00–18:00)

O-01-II04

Introgression of a Y chromosome haplotype into a neo-Y karyotype in Rumex

Felix E.G. Beaudry[1], Stephen I. Wright[1]

[1]University of Toronto (Canada)

Chromosome fusions involving the sex chromosomes can play an important role in reproductive isolation between diverging populations. Furthermore, fusions may spread rapidly through populations if they are subject to directional selection or are favoured by meiotic drive. Here, we examine the population genomics of Rumex hastatulus, where a sex-chromosome autosome fusion has given rise to populations with two sex chromosome karyotypes in respective geographic ranges. Using genomic and transcriptome sequencing, we find evidence of limited gene flow between the two karyotype-populations, a pattern that is more pronounced on the X chromosome than the autosomes. We use demographic modelling with polymorphism data to estimate the extent of gene flow between the karyotype-populations. Interestingly, we find evidence for introgression of an entire Y haplotype across the hybrid zone despite limited gene flow. Polymorphism data support that this is unlikely to be due to incomplete lineage sorting. We explore whether these patterns could instead be due to sex- or pollen- specific selection, or by other processes such as female meiotic drive. Our study suggests that the relationship between demography, reproductive isolation and sex chromosomes in R. hastatulus is complex.

O-01-II05

Distinction between ancient introgression and incomplete lineage sorting in modern human genomes

Makoto Shimada[1]

[1]Institute for Comprehensive Medical Science, Fujita Health University (Japan)

Recent advancements of massive individual human genome and high-quality ancient genomes of other hominins have provided the evidences of diverse population structure in ancestral modern human before out-of-Africa, as well as ancient hybridization between modern human and archaic hominins such as Neanderthals. This brings a difficulty to distinguish current haplotype origins between ancient introgression from archaic hominins and incomplete lineage sorting before out-of-Africa.

To clarify origin of diversified haplotypes in modern human, I focused on eight loci (23 kb - 550 kb in length) containing diversified haplotypes that have been discussed whether these are exceptions of out-of-Africa model or not.

As many modern human haplotypes obtained from the 1000 genomes project as possible were used for comparison in topology between phylogenetic tree and network to understand actual picture of reticulation for each locus. Using S* and EHH (Extended Haplotype Homozygosity) analyses, we examined patterns of linkage disequilibrium (LD) of haplotypes to distinguished haplotypes origins between ancient introgression and ancestral polymorphism followed by incomplete lineage sorting.

These examination on LD analyses with gene genealogy showed actual state of complexity of human genome diversity and difference in origin of the diversified haplotypes among introgressed from archaic hominins whose genome sequences are available such as Neanderthal and Denisova, that from unknown other extinct hominins, and derived from ancient polymorphism without introgression.

SY14: Improving inference frameworks by accounting for population structure (July 9, 16:00–18:00)

O-01-II06

Genetic risk prediction across diverse populations

Alicia R Martin[1, 2, 3], Patrick Turley[1, 2, 3], Hailiang Huang[1, 2, 3], Raymond K Walters[1, 2, 3], Chia-Yen Chen[1, 2, 4], Max Lam[5], Duncan Palmer[1, 2, 3], Christopher R Gignoux[6], Eimear E Kenny[7], Benjamin M Neale[1, 2, 3], Mark J Daly[1, 2, 3]

[1]Massachusetts General Hospital (United States), [2]Broad Institute of Harvard and MIT (United States), [3]Broad Institute of Harvard and MIT (United States), [4]Massachusetts General Hospital (United States), [5]Institute of Mental Health (Singapore), [6]University of Colorado, Anschutz Medical Campus (United States), [7]Icahn School of Medicine at Mount Sinai (United States)

The vast majority of GWAS are performed in individuals of European descent; their applicability to other populations varies with genetic divergence, differences in LD and allele frequencies, and genetic architecture. Demographic models provide a critical lens into complex trait studies, including the transferability of genetic risk prediction to understudied populations. Previously, we simulated an out-of-Africa model to show that genetic risk predicted using European summary statistics transfers poorly to non-European populations. Here, we empirically evaluate genetic risk prediction across populations using results from the Psychiatric Genetics Consortium. We find that East Asian schizophrenia risk is better predicted by summary statistics from smaller East Asian cohorts (13k cases and 16k controls) than from ~3-fold larger European cohorts (37k cases and 113k controls, Nagelkerke's R^2 = .104 vs .066). Both in simulations and empirical data, we find that heritability varies with demographic history, with higher heritability identified in populations with larger effective population sizes. To improve cross-population genetic risk prediction, we develop a novel statistical method to improve prediction accuracy across populations when GWAS summary statistics are available from multiple populations. Our method computes the covariance of effect size estimates in each population weighted by LD structure in each respective population to more closely approximate causal effect sizes used in prediction. Even under neutral evolution (i.e. no heterogeneity of effect by population), our work cautions that findings from large-scale GWAS may have limited transferability across populations with standard approaches, highlighting the need to include more diverse individuals in medical genomics.

O-02-IA01

Understanding the structure and function of archaic ancestry in present-day humans

Sriram Sankararaman[1]

[1]UCLA (United States)

Analysis of the genomes of archaic hominins, such as Neanderthals and Denisovans, has revealed that these groups have contributed to the genetic variation of modern human populations. Characterizing these admixture events and their impact on the structure and function of human genomes is an important problem in human population genomics.

To systematically understand the structure of archaic admixture, we need to infer the locations of segments within the genomes of present-day humans that trace their ancestry to archaic hominins. I will describe machine learning algorithms that we have developed to infer maps of Neanderthal as well as Denisovan ancestry in diverse populations of present-day humans. Analyses of these maps document fine-scale variation in archaic ancestries across populations (including elevated Denisovan ancestry in populations in south and east Asia) as well as along the genome (driven partly by the action of selection on archaic alleles) and illustrate the impact of archaic alleles on the evolution of complex human phenotypes.

O-02-IA02

Using archaic introgression to infer sequence constraints that are shaping human enhancer evolution

Natalie Telis[3, 2], Kelley Harris[1, 2]

[1]University of Washington (United States), [2]Stanford University (United States), [3]Ancestry.com (United States)

Many enhancers annotated to date appear to modulate gene expression without exhibiting clear patterns of sequence conservation. A possible explanation is that many different sequences can perform similar enhancer functions; however, it is also possible that many sequences biochemically resemble enhancers without performing any essential functions. To assess whether the bulk of enhancers are truly bolstering human fitness, we take advantage of the natural experiment that occurred when Neanderthals interbred with humans, introducing divergent enhancer sequences throughout the genome. We find that enhancers are significantly depleted of Neanderthal DNA compared to control regions matched for background selection strength, though they are less depleted than genes or promoters. Highly pleiotropic enhancers are more depleted of introgression than cell-type-specific enhancers, but even enhancers that are active in just one cell type appear more intolerant to introgression than matched control regions. This trend holds true for every cell type profiled by ENCODE, though the magnitude varies across cell types. Neanderthal DNA depletion is correlated across tissues with skew in the site frequency spectrum toward rare variants, but this correlation is imperfect; for example, fetal enhancers seem to tolerate new mutations better than brain enhancers do, but of the two categories, fetal enhancers seem less tolerant of introgressed Neanderthal DNA. Fetal enhancers that are classified as human accelerated regions (HARs) show the strongest introgression intolerance of all, suggesting that fetal development has experienced more functionally significant changes since the human/Neanderthal split than other pathways and tissues that exhibit more enhancer sequence conservation overall.

O-02-IA03

No Evidence for Protracted Selection Against Neandertal Alleles in Humans

Benjamin Vernot[1], Martin Petr[1], Svante Paabo[1], Janet Kelso[1]

[1]Max-Planck-Institute for Evolutionary Anthropology (Germany)

Approximately 2% of the genomes of present-day non-Africans descends from Neandertal ancestors as a result of interbreeding between Neandertals and modern humans approximately 55,000 years ago. Negative selection against Neandertal alleles in the modern human population has been the subject of many studies.

When the proportion of Neandertal DNA has been estimated in ancient modern human specimens that range in age from 45,000 years to present-day, an apparent monotonic decline over time has been observed, and interpreted to be the result of negative selection against Neandertal alleles [Yang & Fu, 2018]. Here, we use a recently sequenced Neandertal genome from Croatia [Pruefer et al 2017] to better estimate Neandertal ancestry in ancient modern human genomes. In contrast to previous results, we observe no long-term decline of Neandertal ancestry. We show that the previous observation is likely an artifact produced by migration between Europe and Africa that altered relationships between European and African populations in a time-dependent manner.

With simulations using realistic recombination maps and selection coefficients, we show that selection against Neandertal alleles is likely to have occurred within the first 500 generations after introgression, i.e. prior to the age of most ancient human specimens analyzed to date. We note that our simulations agree with other observations supporting negative selection against Neandertal alleles, e.g. a relative depletion of Neandertal alleles around coding sequence and the presence of "deserts" of Neandertal sequence of size 10-15Mb.

O-02-IA04

Herbarium genomics infers the changes underlying the evolution of C4 photosynthesis in the Andropogoneae grasses

Matheus Bianconi[1], Jan Hackel[2], Alexandre Meunier[2], Maria Vorontsova[3], Pascal-Antoine Christin[1], Guillaume Besnard[2]

[1]University of Sheffield (United Kingdom), [2]CNRS-ENSFEA-IRD-Universite Toulouse III (France), [3]Royal Botanic Gardens (United Kingdom)

The Andropogoneae grass tribe includes some of the world's most important crop plants, such as maize, sugarcane and sorghum, and numerous dominant species of tropical grasslands. Its 1,200 species all use C_4 photosynthesis, which explains their fast growth and high productivity in tropical conditions. Inferring the early origins of C_4 photosynthesis in Andropogoneae has been hampered by a lack of closely-related C_3 lineages for comparative analysis. Using nuclear genome-wide data to screen the diversity stored in herbarium collections, we identified rare C_3 species that are sister to Andropogoneae. Phylogenomics revisit the timing of C_4 origin in this group and track genetic changes, showing that some enzymes acquired their C_4 properties independently during the early diversification of the group. This first phylogenetic study of grasses based on nuclear genome-wide data retrieved from both ancient DNA material and living specimens allows the investigation of early events during the evolution of a complex physiological trait in a diverse and ecologically successful group of grasses.

O-02-IA05

Tracking plant phenology and genetic diversity during environmental change using contemporary and historical samples

Patricia L.M. Lang[1], Franziska M. Willems[2], Oliver Bossdorf[2], Hernan A. Burbano[1]

[1]Max Planck Institute for Developmental Biology (Germany), [2]University of Tuebingen (Germany)

Climate change has already induced quantifiable species responses, earlier spring flowering being the prime example. The nature of these reactions however is largely unknown: are plant populations acclimatizing via plastic physiological responses? Or is adaptation by natural selection acting, changing the underlying genetic composition? As only the latter will enable long-term evolutionary responses, answering this question is crucial to understand consequences of global change.

To investigate changes in phenology (flowering time) and genetic diversity, we focus on understory herbs with a distinct, narrow spring flowering period that we believe has recently shifted. We study five such species in the Biodiversity exploratories, a German national project for biodiversity and ecosystem research, ongoing since 2006. In each the southern and central German exploratory, we sampled populations in 50 forest plots. Our approach couples phenology field surveys with analysis of these contemporary samples, and local historical herbaria specimen collected from ~1800 to ~1990.

However, non-model species and ancient herbarium DNA (aDNA) are challenging: Our target-species lack a reference genome, and the fragmentation of aDNA precludes using restriction enzyme-based sequencing strategies. To overcome both, we refined a method merging in-solution hybridization and RAD-sequencing for aDNA (hyRAD). With the resulting data from >300 historic and >1000 modern samples across five species and ~200 years, we can investigate temporal trends in genome-wide genetic diversity. Contrasting them with phenological-, and local environmental changes, we can disentangle plasticity from adaptation - and start defining the role of genetic changes for climate change responses such as accelerated flowering.

O-02-IA06

Selection trajectories of genetic variants underlying domestic animal traits

Evan K. Irving-Pease[1], Laurent A. F. Frantz[1, 2], Greger Larson[1], Joshua G. Schraiber[3, 4]

[1]University of Oxford (United Kingdom), [2]Queen Mary University of London (United Kingdom), [3]Temple University (United Kingdom), [4]Temple University (United Kingdom)

The study of animal domestication is an important model system for understanding adaptive responses to changes in environmental conditions, demography and selective pressures over time. Despite speculation surrounding the existence of 'domestication genes', the underlying genetic basis of traits differentiating domestic animals them from their wild counterparts remains poorly understood. Using genome-wide modern DNA, previous studies have contrasted populations of wild and domestic animals to scan for segregating signatures of selection in their respective genomes. Due to the intensive nature of modern breeding practices, it is unclear which candidate genes identified by these methods were under selection during the initial process of domestication, and which represent more recent improvement traits. Time series data, obtained from ancient DNA, can resolve these questions by directly observing changes in allele frequencies over time. Here, we reconstruct the allelic trajectory of hundreds of variants associated with quantitative trait loci (QTLs) in four key domestic species (cattle, pigs, horses and goats). Using a novel dataset of >300 ancient nuclear genomes, spanning ~12,000 years of evolutionary history, we are able to quantify the temporal origins and strength of selection for genetic variants associated with health, reproductive, performance, production, aesthetic and behavioural traits in domestic animal populations. The resulting timelines allow direct correlation between changes in ecological conditions within the domestic niche and selection for specific adaptive traits. Our results demonstrate the critical importance of time series data in resolving the underlying evolutionary process of animal domestication.

O-03-IH01

The evolutionary dynamics of untreated HIV and the maintenance of the latent reservoir

Fabio Zanini[2], Jan Albert[3], Richard Neher[1]

[1]University of Basel (Switzerland), [2]Stanford University (United States), [3]Karolinska Institute (Sweden)

HIV populations in infected individuals are evolving rapidly in response to selection by the host immune system. We performed whole genome deep sequencing of 6-12 longitudinal plasma samples from 11 HIV infected individuals spanning 5-8 years of infection. These data allowed us to characterize the evolutionary dynamics in detail and quantify adaptation and in-vivo fitness landscapes of the virus. The viral populations evolved at markedly different rates in different regions of the genome, while linkage along the genome extended to only a few hundred bases during chronic infection. We show that the viral population rapidly reverted to a consensus-like genotype at sites that previously underwent immune escape. The rate of reversion depended strongly on conservation (a proxy for fitness cost) in global alignments of HIV genomes. We further estimated the in-vivo fitness landscape of almost the entire viral genome from minor variation in deep sequencing data at single-base resolution.

In addition to samples from untreated infection, we sequenced proviral DNA from blood of these same patients after many years of suppressive therapy. We found that proviral DNA in these cells derived from viruses replicating shortly before the start of therapy with no evidence of persistent evolution during suppressive therapy.

O-03-IH02

Quantifying the evolutionary dynamics of tumor progression and metastasis

Christina Curtis[1, 2]

[1]Stanford University School of Medicine (United States), [2]Stanford University (United States)

Cancer results from the acquisition of somatic alterations in an evolutionary process that typically occurs over many years, much of which is occult. Understanding the evolutionary dynamics that are operative at different stages of progression in individual tumors might inform the earlier detection, diagnosis, and treatment of cancer. Although these processes cannot be directly observed, the resultant spatiotemporal patterns of genetic variation amongst tumor cells encode their evolutionary histories. Whereas it has traditionally been assumed that tumor progression results from ongoing sequential selection for driver mutations that confer a stringent fitness advantage, recently, we described a Big Bang model of tumor evolution, wherein after transformation, the tumor grows as a terminal expansion populated by numerous heterogeneous and effectively equally fit subclones. This new model is compatible with effectively neutral tumor evolution and explains the origins of intra-tumor heterogeneity and the dynamics of tumor growth with implications for earlier detection, treatment resistance and metastasis. Building on these findings, I will discuss the importance of accounting for tumor spatial structure when inferring clonal dynamics and describe an extensible framework to simulate spatial tumor growth under varied levels of selection with implications for defining the mode of evolution in diverse solid tumors. Lastly, I will show describe a computational framework to infer the timing of metastatic dissemination from paired primary and metastatic patient samples, revealing fundamentally new insights into this lethal process.

O-03-IH03

The impact of biodiversity on phage immunity and virulence in Pseudomonas aeruginosa

Ellinor Opsal Alseth[1], Ellie Pursey[1], Angus Buckling[1], Edze Westra[1]

[1]University of Exeter (United Kingdom)

Pseudomonas aeruginosa is an opportunistic pathogen known for its ability to rapidly evolve antibiotics resistance, and is as such increasingly becoming a target of clinical phage therapy trials. However, the rapid evolution of phage resistance mechanisms remains an issue, with approximately half of all *P. aeruginosa* clinical isolates possessing adaptive CRISPR-Cas immune systems to combat viral infection. While the widespread nature of CRISPR systems suggest they are important in clinical settings, lab based studies often find that *P. aeruginosa* almost exclusively evolves surface based resistance. This discrepancy may be explained by differences in the biotic and abiotic environment between *in vitro* and *in vivo* environments. In particular, while *P. aeruginosa* is typically examined in isolation when grown in the lab, during clinical infections *P. aeruginosa* usually coexists with a polymicrobial community of other pathogens. Here, we report how an artificial cystic fibrosis microbial community, consisting of *Staphylococcus aureus, Burkholderia cepacia complex* and *Acinetobacter baumannii*, drives the evolution of phage resistance in *P. aeruginosa* PA14. Using a *Galleria mellonella* infection model, we also show that the evolution of CRISPR-based resistance results in the maintenance of virulence on par with the ancestral, while the evolution of surface based resistance leads to reduced virulence *in vivo*. Collectively, our analyses demonstrate how intra-host biodiversity might propagate the evolution of CRISPR, and that the type of resistance mechanism evolved has important implications for *P. aeruginosa* virulence.

O-03-IH04

Selection and clonal interference in B-cell repertoire response to HIV-1 infection

Armita Nourmohammad[1, 2], Marta Luksza[4], Thierry Mora[6], Jakub Otwinowski[3], Aleksandra Walczak[5]

[1]Max Planck Institute for Dynamics and Self-organization (Germany), [2]University of Washington (United States), [3]University of Pennsylvania (United States), [4]Institute for Advanced Study (United States), [5]CNRS/ENS (France), [6]CNRS/ENS (France)

Chronic pathogens, such as HIV, are able to persist in a host for extended periods of time. Upon infection, the host immune system and virus engage in a coevolutionary arms race: B-cell receptors diversify to neutralize the virus and HIV evolves to evade the immune response. While it is clear that HIV exerts strong selection on the adaptive immune system, the modes of immune response are still unknown. Here, we characterize the dynamics of B-cell repertoires in HIV-1 patients from longitudinal high-throughput B-cell receptor sequences. We identify clonal B-cell lineages and infer evolutionary modes of immune response from the structure of lineages and history of hypermutations in B-cell receptors. We find at least 20% of amino acid changes to be beneficial in pathogen-engaging CDR regions in contrast to only 8% beneficial changes in structurally relevant framework regions. Repertoire response to viral expansion following treatment interruption drives competition among B-cells, leading to strong clonal interference among beneficial CDR3 mutants. We argue that rapid affinity maturation upon viral expansion and a quasi-stationary response during chronic infection characterize the B-cell response to HIV-1.

O-03-IH05

Vibrio cholerae genomic diversity within and between patients

Ines Levade[1], Yves Terrat[1], Jean-Baptiste Leducq[1], Ana A. Weil[2], Jason Harris B.[2], Regina C. LaRocque[2], Firdausi Qadri[3], Jesse B. Shapiro[1]

[1]University of Montreal (Canada), [2]Massachusetts General Hospital (United States), [3]International Centre for Diarrhoeal Disease Research (Bangladesh)

Cholera is a severe, waterborne diarrheal disease caused by toxin-producing strains of the bacterium *Vibrio cholerae*. Comparative genomics has revealed "waves" of cholera transmission and evolution, in which successful clones are successively replaced over decades and centuries. However, the extent of *V. cholerae* genetic diversity within an individual patient is poorly understood. Here we characterized *V. cholerae* genomic diversity at a micro-epidemiologic level: within and between individual patients from Bangladesh and Haiti. We isolated multiple (8-20) *V. cholerae* colonies from each of eight patients, sequenced their genomes and identified SNPs and gene gain/loss events. As expected for an acute infection, we found low but detectable SNP diversity within hosts (zero to three SNPs within each patient), most of which appeared to be selectively neutral, with the exception of one putatively selected mutation in a gene encoding for a histidine kinase, which seems to impact biofilm formation. In contrast with the low SNP diversity, we observed higher levels of gene content variation (5-103 variable genes per patient), due mainly to phages, integrative conjugative elements and plasmids segregating within and among hosts. Most of these genes are exchanged among *V. cholerae* strains, but some appear to have been acquired by horizontal transfer from other vibrios, or other members of the gut microbiome. Overall, we show that gene transfer is a more common source of variation than point mutations in the *V. cholerae*genome, but rare point mutations can be targets of natural selection and a source of phenotypic variation within hosts.

O-03-IH06

Within-host evolutionary dynamics of dengue virus in its mosquito vector Aedes aegypti

Sebastian Lequime[1, 2, 3], Vaea Richard[4], Albin Fontaine[2, 3, 5], Meriadeg Ar Gouilh[6, 7], Isabelle Moltini-Conclois [2, 3], Van-Mai Cao-Lormeau [4], Louis Lambrechts[2, 3]

[1]KU Leuven - University of Leuven (Belgium), [2]Institut Pasteur (France), [3]Institut Pasteur (France), [4]Institut Louis Malarde (French Polynesia), [5]Institut de Recherche Biomedicale des Armees (France), [6]Institut Pasteur (France), [7]Universite de Caen Normandie (France)

Dengue viruses (DENV), as other RNA viruses, evolve in their hosts as a diverse population of variants, referred to as quasispecies. This 'swarm' of variants has been shown to be critical for RNA viruses' fitness, pathogenesis and adaptive potential. For arboviruses, like DENV, this variant swarm can explore the 'genetic space' and allows simultaneous maintenance in the population of genetic variants adapted to their different hosts. The maintenance of genetic diversity is therefore crucial for long-term arbovirus transmission.

As a decisive host for transmission, mosquitoes can influence DENV genetic diversity. However, only a few studies have explored the dynamics of within-host variants upon mosquito infection. Our work took advantage of high-throughput sequencing to study DENV-1 within-host genetic diversity upon infection of *Aedes aegypti* tissues such as midgut and salivary glands, and in the saliva.

Our results suggest that the overall within-host genetic diversity is not dramatically changed upon a single passage in the mosquito but is clearly shaped by genetic drift, due to a population bottleneck during initial infection of the digestive tract, and purifying selection. Moreover, we identified significant modulatory effect of mosquito genotype on the breadth of within-host viral genetic diversity in the mosquito tissues, indicating that DENV genetic diversity, and therefore its fitness and evolution, is inextricably linked to its vector genetic variation.

By identifying these evolutionary forces shaping DENV population within its vector, our work brings to light new aspects of vector-virus interactions' underlying processes and its impact on DENV evolution.

SY17: Linking the mitochondrial genotype to phenotype: a complex endeavour (July 11, 12:30–14:30)

O-03-LM01

The function of genomes in bioenergetic organelles

John Allen[1]

[1]University College London (United Kingdom)

Mitochondria and chloroplasts are subcellular bioenergetic organelles with their own genomes and genetic systems. DNA replication and its transmission to daughter organelles produces cytoplasmic inheritance of characters associated with electron transport and energy coupling. The prokaryotic ancestors of mitochondria and chloroplasts were endosymbionts whose genes became copied to the genomes of their cellular hosts. These copies gave rise to nuclear chromosomal genes that encode precursor proteins that are synthesized in the cytosol for import into the organelle into which the endosymbiont evolved. So what accounts for retention of organellar genes for the complete synthesis within mitochondria and chloroplasts of a tiny minority of their protein subunits? One hypothesis is that expression of genes for protein subunits of energy-transducing enzymes must respond to environmental change by means of direct and unconditional regulatory control - control exerted by change in the redox state of the corresponding gene product. This hypothesis proposes that, to preserve function, an entire redox regulatory system is retained within the original membrane-bound compartment. Co-location of gene and gene product for Redox Regulation of gene expression (CoRR) is an hypothesis in agreement with the results of experiments designed to test it and that seem to have no other satisfactory explanation. I present evidence relating to the CoRR hypothesis, and consider mechanisms of its predicted redox regulation. I discuss the development, conclusions, and implications of the CoRR hypothesis, and identify predictions concerning the results of experiments that may yet prove it to be incorrect.

O-03-LM02

Transmission of mitochondrial heteroplasmy across multigenerational pedigrees

Kateryna Makova[1], Arslan Zaidi[1], Peter Wilton[2], Kate Anthony[1], Barbara Arbeithuber[1], Marcia Su[1], Anton Nekrutenko[1], Rasmus Nielsen[2]

[1]Penn State University (United States), [2]University of California Berkeley (United States)

Mutations in mitochondrial DNA, mtDNA, lead to heteroplasmies, the presence of more than one allele at a locus in an individual. mtDNA undergoes a bottleneck during oogenesis that can lead to fixation, in children, of deleterious variants that are present at nonpathogenic frequencies in the maternal germline. It is important to investigate mother to child mtDNA transmissions, and to estimate the germline bottleneck size. Recently we developed a population-genetic framework for modeling mitochondrial heteroplasmy as a process that occurs on an ontogenetic phylogeny, with genetic drift and mutation changing heteroplasmy frequencies during the various developmental processes represented in the phylogeny. Applying the model to previously published heteroplasmy frequency data, we demonstrated a severe effective germline bottleneck comprised of the cumulative genetic drift occurring between the divergence of germline and somatic cells in the mother and the separation of germ layers in the offspring. Additionally, we demonstrated that the two somatic tissues analyzed undergo bottlenecks during embryogenesis, less severe than the effective germline bottleneck, and experience little additional genetic drift during adulthood. Using this novel method, we have analyzed heteroplasmies in high coverage mitochondrial genome sequences from buccal and blood tissues of 363 individuals from 101 multigenerational families, including 212 transgenerational transmissions. The number of heteroplasmies carried by a child is significantly associated with maternal age at the time of giving birth, suggesting that older mothers transmit more mutations to their children. Our results demonstrate the effects of maternal age and germline bottleneck on mtDNA mutational load.

O-03-LM03

Role of competition and Ne in the maintenance of heteroplasmic, selfishly acting mitochondrial mutations in Caenorhabditis elegans

Joseph Dubie[1], Vaishali Katju[1], Ulfar Bergthorsson[1]

[1] Texas A&M University (United States)

It is not well understood how deleterious mitochondrial mutations increase in frequency. One mechanism that has been proposed is selfishness. Under this model mutant mitochondrial DNA increase their frequency within a host cell despite being detrimental to host fitness by taking advantage of some form of replicative advantage.

The Katju lab has spontaneously generated a variety of seemingly selfish mitochondrial mutations at relatively high frequencies within the model organism *Caenorhabditis elegans* during a long term experimental evolution experiment. We then isolated the mitochondrial mutations in a wild-type N2 nuclear background to remove any nuclear mutations. This system has provided us with a unique tool to experimentally test what forces drive the evolution and inheritance of mitochondrial mutations.

Because these mutations arose in populations with low N_e, these mutations may only reach high frequency when selection is relaxed and competition is not present. To determine if this was the case we competed heteroplasmic experimental worm lines with wild-type N2 worms and maintained separate populations of mutant worm lines at large N_e. To determine the effect that competition has on the frequency of heteroplasmy in each population the frequency of heteroplasmic worms and individual worm level of heteroplasmy was assessed each generation. We found that competition with wild-type worms was enough to purge the populations of worms harboring mitochondrial mutations but that in populations of heteroplasmic worms with large Ne the presence of heteroplasmy was maintained.

O-03-LM04

Direct Estimates of Mitochondrial Mutation Rates Across Genotypes and Populations in Daphnia

Fenner Macrae[1], Leigh Latta IV[2], Maia Benner[1], Dieter Ebert[3], Sarah Schaack[1]

[1]Reed College (United States), [2]Lewis-Clark State College (United States), [3]Universitat Basel (Switzerland)

There are very few species for which there are direct estimates of mitochondrial (mtDNA) mutation rates, and in those cases where estimates have been reported, they are usually derived from experiments using a single genotype. Despite this paucity of data, the range of rates reported across eukaryotes is wide, and there is now interest in understanding whether such variation can be observed even among closely related species or among genotypes within a species. Here, we report direct estimates for the mtDNA mutation rate for 9 genotypes of *Daphnia magna* from three geographic regions. We not only observe remarkable levels of variation among genotypes, the rate estimates (on average) for *D. magna* are much higher than those published previously for the congener *D. pulex*, which (when first reported) were the highest observed in animals thus far. We discuss the implications of mtDNA rate variation, as well as potential explanations for the differences in estimates for *D. magna* and *D. pulex* based on methodological differences. We also discuss challenges for accurately estimating mtDNA mutation rates, including several important unknowns: the size of bottlenecks determining effective population sizes for the mtDNA genome at fertilization, numbers of mtDNA and mitochondria determining the likelihood of fixation for heteroplasmic mutations, and unknown rates of somatic mutation compounding those that occur in germ cells.

O-03-LM05

Mitochondrial genomics of exceptional longevity in bats

David Jebb[1], Nicole M Foley[1], Conor V Whelan[1], Gareth Jones[2], Roger D Ransome[2], Gerald Kerth[3], Sebastien J Puechmaille[1, 3], Emma C Teeling[1]

[1]University College Dublin (Ireland), [2]University of Bristol (United Kingdom), [3]University of Greifswald (Germany)

Bats are exceptional among mammals. They are the only mammals capable of powered flight and the majority echolocate. Astonishingly, bats can live up to almost ten times longer than expected given their body size and metabolic rate (Myotis brandtii: max. lifespan 41 years, ~7g). This defies the Free Radical Theory of Ageing and other mitochondrial theories of ageing. These theories posit that reduced mitochondrial quality through time, primarily driven by mitogenome mutations, or heteroplasmies, leads to the progressive ageing phenotype. Next generation sequencing technologies have found increased heteroplamies with age in humans. The exceptional longevity phenotype of bats suggests they may have evolved enhanced mitochondrial quality control possibly due to the energetic demands of flight. Deep sequencing of mitogenomes from 3 different populations and species of long-lived bats, from two tissues, found similar numbers of heteroplasmies as those reported in humans. Interestingly, there was no significant increase with age in any tissue or species. Unique longitudinal sampling from recaptured individuals over 4 years showed heteroplasmy is dynamic, with individuals gaining and losing heteroplasmies through time. This analysis was coupled with phylogenetic selection analyses of ~1500 nuclear encoded mitochondrial proteins, looking at the ancestral bat lineage, and two long lived genera. We found no evidence for selection upon mitochondrial quality control genes at the base of all bats. Rather, the two long lived genera of interest exhibited separate adaptations in distinct quality control pathways, suggesting convergent evolution of enhanced quality control at the phenotype but not the genotype level.

O-03-LM06

Climate-related Mitochondrial Lineages Correlate with Functional Differences in Energy Utilisation

Stephanie Falk[1], Paul Sunnucks[1], Alexandra Pavlova[1], Craig White[1], Antoine Stier[2]

[1]Monash University (Australia), [2]Glasgow University (United Kingdom)

Oxidative phosphorylation (OXPHOS) provides >90% of available chemical energy (ATP) in most eukaryotic cells. OXPHOS complexes are encoded jointly by nuclear and mitochondrial genes, so are potentially subject to mito-nuclear evolutionary interactions. ATP synthesis via OXPHOS can be conducted efficiently with the low release of heat, or less efficiently with more heat produced. The balance of heat-to-ATP production is a possible mechanism of climate adaptation because in hot environments limitation of metabolic heat can prevent heat damage, whereas in cold environments a higher heat-to-ATP ratio can be beneficial for warming and reduction of oxidative stress.

The Eastern Yellow Robin is an excellent system to test hypotheses regarding climate adaptation via mito-nuclear interactions. Its two divergent mitochondrial lineages (mitolineages) occur in different climatic environments and show divergent selection in some mtDNA-encoded OXPHOS genes that potentially convey physiological differences in energy expenditure.

We tested if birds of the inland mitolineage, which occur in variable climates with hotter summers, manipulate their heat-to-ATP ratio differently compared to birds of the coastal mitolineage, which inhabit more stable climates with cooler summers.

Preliminary analysis shows large seasonal changes for both mitolineages in OXPHOS activity and the heat-to-ATP ratio. OXPHOS activity is explored in several states by altering which complexes are active. For most states, the mitolineages are more distinct in colder months and converge in warmer months. Furthermore, phosphorylating OXPHOS states show sex-specific differences.

Thus, mitolineages and sexes optimise energy utilisation differently, which might be caused by mitochondrial genotypic variation.

SY18: Looking beyond the genome: cultural and behavioral drivers of biological evolution (July 11, 9:30–11:30)

O-03-LB01

Cultural traits that shape genetic diversity: a case study in Inner Asia

Evelyne Heyer[1, 2], Laure Segurel[2], Nina Marchi[1], Tatyana Hegay[3]

[1]National Natural History Museum (France), [2]CNRS (France), [3]Ministry of Public Health (Uzbekistan)

Inner Asia, by its central location in the Eurasian heartland, is a migratory crossroads of human populations and represents a region where many ethnic groups with different languages and social organizations co-exist. We took advantage of this cultural diversity to assess the influence of different cultural traits on the genetic structure of these populations.

First, we aim to get insights into the sex-specific cultural traits that shape genetic diversity by analyzing uniparental markers. We show that patrilineality, patrilocality, and the cultural transmission of reproductive success leave particular detectable signatures on the sex-specific genetic structure.

We then explore the hypothesis that migration can be a strategy to limit inbreeding. We collected both ethnological and genomic data in 16 populations across Inner Asia. Based on genome-wide SNP data, we estimated individuals and populations inbreeding level, and based on ethnological data, we estimated geographical exogamy for each couple. We found that the population social organization) correlates to differences in dispersal behaviors, but the exogamy rates does not affect the intra-population genetic diversity. Surprisingly, when focusing on the individual level, we showed that descendants from exogamous couples (4<d

In conclusion, Inner Asia is a good illustration of how cultural behaviors, notably matrimonial rules, can impact the genetic diversity of our species.

O-03-LB02

The evolutionary consequences of sociality and culture: is there gene-culture interaction in non-human animals?

Susanne Shultz[1]

[1]University of Manchester (United Kingdom)

The evolution of complex animal societies represents a major evolutionary transition. At its essence, sociality involves introducing hierarchical structuring of individuals within populations. However, the nature of this structuring varies widely, from loose and ephemeral aggregations, to bonded pairs of reproductive individuals, to multi-level societies. Although social organisation across different taxonomic groups appears analogous, whether the selective pressures and evolutionary pathways leading to social living are universal across groups is unknown. Here we use Bayesian phylogenetic models to reconstruct these evolutionary transitions to different forms of sociality in mammals. It is striking that stable structured sociality is far from an evolutionary labile trait and suggests strong there is evolutionary inertia. Structured sociality is also associated with life history shifts, brain size evolution and the emergence of cultural and behavioural richness in multiple groups. That the most likely transitions into social living vary across taxonomic groups suggests that sociality may solve different ecological problems and the search for universal explanations may be misguided. Our analyses provide insight into not only the transitions to sociality but as the likely timing of these switches and mechanisms underpining them across different mammalian orders.

Invited speaker but no option to upload c.v.

O-03-LB03

Ancient Genomic Diversity Reveals Differences in Cultural Practices and Cultural Barriers between Prehistoric Farmers and Hunter-gatherers in Europe

Zuzana Hofmanova[1, 2], Vivian Link[1, 2], Ilektra Schulz[1, 2], Jens Bloecher[3], Laura Winkelbach[3], Sofija Stefanovic[4, 5], Joachim Burger[3], Daniel Wegmann[1, 2]

[1]University of Fribourg (Switzerland), [2]Fribourg (Switzerland), [3]Johannes Gutenberg University Mainz (Germany), [4]University of Belgrade (Serbia), [5]University of Novi Sad (Serbia)

Humans differ from most other species in that we create our own ecological niche. Culture has thus shaped human genetic variation over millennia. While surprisingly little is known about prehistoric cultural practices, there is vested hope that patterns of ancient genetic diversity will elucidate how past societies were organized and interacted with each other. Yet such inferences remain challenging due to generally low numbers of individuals and especially the lack of population-level samples. Here we present novel samples from the region of the Danube Gorges (Balkans), located in the heart of the migration corridor through which farming was brought from Anatolia to Central Europe. Our archaeologically well-defined samples (~10000-5500 calBC) represent multiple closeby-settlements of a sedentary society before and during Neolithisation. Contrasting population-genomic and cultural affinities of our samples revealed that settlements differed strikingly in their interaction with immigrating farmers: while some exhibited strong barriers to gene flow, others incorporated multiple individuals of genetic ancestry common to Aegean farmers. To elucidate important aspects of social practices before, during and after this demographic shift, we accurately inferred within and between individual genetic diversity of our population sample by sequencing either whole genomes or many putatively neutral regions, and by using novel methods that account for *post-mortem* damage and the heterochronous nature of our reference panel. Notably, we found a lower within-individual diversity as well as a lower X to autosomes diversity in hunter-gatherers than farmers prior to their contact, consistent with an elevated population size and stronger patrilocality in farmers.

O-03-LB04

Complex human histories of Northeast Asia revealed by correlations between genes, language, and music

Hiromi Matsumae[1, 2], Patrick E Savage[3], Balthasar Bickel[4], Thomas E Currie[6], Takahiro Sato[5], Atsushi Tajima[5], Mark Stoneking[7], Kentaro K Shimizu[1, 2], Matt Gillan[8], Steven Brown[9], Hiroki Oota[10]

[1]Yokohama City University (Japan), [2]University of Zurich (Switzerland), [3]Keio University (Japan), [4]University of Zurich (Switzerland), [5]Kanazawa University (Japan), [6]University of Exeter (United Kingdom), [7]Max Planck Institute for Evolutionary Anthropology (Germany), [8]International Christian University (Japan), [9]Neuroscience and Behaviour, McMaster University (Canada), [10]Kitasato University (Switzerland)

Culture evolves in ways that are analogous to, but distinct from, genetic evolution. Previous studies have shown that different genetic loci and different aspects of culture have different evolutionary histories, but few studies have empirically investigated parallels between genetic and cultural evolution using a diverse range of genetic and cultural markers. Here we report the first analysis comparing multiple cultural (language and music) and genetic (autosomal SNPs, mtDNA, Y-STR) markers, using a diverse sample of 13 Northeast Asian populations for which all types of data were available. First, we constructed distance matrices for language (grammar, phonology [AUTOTYP], lexicon [ASJP]), music (song structure [CantoCore], performance style [Cantometrics]), and genes (SNPs, mtDNA, Y-STR). We used these distance matrices to test for correlations among distance matrices and visualized the relationships using network diagrams. Significant correlations between language-genes and language-music were observed, but not for music-genes - a result different from the only previous study to investigate gene-language-music correlations. Exploratory analyses of different sub-types showed that the correlations with language were driven by grammar, which was significantly correlated with both musical markers and with SNPs. Our results suggest that grammar might be one of the strongest cultural indicators to infer human population history, while also demonstrating divergences among different cultural and genetic markers that highlight the complex nature of human history.

O-03-LB05

The Genomics of Megaliths: An Irish case study into the reconstruction of prehistoric societal landscapes through ancient DNA analysis

Lara M Cassidy[1]

[1]Trinity College Dublin (Ireland)

The Irish Neolithic (circa 3,800-2,500) marks the emergence of complex civilization on the island, alongside the establishment of continued contacts with other Atlantic regions, which intensify in the succeeding Copper and Bronze Ages. In addition to these cultural upheavals, the Neolithic period has been demonstrated to both begin and end with mass migration into the island, potentially from multiple external sources. However, the variable interplay between geography and culture in the catalyzation of these population movements has remained an open question. Indeed, the archaeological record would suggest regional heterogeneity in the uptake of British and continental traditions at both transition points.

Here, the potential social and cultural implications of such events are explored through the prism of ancient genomics. Imputed diploid genotypes for over 50 individuals sampled from the Mesolithic to Bronze Age periods, encompassing a diversity of megalithic structures, are presented and dissected through the use of haplotypic-sharing methods, as well as estimations of kinship and inbreeding. Combined with Y chromosome analysis these provide the first evidence of genetic structure on the island during specific prehistoric time intervals, which can be interpreted along both geographical and cultural lines. Furthermore, candidate refugiums that may recurrently act as reservoirs for older traditions and genetic ancestries are identified, as well as hub regions, which appear more susceptible to demographic disturbances on the continent, highlighting the immovable constraints of geography on both cultural and genomic evolution.

O-03-LB06

Waves of history in Remote Oceania: language continuity despite population replacement in Vanuatu

Cosimo Posth[1], Kathrin Naegele[1], Heidi Colleran[2], Frederique Valentin[3], Stuart Bedford[4], Kaitip W Kami[2, 5], Johannes Moser[6, 7], Julia Gresky[7], Fiona Petchey[8], Patrick Roberts[9], Choongwon Jeong[1], Russell Gray[2], Johannes Krause[1], Adam Powell[1, 2]

[1]Max Planck Institute for the Science of Human History (Germany), [2]Max Planck Institute for the Science of Human History (Germany), [3]CNRS (France), [4]The Australian National University (Australia), [5]Port-Vila (Vanuatu), [6]German Archaeological Institute (Germany), [7]German Archaeological Institute (Germany), [8]The University of Waikato (New Zealand), [9]Max Planck Institute for the Science of Human History (Germany)

Recent genomic analyses show that the earliest peoples reaching Remote Oceania - associated with Austronesian-speaking Lapita culture - were almost completely East Asian. Yet, Papuan-related genetic ancestry is found across all present-day Pacific populations, indicating that peoples from Near Oceania have played a significant - but largely unknown - ancestral role. Here, 19 new ancient genomes from the Southwest Pacific provide the first direct evidence of a so-far undescribed Papuan expansion into Remote Oceania starting at ~2,500 years before present, far earlier than previously suggested. Additional genome-wide data from 27 contemporary ni-Vanuatu demonstrate a subsequent and almost complete turnover of the Lapita-related population by Near Oceania ancestry. Despite this massive demographic change, incoming Papuan languages did not replace Austronesian languages. Our analyses show that rather than one large-scale event, this process was incremental and complex, comprising repeated migrations and sex-biased interactions with populations from the Bismarck Archipelago. This study thus provides a compelling explanation for a population replacement that retains the original languages, a phenomenon that is extremely rare - if not unprecedented - in human history.

SY19: Machine learning approaches in evolutionary genetics (July 10, 13:30–15:30)

O-02-ML01

Supervised learning for analyzing large-scale genome-wide DNA polymorphism data

Ziqian Hao[1], Haipeng Li[1]

[1]Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences (China)

Supervised learning has been extensively applied in many fields; Alpha-GO and autopilot might be two of the most well-known cases. However, its application in population and evolutionary genetics is still in childhood. Recently, we introduced the boosting, a supervised learning approach, to identify positive Darwinian selection in natural populations and estimate recombination rate along the human genome. We further analyzed the genome-wide DNA polymorphism data from nearly 10,000 human individuals (UK10K) and obtained a fine-scale genetic map for humans. These results indicate that supervised learning approaches, together with deep learning and reinforced learning, could play essential roles when analyzing large-scale genome-wide DNA polymorphism data.

SY19: Machine learning approaches in evolutionary genetics (July 10, 13:30–15:30)

O-02-ML02

New methods for measuring natural selection and predicting deleterious variants in the human genome.

Adam Siepel[1]

[1]Cold Spring Harbor Laboratory (United States)

Many genetic variants that influence phenotypes of interest are located outside of protein-coding genes, yet existing methods for identifying such variants have poor predictive power. I will describe a new computational method, called LINSIGHT, that substantially improves the prediction of noncoding nucleotide sites at which mutations are likely to have deleterious fitness consequences, and which, therefore, are likely to be phenotypically important. LINSIGHT combines a generalized linear model for functional genomic data with a probabilistic model of molecular evolution. The method is fast and highly scalable, enabling it to exploit the 'big data' available in modern genomics. I will show that LINSIGHT outperforms the best available methods in identifying human noncoding variants associated with inherited diseases. In addition, I will describe an application of LINSIGHT to an atlas of human enhancers and show that the fitness consequences at enhancers depend on cell type, tissue specificity, and constraints at associated promoters. Finally, I will describe an extension of LINSIGHT that considers the full site frequency spectrum and allows for the estimation of position- and allele-specific selection coefficients. So far, we have applied this method to coding sequences in the human genome, where it reveals surprisingly strong selection on synonymous sites, classes of genes that have undergone relaxed and enhanced selection in recent human evolution, and other aspects of natural selection on coding regions. Work is underway to extend this method to noncoding regions.

O-02-ML03

Real-time Phenotype Prediction From Unaligned Whole Genome Sequencing Data Using Deep Learning

Will Landecker[1], Samuel V Scarpino[2, 3, 4]

[1]patho.me (United States), [2] Northeastern University (United States), [3] Northeastern University (United States), [4] Northeastern University (United States)

Whole genome sequencing methods typically generate large, heterogeneous datasets, which must be heavily processed using established bioinformatic pipelines. As a result, and despite a greater availability of data, using genomic data to predict phenotypes in real-time and without the use of dedicated computational resources remains a challenge. Specifically, traditional statistical or algorithmic tools have limited application when analyzing multiple, large and non-linear data sets from genetic surveillance activities, while in the field, or for use in clinical settings and often require sequence data that have been assembled to a reference genome and aligned. Convolutional Neural Networks (CNN), known colloquially as "deep learning", have recently emerged as a solution to analyzing similarly large and complex datasets. Here, we demonstrate that a CNN model can predict influenza antiviral resistance with $>99¥%$ out-of-sample accuracy using either hemagglutinin (HA) or neuraminidase (NA) sequences linked to antiviral resistance data. Critically, the model is able to accurately predict this antiviral resistance phenotype from genomic data in an on-line manner and without positional information coming from an alignment. We also demonstrate how cloud-based storage and computational resources can facilitate real-time phenotype prediction without dedicated computers. This CNN toolkit is now deployed at the US CDC and we anticipate that it will increase the utility of genomic data in molecular surveillance, field research, and real-time clinical diagnostics. Lastly, because the structure of resulting CNNs also recapitulates the evolutionary trajectory of organisms, these methods provide an exciting mechanism for future research into reticulate and non-linear evolution.

SY19: Machine learning approaches in evolutionary genetics (July 10, 13:30–15:30)

O-02-ML04

A Machine-Learning Approach for Phylogenetic Model Selection

Shiran Abadi[1], Dana Azouri[1], Tal Pupko[2], Itay Mayrose[1]

[1]Tel-Aviv University (Israel), [2]Tel-Aviv University (Israel)

Over the years, a plethora of DNA substitution models have been developed, which serve the basis for phylogeny reconstruction. Choosing an inadequate model may lead to wrong phylogenetic inference and inaccurate deductions regarding the evolutionary processes. Current methods for model selection, whether performed within the maximum likelihood or Bayesian inference paradigms, frequently result in contrasting selections of the best fitting model. In this work, we conduct an in-depth analysis of widely-used model selection criteria (AIC, AICc, BIC, DT, hLRT, and Bayes Factor). First, our analysis over 1500 empirical alignments reveals massive incongruencies between the criteria. Second, examination of thousands of simulated alignments demonstrates that the true model is recovered in ~50% of the datasets by all criteria. Yet, whereas hLRT and BIC tend to select simpler models and are thus highly accurate when the generating model is simple, the accuracy of AIC is moderate for all models and exceeds the other criteria when the generating model is complex. Further analysis demonstrates that using different criteria may lead to conflicting inferences. We present an alternative approach for model selection within the machine-learning framework. In this scheme, each alignment is characterized by features that are used to train a Random Forest classifier for predicting the true evolutionary model. In addition, the learning process can pinpoint the features that are most important for model fitting. Analysis of thousands of empirical alignments demonstrates that the predictions made by the proposed strategy are similar to AIC, yet with substantial improvement in running time.

O-02-ML05

Statistical inference frameworks for detecting adaptive evolution of variants and genes

Lauren A Sugden[1, 2], Elizabeth G Atkinson[3], Annie P Fischer[4], Stephen Rong[1, 2], Brenna M Henn[3], Sohini Ramachandran[1, 2]

[1]Brown University (United States), [2]Brown University (United States), [3]Stony Brook University (United States), [4]Brown University (United States)

Identifying evidence for selection in the human genome is a longstanding challenge in evolutionary biology, and requires robust statistical methodology. Here we introduce two probabilistic frameworks that provide insight into adaptation at the level of variants, genes, and pathways.

First, we describe a supervised composite classification framework, SWIF(r), that explicitly models joint distributions of chosen selection statistics under adaptive scenarios of interest, returning interpretable posterior probabilities for each genomic site. In simulations and in 1000 Genomes Project data, SWIF(r) outperforms other composite methods at identification of regions and precise variants undergoing hard sweeps. We apply SWIF(r) to genotype data from the Khomani San of southern Africa, an understudied population representing the most basal human population divergence, and find an overrepresentation of adaptive variants in genes associated with metabolism and obesity. In many cases, SWIF(r) signals tag nearby, highly differentiated, functional variants. This multi-gene signature suggests that fat storage has played an important role in the evolution of Khomani San hunter-gatherers.

Second, we develop a hidden Markov model (HMM) designed to detect evidence for selection at the genic level. With hidden states corresponding to adaptive, linked neutral, and unlinked neutral sites, we leverage stochastic backtrace to sample thousands of probabilistically representative paths through the HMM. By calculating the fraction of these paths that pass through the adaptive state, we can calculate gene-level adaptive probabilities. These probabilities facilitate the identification of polygenic selection on genes and pathways, furthering our understanding of the evolution of complex traits.

SY19: Machine learning approaches in evolutionary genetics (July 10, 13:30−15:30)

O-02-ML06

A Likelihood-Free Inference Framework for Population Genetic Data using Permutation-Invariant Neural Networks

Sara Sheehan Mathieson[1], Jeffrey Chan[2], Valerio Perrone[5], Jeffrey P. Spence[4], Paul A. Jenkins[5, 6], Yun S. Song[2, 3]

[1]Swarthmore College (United States), [2]University of California, Berkeley (United States), [3]University of California, Berkeley (United States), [4]University of California, Berkeley (United States), [5]University of Warwick (United Kingdom), [6]University of Warwick (United Kingdom)

Inference in population genetic models is difficult because exact likelihoods are computationally intractable. While this issue is avoided by likelihood-free methods such as approximate Bayesian computation, these approaches currently require handcrafted summary statistics of the data. In complex settings, designing and selecting suitable summary statistics is both challenging and limiting. Valuable information is lost when converting to summaries, new statistics must be selected for different datasets and parameters, and complicated statistics can be a computational bottleneck when large-scale simulations are required.

Convolutional neural networks (CNNs) provide an attractive alternative to summary statistic based methods, as they have been enormously successful in domains with correlated data such as image recognition. However, in population genetics correlation structure exists along the genome but not across haplotypes, which are permutation-invariant.

To overcome these limitations, we developed a permutation-invariant CNN to learn an exchangeable feature representation for "raw" population genetic data. We use this representation in a novel Bayesian likelihood-free inference framework to learn the inverse functional relationship from the data to the posterior. To train our network, we introduce simulation "on-the-fly", where we simulate new data for every training iteration. This helps avoid common deep learning issues such as over-fitting and poorly calibrated posteriors. Finally, we demonstrate the power of our method on the problem of identifying recombination hotspots. After training our network on data simulated under the HapMap recombination map, we show that we can detect hotspots in multiple populations from the 1000 Genomes Project, outperforming the state-of-the-art method LDhot.

SY20: Mapping genotype and phenotype to fitness (July 12, 12:30–15:30)

O-04-MG01

What can we learn from experimental fitness landscapes?

Claudia Bank[1]

[1]Instituto Gulbenkian de Ciencia (Portugal)

Fitness landscapes, which map genotypes or phenotypes to fitness, have developed from a theoretical metaphor into a popular subject of experimental study. The quantification of the shape of fitness landscapes carries the promise to inform us about the nature of adaptation and the predictability of evolution; yet, the true dimensionality of fitness landscapes is so immense that even large data sets can only cover a small area of the total sequence or phenotype space. Thus, the question emerges whether we can use theoretical or statistical approaches to extrapolate from the observed landscape to the surrounding area or even the whole organism's fitness landscape, and what a given experimental landscape can teach us about evolution in natural populations. I will discuss these questions under consideration of various experimental data sets, and highlight the challenges when trying to bridge our theoretical and empirical knowledge of fitness landscapes.

O-04-MG02

An experimental test of the genomic consequences of local adaptation in deer mice

Rowan Barrett[1], Stefan Laurent[2], Riccardo Mallarino[5], Charles Xu[1], Susanne Pfeifer[3], Jean-Marc Lassance[4], Jeffrey Jensen[3], Hopi Hoekstra[4]

[1]McGill University (Canada), [2]Max Planck Institute for Plant Breeding Research (Germany), [3]Arizona State University (United States), [4]Harvard University (United States), [5]Princeton University (United States)

Although natural selection is a deterministic process, the genomic consequences of local adaptation can be hard to predict because connections between genotype, phenotype, and fitness are complex. Here, we show that natural selection can have predictable effects on a specific phenotype and its causative mutation while driving extensive but idiosyncratic allele frequency change across the rest of the genome. We directly estimated the genetic consequences of natural selection on cryptic pigmentation phenotypes of deer mice using a manipulative field experiment replicated in two divergent habitats. We found that selection repeatedly favoured divergent cryptic phenotypes in each habitat. In one habitat this phenotypic selection was accompanied by rapid shifts in allele frequency at a deletion mutation in the *Agouti* gene that we show causes a functional change in pigmentation by altering protein binding. Finally, we use simulations to show that widespread changes in allele frequency across many other regions of the genome are likely to be the result of correlational selection. Our results demonstrate that strong phenotypic selection can have rapid and pervasive effects on patterns of both gene- and genome-scale variation during local adaptation.

O-04-MG03

Inference of changes of HIV-1 gp160 protein fitness landscape from sequence data with single-position resolution

Galya V. Klink[1], Georgii A. Bazykin[1, 2]

[1]Institute for Information Transmission Problems (Kharkevich Institute) of the Russian Academy of Sciences (Russian Federation), [2]Skolkovo Institute of Science and Technology (Russian Federation)

Fitness conferred by the same allele may differ between genotypes, and these differences shape variation and evolution. While changes in amino acid propensities at protein sites have been inferred from sequence data statistically, the existing methods are data-intensive, and aggregate data from multiple sites. Here, we develop a statistical method for detection of individual amino acids that confer different fitness values in different groups of species from combined sequence and phylogenetic data. The method is based on the distribution of homoplasic (convergent or parallel) and divergent substitutions across a phylogeny, and uses the fact that a higher rate of homoplasic evolution at a clade signifies high fitness of the derived allele. Using the phylogeny of the env gene of two distinct HIV-1 subgroups, we show that the inferred fitness changes are consistent with the fitness differences observed in deep mutational scanning experiments.

O-04-MG04

Genotype-Fitness Mapping in Cancer Cell Lines using CRISPR-Cas9

Elizabeth Hutton[1], Xiaoli Wu[1, 2], Yuhan Huang[1], Sofya Polyanskaya[1], Bin Lu[1], Timothy Somerville[1], Yusuke Tarumoto[1], Yali Xu[1, 2], Christopher Vakoc[1], Adam Siepel[1]

[1]Cold Spring Harbor Laboratory (United States), [2]Stonybrook University (United States)

Inference of the fitness of genetic variants often relies upon the existing genetic variation of the population. High-throughput CRISPR-Cas9 screens allow the fitness of any arbitrary variant to be directly evaluated in cell culture, but these assays present unique challenges in signal extraction. Our work examines patterns of cell growth and proliferation in modified cell lines using a novel statistical framework. Our approach allows us to directly test the fitness effects of variants within their endogenous genomic context, identifying essential gene domains and epistasis. Furthermore, these CRISPR-Cas9 screens allow the assessment of fitness effects of extremely deleterious variants which are not present in viable organisms. We have examined over 100 separate high-throughput CRISPR experiments, using our hierarchical framework to capture multiple levels of CRISPR-specific technical noise. Our improved estimation of tissue-specific fitness effects provide information on the extreme end of the fitness distribution, and can be complemented with classical signatures of natural selection for essential pathway analysis, tissue-specific function prediction, and epistasis specific to different cancer types.

O-04-MG05

Uncovering the genotype-phenotype-fitness map of microbes adapting to novel environments

Grant Kinsler[1], Kerry Geiler-Samerotte[1], Dmitri Petrov[1]

[1]Stanford University (United States)

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

O-04-MG06

Population genomic, ecological and physiological roles for the sodium transporter *HKT1* in *Arabidopsis thaliana* populations adapted to fluctuating coastal habitats

Silvia Busoms[1, 2], Charlotte Poschenrieder[2], David Salt[3], Levi Yant[1]

[1]John Innes Centre, Norwich (United Kingdom), [2]Universitat Autonoma de Barcelona (Spain), [3]University of Nottingham (United Kingdom)

Environmental stressors such as soil salinity and dehydration are major constraints on plant growth, causing world-wide crop losses. Compounding these insults, increasing climate volatility requires adaptation to fluctuating conditions. Salinity stress is relatively well understood in *Arabidopsis thaliana*, making this system well-suited for the rapid molecular dissection of evolutionary mechanisms. In a large-scale genomic analysis of Catalonian *A. thaliana*, we resequenced 77 individuals from multiple gradients along the coast and integrated these data with 1135 world-wide *A. thaliana*genomes for a detailed understanding of the demographic dynamics of naturally evolved salinity tolerance. This revealed that Catalonian varieties adapted to highly fluctuating salinity are not Iberian relicts but instead have more recently immigrated to this region. *De novo* genome assembly of three allelic variants of the *high-affinity K^+transporter* (*AtHKT1*) locus resolved dramatic structural variation that appears to underpin this case of fluctuating selection in response to seasonal changes in soil salinity. Plants harbouring the 'weak' (low expression) *AtHKT1* allele were migrants that have moved into areas where soil sodium levels fluctuate dramatically due to geography and rainfall variation. We demonstrate that the proportion of plants harbouring 'weak *AtHKT1* allele correlates with soil sodium level over time, weak allele plants are better adapted to intermediate levels of salinity, and the weak allele locus clusters with other widespread European high-sodium accumulator accessions. Together our evidence indicates that *AtHKT1* appears to be under fluctuating selection in response to climate volatility and is a worldwide determinant in adaptation to saline conditions.

O-04-MG07

Parallelism of genomic response during rapid seasonal adaptation in Drosophila melanogaster

Sharon Ilana Greenblum[1], Alan Bergland[1, 2], Seth Rudman[3], Subhash Rajpurohit[3, 4], Paul Schmidt[3], Dmitri Petrov[1]

[1]Stanford University (United States), [2]University of Virginia (United States), [3]University of Pennsylvania (United States), [4]Ahmedabad University (India)

Evolution experiments in eukaryotic populations have observed predictable population-wide phenotypic shifts following applied selective pressure in <10 generations, primarily due to rapid shifts in the frequency of segregating variants. Less clear is whether similar shifts occur in natural settings, and whether a predictable set of variants underlie specific phenotypic shifts. Understanding the genomic basis of rapid adaptation in natural settings is critical to addressing long-standing questions regarding the maintenance of functional variation. Previously we sampled wild *Drosophila melanogaster* flies from a Pennsylvania orchard and identified thousands of SNPs with yearly allele frequency shifts of ~20% from spring to fall, suggesting a significant portion of the genome is under seasonal selection. Now we have performed similar experiments in a highly replicated, migration-restricted framework. Each spring for 4 successive years we reconstituted an outbred population from inbred lines, and seeded 6-10 replicate populations in large outdoor cages. Allele frequencies were estimated from pools of individuals sampled from each cage over 15 generations. We find significantly elevated levels of 1) per-cage genome-wide allele frequency shifts and 2) parallel inter-cage frequency shifts, compared to simulated neutral recombination in the same populations. We identify regions with high parallelism across all years and base populations, and regions that exhibit background-dependent parallelism. We also use exponentially growing indoor populations to disentangle ecologically- and environmentally- driven adaptation. Our results suggest that low levels of genomic parallelism do underlie seasonal adaptation, a finding that sheds light on the architecture and predictability of adaptive responses to natural environmental challenge.

O-04-MG08

Integrating functional genetics and demographic life history modelling: PERPETUAL FLOWERING 1 pleiotropically regulates flowering and seed traits in Arabis alpina

P. William Hughes[1, 2], Wim J.J. Soppe[1], Maria Albani C.[1, 2]

[1]Max Planck Institute for Plant Breeding Research (Germany), [2]University of Cologne (Germany)

How can pleiotropic regulation of flowering and seed traits constrain optimal perenniality? While demographic life history models predict that perennial species independently optimize flowering and seed traits, and explain the evolution of perenniality as a consequence of selection on these traits, field studies of model Brassicaceae species suggest that many genes pleiotropically regulate both flowering and seed traits. Like its orthologue in *A. thaliana*, *FLOWERING LOCUS C* (*FLC*), *PEP1* regulates flowering in *A. alpina*; however, it also expressed in seeds, and promotes germination at low temperatures. In this work we use data characterizing these phenotypes to quantify the degree to which pleiotropy constrains independent trait optimization, and therefore results in maladaptive life histories.

We test the hypothesis that pleiotropic regulation constrains optimization of flowering and seed traits in *Arabis alpina* in two ways. First, we use natural accessions and transgenic lines of *A. alpina* to evaluate the phenotypic consequences of over- or under-expression of *PEP1* on flowering and seed traits. Second, we use this data to develop a dynamic state variable model to characterize the tradeoffs that allocate resources to competing functions (i.e. flowers and seeds), and use stochastic dynamic programming to assess the effect of pleiotropy on fitness optima.

By phenotyping plants grown in glasshouse and field conditions, we determine that *PEP1* has substantial pleiotropic effects on seed traits, and that these effects may constrain the degree to which flowering traits and seed traits can be optimized.

O-04-MG09

Gene Expression Drives the Evolution of Dominance

Christian D. Huber[1], Arun Durvasula[2], Angela M. Hancock[3], Kirk E. Lohmueller[1, 2, 4]

[1]University of California, Los Angeles (United States), [2]University of California, Los Angeles (United States), [3]Max Planck Institute for Plant Breeding Research (Germany), [4]University of California, Los Angeles (United States)

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

O-04-MG10

Functional genetic variants revealed by massively parallel precise genome editing

Eilon Sharon[1], Shi-An Chen[1], Neil Khosla[1], Justin Smith[1], Jonathan Pritchard[1], Hunter Fraser[1]

[1]Stanford (United States)

A major challenge in genetics is to determine which sequence differences between species, populations, and individuals drive phenotypic and fitness differences. However, current methods of genetic mapping generally have limited resolution. To address this gap, we developed a high-throughput genome editing approach that allows us to introduce tens of thousands of specific genetic variants in a single experiment. We used this approach to study the fitness consequences of 16,006 SNPs and indels that differ between the yeast strains RM and BY. We identified 572 variants with significant fitness differences in glucose media; many of these have large effect sizes, with 171 estimated to affect fitness by >1%. Most of the significant variants likely have regulatory effects, as only 19.2% affect amino acid sequences. The significant variants are highly enriched in promoters, particularly in and near transcription factor binding sites. Nearby variants nearly always favor the same strain, suggesting that lineage-specific selection on particular genes is often driven by multiple variants. Finally, ribosomal gene promoters were highly enriched for variants whose RM alleles increase fitness, suggesting polygenic adaptation of the ribosome. In sum, our genome editing approach enables direct measurement of the fitness effects of thousands of variants in parallel, and could be adapted to test the effects of genetic variation in any screen for cell survival or cell-sortable markers.

O-03-ME02

Episodic convergent evolution drives dynamic history of passage adaptation and vaccine efficacy in the H3N2 influenza virus

Weiwei Zhai[1], Hui Chen[1], Jacob Alvarez[1], Sock Hoon Ng[2], Rasmus Nielsen[3]

[1]Genome Institute of Singapore (Singapore), [2]DSO National Laboratories (Singapore), [3]Univ of California Berkeley (United States)

As a dominant seasonal virus, H3N2 influenza evolves rapidly in humans and posit constant threats to public health. During vaccine production, Influenza viruses are often propagated in a diverse set of culturing media and additional substitutions known as passage adaptation can cause extra evolution in the target strain, leading to ineffective vaccines. Using a powerful probabilistic approach, we were able to analyse an unprecedented amount of H3N2 hemagglutinin sequences (n=32,278) available from the Global Initiative on Sharing All Influenza Data (GISAID) database. We found that passage adaptation in the embryonated eggs is driven by episodic waves of convergent evolution over just 14 codons and passage adaptation is getting progressive stronger in eggs. Based on allelic states at these sites, we developed a metric of adaptive distance which quantifies the strength of passage adaptation and found strong negative correlation between passage adaptation and vaccine efficacy. Our findings highlight the importance of studying adaptive evolution in disease control and shed light on strategies reducing Darwinian evolution for effective vaccines in the coming future.

O-03-ME03

Population-level, Genome Wide Association Study of Burkholderia multivorans in the Cystic Fibrosis Lung Reveals a Role for Recombination in the Evolution of Antimicrobial Resistance

Julio Felix Diaz Caballero[1], Shawn T Clark[2, 3], Pauline W Wang[4], Sylva L Donaldson[4], D Elizabeth Tullis[5], Yvonne CW Yau[3, 6], Valerie J Waters[7], David M Hwang[2, 3, 8], David S Guttman[1, 4]

[1]University of Toronto (Canada), [2]University of Toronto (Canada), [3]University of Toronto (Canada), [4]University of Toronto (Canada), [5]St. Michael's Hospital (Canada), [6]The Hospital for Sick Children (Canada), [7]The Hospital for Sick Children (Canada), [8]University Health Network (Canada)

Cystic fibrosis (CF) lung infections caused by *Burkholderia cepacia* complex species are responsible for high rates of mortality and morbidity. We performed a population genomic study of *B. multivorans* isolates obtained from a CF patient through three stages of infection including the initial incident infection, deep sampling of a one year period of chronic infection, and deep sampling of a post-transplant recolonization. We used evolutionary genomics to reconstruct the evolutionary history of this clone, and a lineage-controlled genome-wide association study (GWAS) approach to identify genetic variants associated with antibiotic resistance. We found that the incident isolate had a much higher level of antimicrobial susceptibility than any other isolate, and the chronic isolates diversified into distinct genetic lineages that showed distinct patterns of antibiotic susceptibility. The post-transplant reinfection isolates did not originate from the chronic population, but are clonal descendants of the incident strain. There were examples of parallel pathoadaptation, in which individual loci, or even the same codon independently mutated multiple times. We employed a GWAS approach to find resistance-associated variants and identified one variant in ampD associated with resistance to β-lactams, and two variants associated with both aminoglycosides and quinolones. Through identification of recombination breaks, we observed that variants linked to β-lactam antibiotics tend to occur in recombinogenic regions. We also observed that parallel pathoadapted loci were overrepresented in recombinant regions. This study illustrates the power of deep, longitudinal coupled with evolutionary and lineage-corrected GWAS analyses to reveal how pathogens adapt to their hosts.

O-03-ME04

Consequences of European arrival on the microbiota of ancient Native Americans

Laura S Weyrich[1], Keith Dobney[2], Lars Fehren-Schmitz[3], Wolfgang Haak[4], Bastien Llamas[1], Alan Cooper[1]

[1]University of Adelaide (Australia), [2]University of Liverpool (United Kingdom), [3]University of California Santa Cruz (United States), [4]Max Planck Institute for the Science of Human History (Germany)

In the New World, European arrival brought many new infectious diseases to Native American populations, including small pox, measles, and cholera. However, the long-term influences on health, especially on the trillions of commensal microorganisms (microbiota) that live within the body, has not yet been explored in the context of European arrival. Here, we sequenced ancient bacterial DNA preserved within calcified dental plaque (calculus) from 172 ancient North and South Americans, as well as nine post-Columbian individuals, to reconstruct ancient oral microbiota within the New World and explore evolutionary changes in commensal microbial species that may have been influenced by European arrival. The oral microbiota of North and South Americans prior to European arrival was remarkably more diverse than that of their contemporaneous European counterparts and fell outside of known ancient and modern European variation, indicating a robust oral microbiota in the Americas that was likely linked to their unique lifestyles and environments. The oral microbiota also remained similar over time in several single locations, despite cultural or regime changes. Post-Columbian oral microbiota diversity also fell within ancient American diversity, suggesting that the immediate arrival and disease spread by Europeans in the Americas had few, rapid effects on the overall structure of oral microbiota. This also suggests that the assimilation of oral microbiota amongst modern Americans of both Native and European descent likely occurred due to modernization after 1850. Ancient dental calculus provides a wealth of information to better understand how past events may have contributed to modern human health.

O-03-ME05

Adaptive landscape and evolutionary history of the multidrug resistant W148 Russian clone

Thierry WIRTH[1], Maxime BARBIER[1], Philip SUPPLY[2], Matthias MERKER[3], Stefan NIEMANN[3]

[1]Natural History Museum - EPHE (France), [2]Institut Pasteur de Lille (France), [3]Research Center Borstel (Germany)

Mycobacterium tuberculosis is a major human pathogen and still belongs, according to the WHO, to the top 10 causes of death in 2015, being the sole microorganism on the list. Yet, all MTBC lineages are not equal. Some of them, especially the Beijing one, harbour strains with relatively high resistance profiles that go through unprecedented waves of propagation. Here, we leveraged a unique, genetically diverse strain collection of the clade B0/W148, also called Russian clone. This is a twig emerging from the Beijing clade that significantly contributes to MDR epidemics in Russia and Eastern Europe. We determined its evolutionary history and the reasons behind its success. Genome sequencing and phylogenetic analyses of 731 isolates, covering the full Eurasian distribution of the clone, highlighted a deep split that confirmed a central Asian origin some 250 years ago. Bayesian demogenetic analyses confirmed the success of W148, revealing two successive expansions, a first in the late 1970s, followed by a second in the late 1980s. The demographic surge and geographic westward spread were accompanied by numerous adaptive mutations that sharply differentiate W148 from its direct progenitor. This translates into a compensatory mutations acquisition without detectable fitness costs and higher transmissibility, as well as sharp antigenic set-up changes, amino-acid replacements in type VII secretion system genes and modulation in DNA damage response. These swift changes are scars of the selective pressures imposed by the host and contribute to our understanding of the adaptive landscape of a successful multidrug-resistant lineage.

O-03-ME06

MMC-ABC: Inference of population genetic parameters from time-sampled allele frequency data in populations with sweepstakes reproduction

Andrew Sackman[1], Jeffrey Jensen[1]

[1]Arizona State University (United States)

Recent advances in sequencing technology have made time-series population genomic data more accessible than ever before, particularly in the fields of ancient, experimental, and clinical genomics. In conjunction with the increasing abundance of time-sampled data, several methods have been developed for estimating selection coefficients from allele frequency trajectories. However, it is unclear whether these methods can be reliably applied to data from populations with highly skewed offspring distributions (so-called sweepstakes reproduction), which are known to violate assumptions of the Wright-Fisher model and the Kingman coalescent. The development of methods applicable to populations with large reproductive skew has particular importance for the study of viral populations and the evolution of drug resistance. We demonstrate that current methods of time-serial population genomic inference perform very poorly when applied to data from such populations. Making use of recent theoretical advances in the field of multiple merger coalescent theory, in conjunction with approximate Bayesian computation, we address this weakness and propose a novel method for the joint characterization of genome-wide effective population size, offspring distribution, and site-specific selection coefficients from time-sampled mutational frequency data.

O-03-ME07

Colliding histories and opposing forces: ancient DNA of vaccination and smallpox

Ana Duggan[1], Jennifer Klunk[2], Ashleigh Porter[3], Anna Dhody[4], Robert Hicks[4], Edward Holmes[3], Hendrik Poinar[1]

[1]McMaster University (Canada), [2]McMaster University (Canada), [3]University of Sydney (Australia), [4]The College of Physicians of Philadelphia (United States)

Smallpox was eradicated less than 200 years after the development of the vaccination process. A disease with significant mortality and morbidity rates ceased to exist, yet we know remarkably little about the viruses that were used for protection against the causative agent of smallpox, variola virus. While variola virus is specific to humans, most of the related viruses in the Orthopoxvirus family are capable of infecting many mammalian species though their reservoirs and host ranges are poorly understood. For most of the 20th century, until its eradication at the end of the 1970s, smallpox vaccination was achieved by inoculation with vaccinia virus, however vaccinia does not appear to have a natural host or reservoir and has been hypothesised to be a recombinant virus inadvertently created through human vaccination processes. Prior to the 20th century, the method and source of smallpox vaccinations remained unstandardized with vaccination material often collected from anyone suffering from an "eruptive skin disease". This project examines strains of Orthopoxviruses associated with smallpox vaccination processes from the 19th century from historical artefacts for genomic identification and contextualization. In sequencing archival vaccine materials, and comparing their genomic composition with historical documents, we use ancient DNA to reconstruct the history of vaccination and clarify the phylogeny of smallpox vaccination-related viruses.

O-03-ME08

The evolution of pathobiology in the genus Acinetobacter

Bardya Djahanschiri[1], Sachli Zafari[1], Ruben Iruegas Lopez[1], Adrian Podolski[1], Ingo Ebersberger[1, 2]

[1]Institute for Cell Biology and Neuroscience, Goethe University Frankfurt, Frankfurt am Main (Germany), [2]Senckenberg, Frankfurt am Main (Germany)

Within only a few decades, *Acinetobacter baumannii* has evolved from an exotic pathogen to a top-priority organism for R&D on the WHO list of 2017. As novel antibiotics become increasingly hard to obtain, a deeper understanding of how the pathogen adapts to, and interacts with, the human host is key to the identification of new druggable targets.

Here, we shed light on the trajectory of genomic innovations that underlie *Acinetobacter* pathobiology. We base our analyses on a comparative genomics framework that extends from bacterial genome assembly and annotation, via phylogenomics and ancestral gene set reconstructions, up to an automated identification of proteins whose lineage-specific changes in their domain architectures suggest the adoption of new or at least modified functions. Our analysis of 2,474 *Acinetobacter ssp.* genomes reveals that a substantial fraction of *A. baumannii* strains group outside of the monophyletic clades representing the eight international clone types. Thus, a considerable fraction of *A. baumannii* diversity remains to be charted. An orthology-based reconstruction of ancestral gene sets along the *Acinetobacter*phylogeny identifies then acquisitions and losses of genes and associated functions that reflect adaptions to different hosts, and species or strain-specific modifications in virulence. Eventually, we demonstrate how a high-resolution candidate approach, focusing on proteins that are in direct contact with the bacterial environment, identifies potential novel virulence factors. We discuss in detail two promising candidates, ComC and Ata that have modified their domain architecture in *A. baumannii* presumably facilitating a tighter adhesion to the host membrane.

SY22: Molecular bases of the different forms of flowers on plants of the same species (July 10, 16:00−18:30)

O-02-MB01

Diversification of sexual system: insights from the persimmon genome

Takashi Akagi[1, 2]

[1]Kyoto University (Japan), [2]JST (Japan)

Sexual polymorphism, a main strategy to maintain genetic diversity within a species, has long been a major focus in biology. In contrast to the situation in higher vertebrates, the evolution of sexual systems in plants likely occurred independently in multiple sexual lineages and only few of the mechanisms underlying these transitions have been unveiled. Previously, we elucidated the molecular mechanism of sex determination in persimmons (*Diospyros* spp.) and demonstrated the action of a Y-encoded sex determinant pseudo-gene called *OGI*, via the production of small-RNAs against the gene encoding the female-determining factor *MeGI* and resulting in separate male and female individuals (dioecy). Here, we present evolutionary pathways of these two key sex determining genes, triggered by paleo-ploidization/duplication events. Full draft genome sequences of a dioecious persimmon revealed a lineage-specific whole genome duplication event on the K-Pg boundary associated with strong positive selection on a small subset of the duplicated genes, which include *MeGI*. Evolutionary and physiological analyses indicated that *MeGI* underwent neofunctionalization via positive selection after duplication from its sister gene, *Sister of MeGI* (*SiMeGI*), and acquired a new function as a repressor of male organ development. On the other hand, *OGI* was derived from further lineage-specific local duplications from *MeGI*, which established an inverted repeat co-evolved with a forwarded repeat, to produce small-RNAs indispensable for expression of dioecy. These findings exemplify how plant-specific numerous genome duplication events can contribute flexible genetic material whose variation can be selected for development of new sexual systems.

O-02-MB02

Characterisation of the S locus that controls heterostyly in Primula

Philip Gilmartin[1, 2], Jinhong Li[1, 2], Jonathan Cocker[1, 2], Calum Raine[1, 2], Yasuko Yoshida[1, 2, 3], Cock van Oosterhout[4]

[1]University Of East Anglia (United Kingdom), [2]The Earlham Institute (United Kingdom), [3]Kobe University (Japan), [4]University Of East Anglia (United Kingdom)

Heterostyly was described by Charles Darwin as a mechanism in Primula species that promotes cross-fertilization via insect-mediated pollination. This outbreeding mechanism has evolved independently in different angiosperm families to produce different forms of hermaphrodite flower of different individuals. The most common arrangement involves distyly with two forms of heteromorphic flowers that exhibit reciprocal herkogamy with reciprocal stigma and anther heights. The phenomenon has been studied in a small number of species from different taxa. In the Primulaceae, heterostyly is widespread and the majority of species exhibit heterostyly. we have focused on the common primrose, Primula vulgaris where the two form of flower are known as pin and thrum. Pin flowers have a long style, low anthers and produce small pollen grains; thrum flowers have high anthers, a short style and produce fewer but larger pollen grains. An incompatibility system reduces self-pollination between flowers of the same floral morph. These different characteristics are controlled by the S locus, a co-adapted linkage group of genes that define the archetypal supergene. We have applied a combination of classical and molecular genetics, together with genomic approaches, to identify and characterise the Primula S locus. This presentation will focus on our function analysis of the five genes at the S locus and comparisons between different Primula species.

SY22: Molecular bases of the different forms of flowers on plants of the same species (July 10, 16:00−18:30)

O-02-MB03

Identification of genes at the S-locus controlling heteromorphic self-incompatibility in buckwheat through genetic and genomic analyses

Yasuo Yasui[1], Nobuyuki Mizuno[1], Jotaro Aii[2], Tatsuya Ota[3]

[1]Kyoto University (Japan), [2]Niigata University of Pharmacy and Applied Life Science (Japan), [3]The Graduate University for Advanced Studies (Japan)

Identification of genes at the S-locus controlling heteromorphic self-incompatibility in buckwheat through genetic and genomic analyses

In common buckwheat (*Fagopyrum esculentum* Moench; $2n = 2x = 16$), dimorphic flowers, namely short-styled and long-styled flowers, is present with exhibiting self-incompatibility. The floral morphology and intra-morph incompatibility are both determined by a single genetic locus named *S*-locus, where plants with short-styled and long-styled flowers are heterozygous (*S*/*s*) and homozygous recessive (*s*/*s*), respectively. By conducting RNA-seq analyses for the pistils of the two floral morphs and examining differentially expressed genes, in particular, ones specific to *S-allele*, we identified a novel gene named *S-ELF3*, which was present in the genome of only short-styled buy not long-styled plants in world-wide landraces of buckwheat as well as in two distantly related Fagopyrum species with exhibiting heteromorphic self-incompatibility. Independent disruptions of *S-ELF3* were detected in a recently emerged self-compatible Fagopyrum species and a self-compatible line of buckwheat. The nonessential role of *S-ELF3* in the survival of individuals and the prolonged evolutionary presence only in the genomes of short-styled plants of genus *Fagopyrum*exhibiting heteromorphic SI suggests that *S-ELF3* is a suitable candidate gene for the control of the short-styled phenotype of buckwheat plants. Furthermore, a huge genomic region (5,393,196 bp) tightly linked to *S*-allele was identified from draft genome of buckwheat through association analyses using world-wide buckwheat landraces. In the presentation, we will update our knowledge on genomic structure of *S*-locus and discuss about the evolutionary history of *S*-linked genes in *Fagopyrum*species.

O-01-OS01

Whole genome integration of giant virus to chlorella and its implication to the evolution of giant virus and the symbiosis of chlorella and ciliates

Atsushi Ogura[1], Ryuhei Minei[1], Yoshino Murase[1], Ryo Hoshina[1]

[1]Nagahama Institue of Bioscience and Technology (Japan)

Paramecium burusaria Chlorella virus (PBCV) belonging to the family Phycodonaviridae is a double-stranded DNA giant virus with a genome size of approximately 330 kb. PBCV-1, a representative strain of PBCV, is known to infect a symbiotic chlorella Chlorella variabilis (Cva) that can coexist with the paramecium burusaria. On the other hand, no infection has been reported to Chlorella vulgaris (Cvu), free-living chlorella that does not have a symbiotic relationship with the ciliates. We sequenced the genome of Cvu and conducted comparative genomic analysis using Cva, Cvu, and PBCV-1, and found that whole genome of the giant virus has been integrated into the Cvu that possess resistance to PBCV-1. We confirmed that the insertion of virus integration site (VIS) was caused by horizontal transfer by phylogenetic analysis using Nucleo-Cytoplasmic Virus Ortholog Genes (NCVOGs) in VIS, and identified the boundary between Chlorella genome and VIS by a long-read sequencer. In the study, we also found that GC content and codon usage frequency of the entire gene in VIS are closer to Cvu than PBCV-1, suggesting that the insertion timing of VIS is in ancient age. These findings provide unique opportunity to study both the evolution of giant virus using "fossil" virus genome and evolutionary role of virus integration for the host chlorella regarding capability of symbiosis against the ciliates.

O-01-OS02

Estimating a time tree of the Australian radiation of Eugongylus group lizards by applying StarBEAST2 to a genome-scale data set

Huw Ogilvie[1], Jason Bragg[1, 2], Ana Afonso Silva[1], Mozes Blom[1, 3], Sally Potter[1], Craig Moritz[1]

[1]Australian National University (Australia), [2]The Royal Botanic Gardens and Domain Trust (Australia), [3]Swedish Museum of Natural History (Sweden)

Eugongylus group lizards form a large clade of skinks, and a massive Miocene radiation centered in Australia includes approximately 120 recognized extant species. These species are adapted to an extraordinary range of climates, from arid deserts (e.g. *Cryptoblepharus ochrus* and *Carlia triacantha*), to rainforests (e.g. *Saproscincus* spp.), to freezing mountaintops (*Niveoscincus*). To resolve species relationships, divergence times and macroevolutionary patterns of the Australian radiation, we inferred a time tree from 580 individuals representing almost all Australian lineages.

To generate sequence data for the tree we used exon capture, which has the advantage of being genomic in scale, but more cost-effective than whole genome sequencing and usable at greater phylogenetic depth than RAD-seq. To automate quality control of the aligned loci, we developed a method to detect the presence of anomalous sequences using correlation coefficients.

In addition to clean data, multispecies coalescent (MSC) methods are essential for accurately estimating species trees at the time scale of this radiation. However the numbers of loci and individuals are both too large to be used as-is with StarBEAST2, a start-of-the-art Bayesian MSC method. So to estimate a time tree from hundreds of individuals, we used a divide-and-conquer approach which combined StarBEAST2-estimated trees of smaller, well supported clades. We took advantage of the excessive number of loci by replicating our analysis across non-overlapping subsets, which confirmed the robustness of our results. Our study demonstrates an approach that makes it practical to accurately estimate large time trees from genome-scale data.

O-01-OS03

Diversity and fitness impacts of extracellular vs. intracellular bacterial associates of butterflies

Kruttika Phalnikar[1], Krushnamegh Kunte[1], Deepa Agashe[1]

[1]NCBS-TIFR (India)

A central focus of biological research is to understand the mechanisms leading to the enormous diversity of life. Host-microbe interactions are a potential driver of animal diversification because microbes can strongly impact crucial parameters such as host diet, mate-choice, fecundity and survival. Thus, studying host-microbe interactions is important to understand host ecology, evolution and behavior. Butterflies represent an excellent model to study host-microbe associations as they show great diversity in their diets, ecological niches and behavior. We tested bacterial community composition across life stages and across species of 12 butterfly hosts. We found that bacterial communities do not change significantly with development, despite vastly different diets across stages. In contrast, dietary variation between species is strongly associated with distinct bacterial communities. Surprisingly, more specialist larvae have relatively similar microbiomes whereas the more generalist adults harbor very distinct bacterial flora. Thus, unexpectedly, adult butterflies seem to impose a stronger filter on their gut communities. Next, we tested the impacts of extracellular and intracellular bacteria on butterfly fitness. We found that in two butterfly species, eliminating dietary microbes had weak and variable fitness impacts indicating that butterflies do not have a strong beneficial association with their gut-microbes. In contrast, we found that in one of these two host species, intracellular *Wolbachia* bacteria distorted adult sex ratio giving a strong female bias. Overall, in butterflies, obligate endosymbionts appear to play a larger role in governing the ecology and evolution of specific hosts, compared to very weak effects of non-obligate extracellular bacterial associates.

O-01-OS04

Drivers of genome reduction in free-living marine bacteria

Haiwei Luo[1]

[1]The Chinese University of Hong Kong (China)

Surface ocean waters are dominated by free-living bacterial lineages with highly reduced genomes. The best examples are the cyanobacterial genus *Prochlorococcus*, the alphaproteobacterial clade SAR11, and the gammaproteobacterial clade SAR86, together representing over 50% of the cells in surface oceans. Several studies identified signatures of selection on these lineages in today's ocean, and postulated selection as the primary force throughout their evolutionary history. However, massive loss of genomic DNA in these lineages often occurred in distant past, and selective pressures underlying these ancient events have not been assessed. Here we probed ancient selective pressures by computing %GC-corrected rates of conservative and radical nonsynonymous nucleotide substitutions. Surprisingly, we found an excess of radical changes in several of these lineages, in comparison to their relatives with larger genomes. Furthermore, analyses of allelic genome sequences of several populations within these lineages consistently supported that radical replacements are more likely to be deleterious than conservative changes. The coincidence of the massive DNA losses and the accelerated fixation of more deleterious variants supports a primary role of genetic drift driving ancient genome reduction of marine bacterioplankton lineages.

O-01-OS05

Implications of population structure for site frequency spectra

Marcy Uyenoyama[1], Seiji Kumagai[1], Naoki Takebayashi[2]

[1]Duke University (United States), [2]University of Alaska, Fairbanks (United States)

We extend the Ewens Sampling Formula under infinite-alleles mutation to structured populations comprising two intermigrating demes. We also derive the moment generating function of the number of alleles observed in a sample of a specified size and the probability that the next-sampled gene represents a novel allele. Conditioning on biallelic samples, we determine the exact distribution for the folded site frequency spectrum (SFS), a summary statistic widely used in explorations of genome-scale sequence data. We find that unlike the panmictic case, the SFS depends on theta, the scaled rate of substitution. Variation in theta across segregating sites can generate both excesses and deficiencies in rare alleles relative to expectation assuming panmixis, even in samples derived entirely from a single deme of a structured population. As a consequence, patterns often interpreted as signatures of selection or changes in effective number may in fact reflect population structure alone.

O-01-OS06

Developmental delay caused by mitochondrial replacement reveals co-evolutionary divergence of nuclear and mtDNA

Xu Wang[1, 2], Ting Li[1], Rachel Edwards[3], John H. Werren[3]

[1]Auburn Universiy (United States), [2]HudsonAlpha Institute for Biotechnology (United States), [3]University of Rochester (United States)

Two closely-related parasitoid wasp species, *Nasonia vitripennis* (Nv) and *N.girault*i(Ng) diverged one million years ago. Nv-mitochondria replacement strain in Ng nuclear background(designated GG[V]) has is a 2-day developmental delay compared to Ng(GG[G]). After maintaining in the laboratory for 25 years, 4320 nucleotide substitutions occurred (current genetic background represented as G^sG^s[V]). RNA-seq experiments in all possible crosses between Nv(VV[V]), Ng(GG[G]) and G^sG^s[V] discovered that the expression of G^s nuclear genome resembles G in [G] mitochondrial background, with only 45 differentially expressed genes (DE-genes) between GG[G] and GG^s[G]. Surprisingly, ~3000 DE-genes were found in G^sG^s[V]-G^sG[V] and VG^s[V]-VG[V] comparisons, indicating huge G^s-G expression divergence under [V], comparable to the number of DE-genes between Nv(VV[V]) and Ng(GG[G]). Half of the DE-genes overlap in the two comparisons, suggesting shared incompatibility between G and [V]. We hypothesized that due to concerted nuclear-mitochondria substitutions in the ancestral Ng(GG[G]) lineage, G is no longer compatible with [V], and G-to-G^s changes in G^sG^s[V] reverted the ancestral V allele to resolve the nuclear-mitochondria incompatibility, at a cost of developmental delay. Among 4320 G-G^sdifference, 2036 are also Nv-Ng SNPs. These substitutions were polarized and 98.3%(2002/2036) have V as the ancestral and G as the derived allele (P<1e-16 compared to ~50% in the entire genome), suggesting G^s is re-adapting [V] mitochondria. Genes with exonic substitution in G^s are significantly enriched in developmental genes with early embryo-biased expression, which is consistent with the developmental trade-off. Our study will shed light on the mitochondrial evolution and the nuclear-mitochondrial interaction.

O-01-OS07

Drift robustness and the evolution of genome architecture in small populations

Thomas LaBar[1, 2, 3], Christoph Adami[1, 2, 3]

[1]Michigan State University (United States), [2]Michigan State University (United States), [3]Michigan State University (United States)

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

O-01-OS08

Weak selection primes non-coding sequences for de novo evolution

Luke Kosinski[1], Joanna Masel[2]

[1]University of Arizona (United States), [2]University of Arizona (United States)

Weak forces of selection may help explain *de novo* evolution of protein-coding sequences from non-coding regions. Polypeptides from non-coding sequences are usually assumed to comprise poor material for adaptive evolution because they are not exposed to selective pressures. However, molecular errors translate non-coding sequences in small amounts, exposing their polypeptide products to selection, albeit much weaker than selection on coding regions. The "pre-adapting selection" hypothesis states that selection is sufficient to purge deleterious non-coding sequences and leave behind benign sequences, i.e. that the effects of translating deleterious sequences are strong enough and their translation by error is frequent enough. We tested this hypothesis in *Saccharomyces cerevisiae* by examining the non-coding sequences downstream of stop codons, which are translated following stop codon readthrough errors. We inferred per-gene levels of readthrough both directly from ribosome profiling data and indirectly from protein abundance. We inferred benign effects from high predicted intrinsic structural disorder, low predicted aggregation propensity, and short extension length. We confirmed that highly translated extensions showed more benign characteristics, providing the first strong empirical support for the pre-adapting selection hypothesis. High intrinsic structural disorder and low aggregation propensity predict benign vs. deleterious effects not only for C-terminal extensions, but also for the effects of random polypeptides on *Escherichia coli* growth rates as measured by Neme et al. (2017). These results show that the weak selective force of pre-adapting selection purges deleterious non-coding sequences, enriching for the pool of benign sequences that better form the raw material of adaptive evolution.

O-01-OS09

Retrotransposons spread potential sources of cis-regulatory elements for mammary gland evolution

Hidenori Nishihara[1]

[1]Tokyo Institute of Technology (Japan)

Acquisition of *cis*-elements is a major driving force for rewiring a gene network, but the evolutionary process is largely unknown. Several kinds of transposable elements (TEs), mostly retrotransposons that propagate via a copy-and-paste mechanism, possess transcription factor binding motifs and have provided source sequences for enhancers/promoters, followed by alterations in gene regulatory networks. It remains largely unknown whether retrotransposons have spread the binding sites of master regulators of morphogenesis and accelerated *cis*-regulatory expansion involved in common mammalian morphological features during evolution. Here, I demonstrate that no less than thousands of binding sites for ERα, FoxA1, GATA3, and AP2γ that are essential regulators of mammary gland development arose from a spreading of the binding motifs by TEs. In particular, L2, MIR, and endogenous retrovirus 1 (ERV1) families made a large contribution to facilitate this regulatory evolution. The TE-derived elements serve primarily as distal enhancers and are enriched around genes associated with mammary gland morphogenesis. The source TEs occurred via a two-phased expansion consisting of L2/MIR in a eutherian ancestor and ERV1 in simian primates. Thus TEs have had a step-by-step but dramatic impact on both the establishment and diversification of the gene regulatory system during mammalian evolution. The build-up of potential sources for *cis*-elements by retrotransposons followed by their frequent utilization by the host (co-option/exaptation) may have a general accelerating effect on orchestration of a gene regulatory network, leading to morphological innovation.

O-01-OS10

A large fraction of paralogous genes is not free to diverge independently due to molecular interference

Christian Landry[1, 2, 3], Axelle Marchant[1, 2, 3], Lou Nielly-Thibault[1, 3], Yacine Seffal[1, 3], Angel Cisneros[2, 3]

[1]Universite Laval (Canada), [2]Universite Laval (Canada), [3]Universite Laval (Canada)

Gene duplication is a major source of molecular innovation and contributes to the complexity of biological systems. The contribution of neutral and selective forces to the maintenance of paralogs has been studied at the theoretical and experimental levels. In most scenarios, paralogs are assumed to be contributing independently to their molecular functions and are thus free to diverge from each other. However, in many instances, paralogs can be linked by a shared function, which means that they may interfere which each other. We find that this is the case for instance for paralogs that encode self-interacting proteins. The duplication of these self-interacting proteins leads to pairs of physically interacting paralogs, making their function co-dependent. We have exhaustively examined the extent of this phenomenon in the yeast protein interaction network and found that a large fraction of paralogs likely derives from ancestral self-interacting proteins and that this is affecting their evolution and probability of retention. We are developing theoretical models based on molecular interference between paralogs that would explain this retention. Overall, our work shows that the fate of paralogs is linked by epistatic relationships when they derive from ancestral self-interacting proteins, which means that each paralog of a pair shapes the evolutionary landscape of the other.

O-01-OS11

A reassessment of evolutionary impact of genomic structural variation

Omer Gokcumen[1]

[1]University at Buffalo (United States)

Genomic structural variants (SVs, i.e., deletions, duplications, inversions and translocations of large segments of DNA) comprise a significant part of genetic variation among primates. Several studies have demonstrated adaptive phenotypes that are underlain by SVs. However, the current paradigm is that SVs contribute little to human phenotypic variation as compared to single nucleotide variants. We argue that the underappreciation of SVs stem primarily from multiple inter-related technical barriers inherent in studying these variants. Here, we present data from multiple studies, in which we conducted an empirical and model-based analyses of multiple common SVs and the haplotypes that harbor them, affecting genes with evolutionarily relevant functions: growth hormone pathway (*GHR*), epidermal differentiation complex (*LCE3* and *FLG* gene families), salivary function (*MUC7*), and xenobiotic metabolism *GSTM1* and *GSTT1*). Collectively, our results suggest that complex geography-specific and balancing adaptive forces maintain the most common exonic SVs in the human population. In addition, we confirmed that current methods of detecting selective sweep at the genome-wide level are inadequate to identify adaptive signatures affecting SVs: (i) SVs are located highly complex repetitive regions with high rates of recombination and gene conversion, resulting in weak and short linkage disequilibrium between SVs and flanking single nucleotide variants, (ii) few, if any, of SVs, have undergone a classical selective sweeps in humans. Our study provides a framework for future studies to adequately elucidate the role of SVs in human evolution.

O-01-OS12

Phylogenomics and comparative genomics of Palaeognathous birds reveal Palaeognath evolutionary history and ZW sex chromosome evolution

Zongji Wang[1, 2, 3], Jilin Zhang[4], Guojie Zhang[3, 5, 6], Qi Zhou[1, 2]

[1]University of Vienna (Austria), [2]Zhejiang University (China), [3]BGI-Shenzhen (China), [4]Karolinska Institutet (Sweden), [5]Kunming Institute of Zoology, Chinese Academy of Sciences (China), [6]University of Copenhagen (Denmark)

The evolution of the Palaeognathae, classified into tinamous and ratites, has intrigued biologists for over a century. Although recent molecular studies strongly supported ratite paraphyly, rheas evolutionary relationship within ratites is still controversial. Ratites, as the most basal bird order, have homomorphic sex chromosomes, which revealed not all old sex chromosome systems have degenerated heteromorphic sex chromosomes. Moreover, the evolutionary trajectory of sex chromosomes across Palaeognathae remains unresolved. Here we sequenced the genomes of 12 Palaeognathous birds and performed whole-genome phylogenetic analyses, which revealed that rhea is the closest relatives of tinamous and is distant from the basal Palaeognathae lineage of ostriches. This unexpected result strongly supports continental vicariance, which suggests similar life history traits between ostrich and rhea have been driven by convergent evolution. We identified about 85Mb Z-linked and different lengths of W-linked scaffolds, indicating that there is a great diversity of evolutionary differentiation across Palaeognathous birds. We found that almost all ratite birds have long pseudo-autosomal regions PARs, suggesting these species have a W chromosome that is not completely degenerated. However, The PARs and non-recombining regions of tinamou species exhibit a complex pattern, resulting from the suppression of recombination in a stepwise and independent manner among the ancestors of species with same genus level. Moreover, our study revealed the abundance of transposable elements probably contributes to the divergence of sex chromosomes in tinamous. Our study supported the hypothesis of rheas have originated through vicariant speciation and uncovered an unexpected complexity of sex chromosomes in Palaeognathae.

O-01-OS13

Cavefish Metabolic Adaptation: Hungry, Fat, and Healthy

Nicolas Rohner[1, 2]

[1]Stowers Institute for Medical Research (United States), [2] KU Medical Center (United States)

Understanding the genetic basis of physiological adaptation has broad implications not only for a basic understanding of evolution, but also for human pathologies given that many human metabolic diseases are a consequence of mis-adaptation to modern societies. The emerging model system Astyanax mexicanus has become an important fish species to address metabolic adaptation to extreme environments due to its unique ecology and the availability of genetic tools and genomic resources.

Cave environments are typically dark and as a consequence nutrient deprived. We have previously shown that cavefish acquired impressive adaptations such as increased appetite, starvation resistance and altered feeding behaviors to cope with these conditions. In this study, we show that cavefish display elevated blood sugar levels and insulin resistance compared to surface fish, but without any effect on the health of the fish. On the contrary, we show that these phenotypes are helping the cavefish to gain weight more quickly as part of their starvation resistance strategy.

Using whole genome information of different cave and surface populations, we identified a structural mutation in the insulin receptor of cavefish underlying the observed insulin resistant phenotypes. We use CRISPR mediated gene editing to show that this mutation is sufficient to cause a similar phenotype in zebrafish and provide evidence for strong selection of this allele in the wild. Interestingly, the same mutation is found in cases of Type 2 diabetic patients in human populations, raising the question whether cavefish can be used to gain insight into human glucose control homeostasis.

O-01-OS14

Recombination Hotspots and Imprinted Genes in Indigenous African Cattle

Kelsey Caetano-Anolles[1], Taehyung Kwon[1], Heebal Kim[1]

[1]Seoul National University (Republic of Korea)

Genetic recombination is the rearrangement of genetic material from different chromosomes or between different regions of the same chromosome which results in novel combinations of alleles that differ from those of either parent. Studies on yeast and mammalian genomes have revealed that some locations on the chromosome, known as hotspots, show higher rates of cross-over than others. The recombination rate in these regions can be up to hundreds of times higher than those of the surrounding non-hotspot regions. Although recombination and recombination hotspots have been previously studied in several mammal species including cattle, no study has yet been performed in indigenous African cattle. The present study aimed to calculate recombination rates and map the location of recombination hotspots in the genome of 162 African indigenous cattle from 15 different breeds. Additionally, we mapped the locations of previously known imprinted genes in cattle in order to determine whether or not a correlation exists between imprinted regions and recombination hotspots in African Cattle. A correlation between recombination hotspots and imprinted genes has been noted in the literature, and the average distance between imprinted genes and recombination hotspots has been previously shown to be around one third that of all annotated genes. However, the relationship between these phenomena is not yet well understood. Results of this study not only expand the limited pool of information available on the genome of indigenous African cattle, but also further elucidates the relationship between imprinted genes and recombination hotspots in the mammalian genome.

O-01-OS15

A comprehensive lipidome map of human brain

Ekaterina Khrameeva[1, 2], Waltraud Mair[1], Anna Tkachev[1, 2], Patricia Guijarro[4], Ilia Kurochkin[1], Olga Efimova[1], Aleksandra Mitina[1], Dmitry Zubkov[1], Philipp Khaitovich[1, 3, 4]

[1]Skolkovo Institute of Science and Technology (Russian Federation), [2]Institute for Information Transmission Problems (Russian Federation), [3]Max Planck Institute for Evolutionary Anthropology (Germany), [4]CAS-MPG Partner Institute for Computational Biology (China)

The lipid composition of brain anatomical structures remains poorly understood, particularly in humans and closely related non-human primates. We describe the generation and analysis of a lipidome atlas of the adult human brain, comprising a large-scale mass spectrometry-based lipidome profiling of 75 anatomically precise subdivisions in four individuals. Lipid composition varies enormously by anatomical location, with different regions displaying robust molecular signatures (modules) that are enriched with particular lipid classes and functional pathways, and are highly conserved between individuals. Modules show striking anatomical specificity, especially for non-cortical brain regions, while neocortex displays a relatively homogeneous lipidome composition, but with distinct features associated selectively with primary cortices. To study evolution of lipid composition of brain anatomical structures, we produce lipidome atlases of adult chimpanzee, bonobo and macaque brains in three individuals per species. In agreement with previous studies [Bozek et al., 2015], the human brain shows the most distinct lipid composition and evolves faster, compared to the other primates. Human-specific concentration changes are supported by human-specific expression changes for corresponding enzymes. This data resource forms a comprehensive baseline for studies of normal and abnormal human brain function and evolution.

References:

Bozek, K. et al. Organization and Evolution of Brain Lipidome Revealed by Large-Scale Analysis of Human, Chimpanzee, Macaque, and Mouse Tissues. Neuron 85, 695-702 (2015).

This study was supported by the Skoltech Systems Biology Fellowship awarded to E.K.

O-01-OS16

Mutation dynamics and fitness effects followed in single cells

Marina ELEZ[1, 2], Lydia ROBERT[1, 3], Jean OLLION[1]

[1]UPMC, Sorbonne University (France), [2]UEVE, University Paris Saclay (France), [3]INRA, AgroParisTech, University Paris Saclay (France)

Mutations have been investigated for more than a century but never witnessed in action in single cells, thus preventing direct characterization of their dynamics and reliable estimation of the distribution of their fitness effects. Here we accomplished this in Escherichia coli by employing microfluidics and time-lapse imaging. We visualized replication errors giving rise to point mutations in single cells using a fluorescent tag of the Mismatch Repair System, revealing a Poissonian dynamics. We also followed the growth and lifespan of individual cells accumulating ~20 000 genome-wide mutations during ~200 generations of growth in >1000 independent microchannels designed to block natural selection, resulting in $>10^5$ fitness measurements. We provide the first non-parametric characterization of the distribution of mutation fitness effects. We found a wide, heavy-tailed distribution dominated by neutral mutations, with a surprisingly weak average cost of ~0.3% for non-lethal mutations, 1% of lethal mutations, and negligible macroscopic epistasis.

O-03-OS01

A comprehensive map of genetic variation in the world's largest ethnic group - Han Chinese

Charleston W.K. Chiang[1, 2], Serghei Mangul[3, 4], Christopher R. Robles[5], Warren W. Kretzschmar[6], Na Cai[7, 8], Kenneth S. Kendler[9], Sriram Sankararam[3, 5], Jonathan Flint[2]

[1]University of Southern California (United States), [2]University of California, Los Angeles (United States), [3]University of California, Los Angeles (United States), [4]University of California, Los Angeles (United States), [5]University of California, Los Angeles (United States), [6]University of Oxford (United Kingdom), [7]Wellcome Genome Campus (United Kingdom), [8]Wellcome Genome Campus (United Kingdom), [9]Virginia Commonwealth University (United States)

As are most non-European populations around the globe, the Han Chinese are relatively understudied in population and medical genetics studies. From low-coverage whole-genome sequencing of 11,670 Han Chinese women we present a catalog of 25,057,223 variants, including 548,401 novel variants that are seen at least 10 times in our dataset. Individuals from our study come from 24 out of 33 administrative divisions across China (including 19 provinces, 4 municipalities, and 1 autonomous region), allowing us to study population structure, genetic ancestry, and local adaptation in Han Chinese. We identify previously unrecognized population structure along the East-West axis of China and report unique signals of admixture across geographical space, such as European influences among the Northwestern provinces of China. Finally, we identified a number of highly differentiated loci, indicative of local adaptation in the Han Chinese. In particular, we detected extreme differentiation among the Han Chinese at MTHFR, ADH7, and FADS loci, suggesting that these loci may not be specifically selected in Tibetan and Inuit populations as previously suggested. On the other hand, we find that Neandertal ancestry does not vary significantly across the provinces, and contrary to a previous report, Neandertal ancestry does not explain a significant amount of heritability in depression. Our findings provide the largest genetic data set so far made available for Han Chinese and provide insights into the history and population structure of the world's largest ethnic group.

O-03-OS02

Whole genome analysis of the Jomon remain reveals deep lineage of East Eurasian populations

Takashi Gakuhari[1, 3], Martin Sikora[2], Simon Rasmussen[2], Morten Allentoft[2], Takehiro Sato[1], Thorfinn Korneliussen[2], Minoru Yoneda[4], Hajime Ishida[5], Yasuhiro Yamada[6], Hiroki Shibata[7], Shigeki Nakagome[8], Eske Willerslev[2, 9, 10], Hiroki Oota[3]

[1]Kanazawa University (Japan), [2]University of Copenhagen (Denmark), [3]Kitasato University (Japan), [4]The University of Tokyo (Japan), [5]University of the Ryukyus (Japan), [6]National Museum of Japanese Histories (Japan), [7]Kyushu University (Japan), [8]the University of Dublin (Ireland), [9]University of Cambridge (United Kingdom), [10]Wellcome Trust Sanger Institute (United Kingdom)

Post late-Paleolithic hunter-gatherers lived throughout the Japanese archipelago, Jomonese, are thought to be a key to understanding the peopling history in East Asia. Here, we report a whole genome sequence (x1.85) of 2,500-year old female excavated from the Ikawazu shell-mound, unearthed typical remains of Jomon culture. The whole genome data places the Jomon as a lineage basal to contemporary and ancient populations of the eastern part of Eurasian continent, and supports the closest relationship with the modern Hokkaido Ainu. The results of ADMIXTURE show the Jomon ancestry is prevalent in present-day Nivkh, Ulchi, and people in the main-island Japan. By including the Jomon genome into phylogenetic trees, ancient lineages of the Kusunda and the Sherpa/Tibetan, early splitting from the rest of East Asian populations, is emerged. Thus, the Jomon genome gives a new insight in East Asian expansion. The Ikawazu shell-mound site locates on 34,38,43 north latitude, and 137,8, 52 east longitude in the central main-island of the Japanese archipelago, corresponding to a warm and humid monsoon region, which has been thought to be almost impossible to maintain sufficient ancient DNA for genome analysis. Our achievement opens up new possibilities for such geographical regions.

O-03-OS03

Divergent mitochondrial phenotypes associated with the Doubly Uniparental Inheritance (DUI) of mitochondria

Stefano Bettinazzi[1], Liliana Milani[2], Enrique Rodriguez[3], Pierre Blier[3], Sophie Breton[1]

[1]University of Montreal (Canada), [2]University of Bologna (Italy), [3]University of Quebec at Rimouski (Canada)

Mitochondrial energy production through oxidative phosphorylation (OXPHOS) depends on the coordinated expression of nuclear and mitochondrial DNA (mtDNA). In animals, strictly maternal inheritance (SMI) of mitochondria optimizes this interaction because it fosters dialogue between the nucleus and only one mitotype. DUI of mitochondria is a unique system characterized by two sex-linked mtDNAs (F- and M-type) associated with gametes, and also sometimes coexisting (heteroplasmy) in male somatic tissue. Considering the impact of a single nucleotide substitution in humans, the co-expression of such different mtDNAs (20-40% of nucleotide divergence) should have a strong, potentially deleterious, phenotypic effect.

What is the adaptive value of mtDNA variants? How heteroplasmy affects mito-nuclear coevolution and cellular fitness? Could differences in the energetic metabolism explain DUI retention and represent an adaptation for male functions?

OXPHOS activity of oocytes, spermatozoa, and gills, of different species of bivalves (DUI and SMI), has been investigated through high resolution respirometry. The overall activity, the capacity of electron transport chain complexes, as well as the presence of alternative respiratory pathways have been measured and compared among species.

Different mitochondrial phenotypes associated with DUI haplotypes give us new insights on how respiratory activity can be modulated, possibly reflecting a different role and energetic demand of cells bearing sex-specific mitochondria. These findings highlight the phenotypic consequences of mtDNA variations, pointing over a common evolutionary meaning of DUI in the light of male-energetic adaptation, mitotype preservation and inheritance. Moreover, specific OXPHOS features are consistent with resistance to both heteroplasmy and ageing.

O-03-OS04

Molecular clocks on Chelicerata suggest an early colonization of land by arachnids and support the monophyly of mites

Davide Pisani[2, 3], Alastair R. Tanner[3], Jakob Vinther[2, 3], Gregory R. Edgecombe[4], Jesus Lozano-Fernandez[1, 2, 3]

[1]University of Barcelona (Spain), [2]University of Bristol (United Kingdom), [3]Bristol (United Kingdom), [4]The Natural HIstory Museum (United Kingdom)

Animal life has marine origins, with only few phyla completing their entire life cycle outside water. The process through which organisms adapt to life on land is known as terrestrialisation, and it is one of the most extreme cases of adaptation to a new habitat. The chelicerates (pycnogonids, horseshoe crabs, spiders, scorpions) are an ancient group of arthropods, with an astonishing fossil record dating back to Cambrian, and includes the second largest clade of fully terrestrial organisms, the arachnids. Morphological phylogenies support a single colonization of land by placing marine horseshoe crabs as the sister group of terrestrial arachnids, but phylogenomic studies nest this aquatic lineage within Arachnida (implying multiple terrestrialisation events). To identify *how* many times and *when* arachnids adapted to life on land we need to assess the phylogeny of chelicerates, and its evolutionary timescale.

Here, we present a timescale for Chelicerata designed to test how many times and when arachnids adapted to life on land. We used an expanded multigene dataset covering most chelicerate diversity and the largest set of fossil calibrations to date. Our results recover monophyly of Chelicerata, Euchelicerata and Arachnida, suggesting a single terrestrialization event. Furthermore, we found Acari as monophyletic (Parasitiformes+Acariformes) and recover Tetrapulmonata (Araneae+Pedipalpi) in alliance to Scorpiones (Arachnopulmonata) or allied to a clade composed by Scorpiones+Pseudoscorpiones. Our results reconcile previous results based on morphology and molecular evidence, suggesting a Cambrian-Ordovician colonization of land by arachnids, substantially predating trace or body fossil evidence.

O-03-OS05

DNA repair in bdelloid rotifers: genome dynamics.

Karine Van Doninck[1], Matthieu Terwagne[1], Boris Hespeels[1], Bernard Hallet[2]

[1]University of Namur (Belgium), [2]Universite Catholique de Louvain (Belgium)

Large genomic rearrangements are hypothesized to be a key component of genomic plasticity enabling adaptation. Genome sequencing of the first bdelloid rotifer species Adineta vaga revealed a peculiar genomic structure as allelic regions are shuffled across chromosomes with no homologous pairs being present. Moreover, it has been previously demonstrated that prolonged periods of desiccation or high doses of ionizing radiation induce DNA double strand breaks that bdelloid rotifers repair upon rehydration. The origin of the genomic rearrangements observed in the Adineta vaga genome and its genome dynamics following cycles of desiccation in their ephemeral habitats or exposure to ionizing radiation remains unknown. The genome integrity and reconstructed genome profile was studied using Pulsed-Field Gel Electrophoresis coupled to restriction enzymes in bdelloid individuals exposed to desiccation and radiation and in their descendants. Interestingly, no modification in the restriction genomic profile was observed in their F1 descendants despite massive DNA damage induced by 500 Gy ionizing radiation. We therefore hypothesize that genome reconstitution may rely on homology-driven de novo synthesis rather than by non-homologous end joining. The mechanisms of DNA repair are studied in detail in bdelloid rotifers by transcriptome analysis and using in vivo labelling of newly produced DNA. The labelled DNA is then detected at each time point by a click chemical reaction with a fluorescent dye and visualized by whole-organism confocal imaging. The preliminary results obtained are promising and all the new data will be presented at SMBE 2018.

O-03-OS06

Evolution of an intratumoral ecology susceptible to successive treatment in breast cancer xenografts

Hyunsoo Kim[1], Pooja Kumar[1], Francesca Menghi[1], Javad Noorbakhsh[1], Eliza Cerveira[1], Mallory Ryan[1], Qihui Zhu[1], Chengsheng Zhang[2], James Keck[2], R Krishna Murthy Karuturi[3], Carol J Bult[3], Charles Lee[1], Edison T Liu[3], Jeffrey H Chuang[1, 4]

[1]The Jackson Laboratory (United States), [2]The Jackson Laboratory (United States), [3]The Jackson Laboratory (United States), [4]UConn Health (United States)

The processes by which tumors evolve are essential to the efficacy of treatment, but quantitative understanding of intratumoral dynamics has been limited. Quantification of evolution is difficult from clinical samples because treatment replicates cannot be performed. To circumvent this we derived and assayed large sets of human triple-negative breast cancer xenografts and cell cultures from two patients, including 86 xenografts from cyclophosphamide, doxorubicin, cisplatin, docetaxel, or vehicle treatment cohorts as well as 45 related cell cultures. We assayed these via exome-seq and/or high-resolution ddPCR, allowing us to distinguish complex therapy-induced subclone selection and drift processes with cellularity uncertainty <3%. For one patient, we discovered two predominant subclones that were granularly intermixed in all 48 co-derived xenograft samples. These two subclones exhibited differential chemotherapy sensitivity -- when xenografts were treated with cisplatin for 3 weeks, the post-treatment volume change was proportional to the post-treatment ratio of subclones on a xenograft-to-xenograft basis. A subsequent cohort in which xenografts were treated with cisplatin, allowed a drug holiday, then treated a second time continued to exhibit this proportionality. In contrast, xenografts from other treatment cohorts, spatially dissected xenograft fragments, and cell cultures evolved unsystematically but with substantial population bottlenecks. These results show that ecologies susceptible to successive retreatment can arise spontaneously in breast cancer in spite of a background of irregular bottlenecks. Intriguingly, in such an ecology the ratio of common subclones is predictive of the state of treatment susceptibility, suggesting that this can be measured to optimize dynamic treatment protocols in patients.

O-03-OS07

The molecular determinants of pheromone divergence and their role in the evolution of reproductive isolation in orchid bees

Philipp Brand[1], Thomas Eltz[2], Santiago Ramirez[1]

[1]UC Davis (United States), [2]Ruhr University Bochum (Germany)

Animal lineages across the tree of life rely on pheromones to locate and identify mates. Variation in pheromone communication systems has been linked to the origin and maintenance of reproductive isolation. Nonetheless, the molecular evolution of pheromone communication systems and its impact on the speciation process remains poorly understood. Male orchid bees collect scents from various environmental sources to concoct species-specific pheromone mixtures that are subsequently exposed to females during courtship. As a result, orchid bees rely exclusively on the sense of smell to accumulate and detect pheromones. Here, we tested the hypothesis that chemosensory gene divergence underlies the evolution of pheromone communication and reproductive isolation in orchid bees. We conducted a population-level analysis of two recently diverged sibling species including 366 bees collected throughout the entire distribution range of both species. We identified that pheromones were species-specific, despite low genetic differentiation and ongoing gene flow. After correcting for population structure, we identified few highly differentiated genomic regions between species. In particular, we found pronounced differentiation in a tandem array harboring 37 odorant receptor (OR) genes. While most of these genes were of low divergence, we identified a selective sweep around OR41 in combination with significantly elevated dN/dS in one species, suggesting it evolved under strong positive selection. Together our results indicate that OR41 may play an important role in species differentiation. The combination of pheromone divergence and rapid evolution of OR41 suggests a magic-trait scenario and predicts that divergent chemosensory genes generate important reproductive barriers during orchidbee speciation.

O-03-OS08

Fifty years of Neutral Theory: Past, Present and Future Perspective

Antonio Barbadilla[1], Sonia Casillas[1]

[1]Universitat Autonoma de Barcelona (Spain)

This presentation addresses the fundamental role that neutral theory has been playing in the field of population genetics and molecular evolution, since its inception, half a century ago, now and which will play in the foreseeable future. In 1968, Motoo Kimura proposed a radically new explanation for the first available descriptions of population variation and substitution rates of proteins. Assuming that most of the existing polymorphisms and the fixed differences between the species are selectively neutral and functionally equivalent, Kimura's neutral theory provides a diaphanous kinetics and dynamics of molecular evolutionary change, encapsulated in one of the most elegant mathematical expressions of science: $K = \mu$. If variants are neutral, then evolutionary rate depends on the mutational rate only. Because of its simplicity, intelligibility, robustness, and feasible theoretical predictions, the (nearly -with Otha's addition) neutral theory of molecular evolution became enthroned as the universal null model against which to test any selective hypothesis. The subsequent advent of polymorphic DNA sequences showing a correlation between recombination rate and amount of polymorphism in most studied species challenged the neutral paradigm, supporting the relevance of recurrent linked selection. If linked selection is common, a neutral framework to analyze genome data would distort the interpretation of variation patterns. In fact, the genomes of some species can be envisaged as mosaics of pieces evolving either under a neutral framework or under recurrent linked selection. Future population genomics should incorporate to the neutral framework integrative models, which allow performing powerful, knowledge-based, population genomics tests.

O-03-OS09

The *E. coli* translation machinery evolves towards minimal total mass concentration at the required protein production rate

Xiao-Pan Hu[1], Hugo Dourado[1], Martin J. Lercher[1]

[1]Heinrich Heine University Duesseldorf (Germany)

The total concentration of dry mass in bacterial cells is approximately constant across environments and growth rates. If the cell allocates more of this total mass concentration to a particular process, than less of the dry mass can be devoted to other important cellular processes. In growing bacterial cells, a large fraction of cellular dry mass is occupied by the translation machinery, including mRNA, ribosomes, charged tRNAs, and elongation factors. We hypothesize that to maximize *E. coli* growth rate in a given environment, natural selection minimizes the summed mass concentration of translation components needed to achieve the required translation rate, thereby freeing up parts of the limited cellular "solvent capacity" for other processes. To test this, we constructed a mathematical model of *E. coli* translation encompassing about 100 biochemical reactions and fully parameterized with literature values. At a given *E. coli*growth rate, we minimize the combined mass concentration of ribosomes, mRNA, elongation factors, and charged tRNAs. Without any free fitting parameters, the predictions agree accurately with measured concentrations and ribosomal elongation rates across multiple growth conditions. Thus, gene regulation of protein translation in *E. coli* indeed evolves towards a predictable, condition-dependent fitness peak.

O-03-OS10

Phenotypic and genomic changes involved in cross-species transmission and emergence of a zoonotic virus

Roman Biek[1], Nardus Mollentze[1], Daniel Streicker[1, 2], Katie Hampson[1], Pablo Murcia[2], Denise Marston[3], Anthony Fooks[3]

[1]University of Glasgow (United Kingdom), [2]University of Glasgow (United Kingdom), [3]Animal and Plant Health Agency (United Kingdom)

Viruses are notorious for their ability to emerge in novel host species following cross-species transmission, a phenomenon with large economic and health implications. However, identifying what constrains viruses in their ability to establish in a new host remains an important but challenging research area. Here we use a combination of meta-analyses, molecular evolutionary studies, and experimental work to shed light on the phenotypic and genetic changes underlying viral cross-species transmission and disease emergence, using rabies virus as a study system. Applying a meta-analysis approach to data collected from published cross-species infection experiments, involving over 900 individuals, we show that rabies incubation time, the period between infection and first symptoms, in the novel host changes as a function of several predictors, including a previously unrecognised effect of host body temperature difference between donor and recipient species. We further present data from an unusual case of rabies virus emerging in a wild ungulate population. Contrary to predictions, we find limited evidence for specific substitutions in virus genomes linked to emergence. However, based on in vitro experiments we are able to show differences in infectivity that are consistent with rapid adaptation to the novel host. Taken together, our results point to subtle but epidemiologically important phenotypic changes that are driving the maintenance of rabies in particular host species as well as its rapid emergence in novel hosts, although the genotypic basis for these processes still remains to be confirmed.

O-03-OS11

A GENOMIC TIMELINE FOR THE EVOLUTION AND DIVERSIFICATION OF SEA SPIDERS (ARTHROPODA: PYCNOGONIDA)

Jesus A Ballesteros[1], Claudia P Arango[2], Georg Brenneis[3], Prashant P Sharma[1]

[1]University of Wisconsin-Madison (United States), [2]Queensland Museum (Australia), [3]Wellesley College (United States)

Sea spiders (class Pycnogonida) are one of the most enigmatic and understudied lineages in the phylum Arthropoda. Pycnogonids harbor ca. 1350 described marine and estuarine species that occur in all oceans of the world, from the shallow intertidal zones to abyssal depths. Their unique and peculiar morphology clearly separates sea spiders from all other arthropods but at the same time has obscured their placement in the tree of life. Current phylogenetic hypotheses place pycnogonids either as the sister group to all other arthropods or within Chelicerata as the sister group to the clade of horseshoe crabs + arachnids. Although the fossil record of pycnogonids is scarce, fossilized sea spiders represent some of the earliest representatives of any of the extant arthropod groups. The oldest crown group fossil dates from the early SIlurian (430 mya), the oldest putative larval fossil dates from the Cambrian (497 mya). However, no dated phylogeny of the group exists, and the age of extant Pycnogonida is entirely unknown. Here, we sequenced and analyzed the transcriptomes of seven species of pycnogonids representing 6 of the 11 known families, and combined phylogenomic and modern dating techniques to produce the first backbone phylogeny of the group. We explored the effect and robustness of the estimates ages with and without internal calibration points. Our dated phylogenomic tree of Chelicerata provides the first timeline for the diversification of this enigmatic group of organisms.

O-03-OS12

A Path Integral Method for Analytically Tractable Inference of Evolutionary Dynamics

John P. Barton[1], Raymond H. Y. Louie[2], Matthew R. McKay[2, 3], Muhammad S. Sohail[2]

[1]University of California, Riverside (United States), [2]Hong Kong University of Science and Technology (Hong Kong), [3]Hong Kong University of Science and Technology (Hong Kong)

Understanding the forces that shape genetic evolution is a subject of fundamental importance in biology and one with numerous practical applications. Modern experimental techniques give insight into these questions, but inferring evolutionary parameters from sequence data, such as how an organism's genotype affects its fitness, remains challenging. Here we present a method to infer selection from genetic time-series data using a path integral approach based in statistical physics. This approach allows us to derive an intuitive, closed-form solution for the most likely selection coefficients underlying an observed evolutionary trajectory, while taking into account the influences of mutation and genetic linkage. We illustrate the effectiveness of this approach using several simple examples, such as disentangling the selection coefficients for hitchhiking mutations. Through extensive numerical tests we also find that our method meets or exceeds the current state of the art in the successful classification of mutations as beneficial or deleterious in a variety of scenarios, while also yielding enormous improvements in run time compared to Monte Carlo-based methods. Our approach can also be extended to jointly infer other evolutionary parameters such as the effective population size and mutation rates.

O-03-OS13

The thermodynamical roots of pair-wise epistasis in alpha helix of beta-lacatamse TEM-1

Andre Birgy[1, 2], Harry Kemble[1], Jimmy Mullaert[1], Herve Jacquier[1, 3], Olivier Tenaillon[1]

[1]University Paris Diderot/ Inserm (France), [2]AP-HP (France), [3]AP-HP (France)

The interactions between mutations on fitness or epistasis can have a large impact on genome or gene evolution. However, the mechanistic bases of epistasis remain largely unclear. At the gene level, epistasis may emerge from the impact of mutation on protein stability. The contribution of protein stability has however never been fully quantified. To have a more quantitative understanding of the molecular bases of epistasis, we have made a comprehensive library of more than 15,000 double mutants in 11 amino-acid alpha helix of beta-lactamase TEM-1. We first observed that epistasis was extremely diverse encompassing both large positive and negative values. Particularly large positive epistasis was observed between beneficial mutations and deleterious ones. Second, once unconditionally lethal mutations were removed, epistasis was pervasive with less than 12% of mutations showing no clear sign of epistasis, most of them being neutral. Third, mutations could be categorized as destabilizing, neutral of stabilizing according to their overall pattern of epistasis. A thermodynamics model of protein stability recapitulated our results and could explain up to 84% of the variance. Nonetheless, clear deviation from the model could be observed especially for sites in direct contact. This suggests that epistatic interactions for residues in contact cannot be simply summarized by the sum of mutation impact on the overall protein stability.

O-03-OS14

Using transcriptomics to study hibernation in a natural primate population.

Jose Luis Villanueva-Canas[3], Sheena Lee Faherty[1, 2], Marina Blanco[2], M. Mar Alba[3, 4], Anne Diane Yoder[1, 2]

[1]Duke University (United States), [2]Duke University (United States), [3]Hospital del Mar Research Institute (IMIM), Universitat Pompeu Fabra (UPF) (Spain), [4]Catalan Institution for Research and Advanced Studies (ICREA) (Spain)

Hibernation is a complex physiological response some mammalian species employ to evade energetic demands. Hibernators conserve energy by shutting down physiological processes; metabolic rate is severely depressed, body temperature plummets to ambient levels, and brain activity is greatly diminished. The lemur *Cheirogaleus crossleyi*belongs to a unique group of primates known to hibernate (genus *Cheirogaleus*). It has a characteristic tail rich in white adipose tissue (WAT) that is burnt during hibernation, a period that can endure up to 7 months in Madagascar.
Using capture-mark-recapture techniques to track the same animals in Madagascar, we used RNA-seq to compare gene expression profiles in WAT during three timepoints with different physiological states: pre-hibernation, torpor, and active. This is the first time that a study like this has been performed in a wild population. Because there is no genome available we first generated and evaluated a *de novo* transcriptome assembly for this non-model species. We applied a state of the art pipeline and focus on pathway analysis to assess the biological significance of transcriptional changes in dwarf lemur WAT.

The hibernation signature is characterized by a suppression of lipid biosynthesis, pyruvate metabolism and mitochondrial-associated functions, and an accumulation of transcripts encoding ribosomal components and iron-storage proteins. One of the key molecules is pyruvate dehydrogenase kinase isoenzyme 4 (PDK4), which regulates the shift in fuel economy during periods of resources scarcity. This PDK4 pattern holds true across representative hibernating species from disparate groups, suggesting that the genetic underpinnings of hibernation may be ancestral to mammals.

O-03-OS15

Insights into the population history of the "Hidden Ones": From oral history to genome-wide analysis

Hiba Babiker[1], Abbie Hantgan[1], Johann-Mattis List[1], Jeffrey Heath[2], Russell Gray[1]

[1]Max Planck Institute for the Science of Human History (Germany), [2]University of Michigan (United States)

Africa contains nearly one-third of the world's languages, and its populations have the highest genetic diversity. The depiction of genetic diversity in Africa is critical for reconstructing modern human origins and demographic history. Exploring the linguistic diversity in Africa is reframing current classification system regarding language families and isolates. The sub-Saharan fringe in Western Africa is becoming ripe for extensive research in linguistics, genetics, archaeology and cultural evolution. However, some populations are under-represented in multifaceted studies.

Our research focuses on the Dogon and Bangande populations of the Bandiagara Escarpment in Central Eastern Mali who claim a "Mande" origin and a collective Dogon ethnicity. We used genome-wide data to understand the genetic structure of these populations and the evolutionary forces that shaped their genomes. We also leveraged lexical data representing the languages spoken by these populations to document historical contacts. We used state of the art analyses in population genetics and computational linguistics as well as simulation modelling.

My talk will explore recent findings and will discuss the impact of cultural and geographical barriers on gene flow and population admixture. These findings shed light on the co-evolution of languages and genes. The results highlight the importance of interdisciplinary research in depicting the complex linguistic and genetic histories, hence, decoding unanswered questions in the human history.

O-03-OS16

Sex Differences in Reference Genome Affect Variant Calling and Differential Expression

The human X and Y chromosomes evolved from a pair of homologous autosomes, but today have vastly different gene content and structure. Curiously, despite tremendous sex-bias in human disease, the sex chromosomes are rarely included in genome-wide analyses of human health and disease. One of the reasons for this exclusion is that the X and Y chromosomes don't follow autosomal patterns of inheritance. However, even when they are included, technical biases resulting from aligning all sequences to a single sex-averaged reference genome can result in erroneous mapping to X and Y. I will present results that failing to account for the ancestral sequence similarity between the human X and Y and their evolution can affect variant calling and inference of gene expression.

O-03-OS17

Towards more accurate phylogenomic inference using IQ-TREE

Bui Quang Minh[1, 2], Olga Chernomor[2], Stephen Crotty[2], Heiko Schmidt[2], Diep Thi Hoang[3], Le Sy Vinh[3], Lars Jermiin[1], Huai-Chun Wang[4], Edward Susko[4], Andrew Roger[4], Arndt von Haeseler[2]

[1]Australian National University (Australia), [2]University of Vienna, Medical University of Vienna (Austria), [3]Vietnam National University (Viet Nam), [4]Dalhousie University (Canada)

Evolutionary biologists now reconstruct phylogenetic trees from many genes, even whole genomes, generated by deep sequencing technologies. While such huge amounts of phylogenomic data should enable us to resolve the tree of life, many phylogenomic studies lead to highly contradictory conclusions (i.e., presenting conflicting branching patterns with nearly 100% statistical support each). Here, we will present novel models and methods recently implemented in the widely-used IQ-TREE software (http://www.iqtree.org) towards more accurate phylogenomic inference. To this end, we introduced advanced mixture and site-specific models in both model selection (ModelFinder) and tree search. Moreover, the ultrafast bootstrap allows to reduce the impact of polytomies or severe model violation. Subsequently, applications will be shown to disentangle contentious animal phylogenetic relationships.

O-03-OS18

Inadvertent paralog inclusion impacts phylogenomic relationships and timetree estimates in the Lissamphibia

Karen Siu-Ting[1, 2, 7], Maria Torres-Sanchez[4], Diego San Mauro[4], Mark Wilkinson[5], Davide Pisani[6], Mary J OConnell[3], Christopher J Creevey[2]

[1]Dublin City University (Ireland), [2]Aberystwyth University (United Kingdom), [3]Leeds University (United Kingdom), [4]Complutense University of Madrid (Spain), [5]Natural History Museum, London (United Kingdom), [6]University of Bristol (United Kingdom), [7]Universidad Nac. Mayor San Marcos (Peru)

Increasingly large phylogenomic datasets driven by transcriptomic data availability from non-model organisms allow controversial and unexplored evolutionary relationships in the tree of life to be addressed but it is unclear how inadvertent inclusion of paralogs from this data affects the results retrieved. The Lissamphibia are susceptible to these issues as only a small number of sequenced genomes are available for this hyper diverse group and transcriptomic data is being increasingly applied to resolve their historically conflicting taxonomic hypotheses.

We tested the impact of paralog inclusion on the relationships and timetree estimates of the Lissamphibia using published and *de novo* sequences including 18 amphibian species, from which 2,656 putative orthologous gene families were identified. Further paralog filtering using a novel phylogenetic approach resulted in four increasingly curated datasets which were each used for Bayesian Inference, Maximum Likelihood, Coalescence and Supertree reconstructions of the phylogeny of the Lissamphibia.

We found that paralog inclusion resulted in retrieving conflicting hypotheses within the Lissamphibia (Procera and Batrachia). All methods, except Bayesian with the CAT model, were found to be sensitive to paralogs, that when removed resulted in convergence to the same answer (Batrachia). Furthermore, paralog inclusion resulted in older divergence time estimates within the Lissamphibia, even within groups where no variation in topology was observed.
Our results demonstrate that careful curation of orthologous genes is a critical factor in phylogenomic studies and emphasises the importance of quality over quantity in large-scale studies, particularly for understanding the Amphibian tree of life.

O-03-OS19

Incomplete lineage sorting in mammalian phylogenomics

Nicolas Galtier[1], Celine Scornavacca[1]

[1]CNRS - University Montpellier (France)

Gene trees can differ from species trees if the time between two successive speciation events is not substantially longer than the average coalescence time in the population that split. This process, called incomplete lineage sorting (ILS), has been increasingly invoked as one of the major sources of phylogenetic conflict in phylogenomic datasets, the need to account for ILS in species tree reconstruction being a matter of intense controversy (Gatesy & Springer 2013 PNAS 110:1179,Liu et al 2015 Ann N Y Acad Sci 1360:36). Focusing on full-genome data in placental mammals, we empirically test two assumptions underlying current usage of tree-building methods that account for ILS. We show that in mammals (i) distinct exons from a common gene do not share a common genealogy, and (ii) ILS is only a minor determinant of the existing phylogenetic conflict (Scornavacca & Galtier 2017 Syst Biol 66:112). We introduce a new approach based on the distribution of branch lengths for distinguishing ILS from other sources of conflict, particularly horizontal gene transfer, while controlling for among-loci variation in mutation rate and effective population size. These results shed new light on the prevalence of ILS and conditions of applicability of ILS-aware methods in phylogenomic analyses.

O-03-OS20

Evolution of bacterial communities associated with termites

Thomas Bourguignon[1], Nathan Lo[2], Carsten Dietrich[3, 4], Jan Sobotnik[5], Sarah Sidek[6], Yves Roisin[7], Theodore Alfred Evans[8]

[1]Okinawa Institute of Science and Technology (Japan), [2]University of Sydney (Australia), [3]Max Planck Institute for Terrestrial Microbiology (Germany), [4]Siemens Healthcare (Germany), [5]Czech University of Life Sciences (Czech Republic), [6]National University of Singapore (Singapore), [7]Universite Libre de Bruxelles (Belgium), [8]University of Western Australia (Australia)

Termites are among the few animals that evolved to feed on the most abundant biomolecule on Earth: ligno-cellulose. Termite do not degrade ligno-cellulose alone, and instead do so with the help of the microbes colonizing their gut. The gut microbiome of termites is a complex assemblage, containing upwards of 1000 species of bacteria and archaea. It has fascinated scientists for over a century, yet, how gut bacterial communities assemble on timescales of several million years is not yet fully understood. To address this issue, we examined the gut microbiota of 94 termite species. We built phylogenetic trees for 211 bacterial lineages, which included bacteria from termite gut and their closest relatives from other environments, which we identified using BLAST. Our results show that the majority of termite gut bacteria form large monophyletic clusters, indicating a high level of niche specialization. However, in many clades, termite gut bacteria were interspersed with foreign bacterial lineages, often originating from the gut of other animals. Our results show that "mixed-mode" transmission, combining vertical colony-to-offspring transmission and horizontal colony-to-colony transfer, has been the dominant force shaping the bacterial communities of termite guts.

O-03-OS21

Concordance and divergence of the VLR-based adaptive immune system in jawless vertebrates: Functional and evolutionary implications

Sabyasachi Das[1], Jianxu Li[1], Masayuki Hirano[1], Yoichi Sutoh[1], Jonathan Rast[1], Max D Cooper[1]

[1]Emory University (United States)

Three types of variable lymphocyte receptor genes, VLRA, VLRB and VLRC, encode antigen receptors in the extant jawless vertebrates, the remarkably diverse repertoire of which is generated by insertion of neighboring leucine rich repeat (LRR) sequences into the incomplete germline genes. In both lampreys and hagfish, the B-cell like VLRB+ cells differentiate into VLRB-secreting plasma cells, whereas the alpha-beta and gamma-delta T cell-like VLRA+ and VLRC+ cells express their VLR products solely as cell surface proteins. In comparative studies of lampreys and hagfish, we find that remarkable functional differences have evolved in these lymphocyte lineages. In contrast to their predominance in lampreys, the VLRB+ cells constitute a minority lymphocyte population in hagfish, wherein VLRC+ cells instead predominate. Whereas the germline VLRB gene in hagfish contains a short non-coding intervening sequence, VLRB genes in sea lampreys and Japanese lampreys have very long intervening sequences containing remnants of many transposable elements, which may influence VLRB expression. In keeping with the relatively low numbers of hagfish VLRB+ cells, we find that antibody responses to the same immunogen, sheep erythrocytes, are less robust in hagfish than in lampreys. Thus, even though the fundamental genetic program for differentiation of two prototypic T-like lymphocyte lineages and one B-like lineage is conserved in both jawless and jawed vertebrates, and therefore must have been present in a common vertebrate ancestor, the genetic programs used for fine tuning of the differentiation of these VLR lymphocyte lineages have undergone notable independent evolutionary changes in lampreys and hagfish.

O-03-OS22

Intra-individual analysis of the MHC-dependent TCR repertoire diversity

Ana Teles[1], Benno Woelfing[1], Philip LF Johnson[2], Manfred Milinski[1], Tobias L Lenz[1]

[1]Max Planck Institute for Evolutionary Biology (Germany), [2]University of Maryland College Park (United States)

Antigen-specific recognition of a pathogen by the adaptive immune system involves two crucial steps: antigen-presentation by the highly polymorphic major histocompatibility complex (MHC) and recognition of the antigen:MHC complex by a corresponding T cell receptor (TCR). The TCR diversity necessary for recognition of a broad antigen spectrum is determined during MHC-mediated T cell selection in the thymus. Theory predicts that, at the individual level, MHC diversity is negatively correlated with TCR diversity: On the one hand, higher individual MHC diversity is expected to allow presentation of a larger number of antigens, facilitating antigen recognition. On the other hand, it is also expected to enhance self-antigen presentation, resulting in enhanced negative selection of self-reactive T cells in the thymus and consequently a reduced diversity of the mature T cell repertoire. This immunological conundrum may select for optimal diversity in the copy number-variable MHC, but the quantitative association between individual MHC diversity and individual T-cell repertoire diversity is still unclear.

Using next generation sequencing, we have characterized complete individual TCR&beta repertoire diversity among naive three-spined sticklebacks from a targeted laboratory cross, harboring copy number-variable MHC IIb genotypes ranging from low diversity (1 allele) to high diversity (7 alleles). Allowing for a direct test of the predicted negative association between MHC and TCR diversity, our results shed light on a crucial but still little understood process of the adaptive immune system and the evolution of a delicate balance between pathogen resistance and self-tolerance in vertebrates.

O-03-OS23

Deciphering The Evolution Of Drought Tolerance In populations Of Silver Fir (Abies alba Mill.) Populations Across Switzerland And Southern France

Katalin Csillery[1, 2], Nina Buchmann[3], Bruno Fady[4], Felix Gugerli[2], Otso Ovaskainen[5, 6], Christoph Sperisen[7], Alex Widmer[8]

[1]University of Zurich (Switzerland), [2]Swiss Federal Research Institute WSL (Switzerland), [3]ETH Zurich (Switzerland), [4]INRA Avignon (France), [5]University of Helsinki (Finland), [6]Norwegian University of Science and Technology (Norway), [7]Swiss Federal Research Institute WSL (Switzerland), [8]ETH Zurich (Switzerland)

Climate change increases the frequency and severity of drought events, altering natural selection acting on forest trees. We studied a notoriously complex trait, drought tolerance, in 19 and 16 autochthonous silver fir populations across Switzerland and Southern France, respectively, combining tools from population and quantitative genetics. For the two regions independently, we estimated drift distances between populations based on 357 single-nucleotide polymorphism (SNP) loci using an admixture F-model. Seedling growth, height and intrinsic water-use efficiency (WUE, in Switzerland) or drought stress response (lack of watering, in France) were measured in common garden experiments. WUE was also measured on adult trees in the seed-source populations using stable isotopes from needle tissue. We obtained fine spatial scale historical climatic data using statistical downscaling and propose a test to evaluate if selection is driven by climate. Our results indicate strong selection on height and growth driven by temperature, both in Switzerland and France: trees from warm sites grow faster and taller than those from cold sites. The evolution of drought tolerance was strongly and similarly correlated with growth in Switzerland and France: populations that evolved slow growth and small stature also have higher drought tolerance. This finding was supported by multiple independent lines of evidence, including response to water stress treatment, heritability of WUE and its genetic correlation with growth and height, comparison of WUE in adults and seedlings, and habitat information. Our study illustrates the potential of combining different data types and methods to reveal the presence of drought adapted populations.

O-03-OS24

Genetics of parallel leaf shape evolution in the Mimulus guttatus species complex

Kathleen Gray Ferris[1], Graham Coop[1], Johanna Schmitt[1]

[1]UC Davis (United States)

Parallel evolution or the independent evolution of similar phenotypes in organisms occupying similar environments is strong evidence of adaptation. Whether convergent phenotypes are controlled by the same genetic loci and mutations and therefore whether evolution is predictable at the molecular level is a central question in evolutionary biology. To address this question we examine the genetic and adaptive significance of parallel leaf shape evolution across and within species in the Mimulus guttatus species complex. Lobed and narrow leaves have evolved from the entire round leaves of M. guttatus in M. laciniatus M. nudatus and a polymorphic serpentine M. guttatus population. In addition to having divergent leaf shapes all three of these taxa occur in harsh rocky habitats. We used (1) population genetics to detect local adaptation in leaf shape across altitudinal clines of M. laciniatus (2) phenotypic selection analysis to test whether lobed leaf shape is adaptive in M. laciniatus granite habitat and (3) quantitative trait locus (QTL) mapping and genome wide association to examine whether leaf shape evolution has a parallel genetic basis across three rocky outcrop Mimulus taxa. We found that (1) leaf lobing appears to be adaptive at high elevations across multiple altitudinal clines in M. laciniatus (2) leaf shape is adaptive in rocky outcrops and (3) leaf shape is controlled by overlapping genetic regions in all three Mimulus species. This overlap in QTLs and harsh rocky habitats suggests that parallel genetic evolution is responsible for adaptive leaf shape evolution across Mimulus.

O-04-PG01

Using equivalent information trees to maximize the power of taxon and gene sampling in phylogenetics and phylogenomics

Jeffrey Peter Townssend[1, 2, 3], Zheng Wang[2, 3], Dylan Greaves[2], Christoph Leuenberger[4], Alex Dornburg[5]

[1]Yale University (United States), [2]Yale University (United States), [3]Yale University (United States), [4]University of Fribourg (Switzerland), [5]North Carolina Museum of Natural Sciences (United States)

Despite persistent efforts and remarkable decreases in the cost of sequencing, some of the most vexing problems in evolutionary biology recalcitrantly defy resolution or are resolved by topologies that are strongly supported but incongruent between studies. While there are a variety of factors that can contribute to this incongruence, it has long been recognized that chance convergence and parallelisms play a major role in misleading topological inferences. A common pattern in phylogenomic studies has been incongruence or lack of resolution for nodes characterized by short divergences followed by long subtending branches. This class of phylogenetic problem is notoriously difficult to resolve, as short branches provide little phylogenetic signal while long branches increase the probability of chance convergences, or homoplasy, misleading inference. Resolution requires insight into how to target loci for analyses that will efficiently improve phylogenetic accuracy and precision. Recognition that rates of character change provide key insight into the probability of correct topological inference has yielded some approaches that are either computationally refractory to large phylogenetics or that make topological simplifications that are biased in the case of high taxon sampling. A comprehensive theory would facilitate calculation of probabilities of resolution based on any set of character rates for any taxon-sampling scheme. Here we present an approach that iteratively collapses trees of arbitrarily complex topology to information-equivalents, enabling analysis of the power of datasets of any topological complexity. We apply it to phylogenomic datasets, demonstrating its flexible utility for phylogenetic experimental design in the era of big data.

O-04-PG02

Enhanced phylogenetic resolution of three highly-diverse flowering plant lineages from the Neotropics using Anchored Hybrid Enrichment

Carolina Granados Mendoza[1], Gerardo A. Salazar[1], Susana Magallon[1], Eric Hagsater[2], Stefan Wanke[3], Itzi Fragoso-Martinez[1], Martha Martinez-Gordillo[4], Marie-Stephanie Samain[5], Emily Moriarty Lemmon[7], Alan R. Lemmon[6]

[1]Biology Institute, National Autonomous University of Mexico (Mexico), [2]Mexican Association of Orquideology (Mexico), [3]Botany Institute, Dresden University of Technology (Germany), [4]Faculty of Sciences, National Autonomous University of Mexico (Mexico), [5]Instituto de Ecologia, A. C. (Mexico), [6]Florida State University (United States), [7]Florida State University (United States)

Large-scale sequence data collection for macro- and micro-evolutionary studies has recently been boosted by the application of novel genomic partitioning strategies (GPS), a set of methods focused on simultaneously sequencing selected portions of the genome that are expected to be highly informative for specific phylogenetic scopes. Anchored Hybrid Enrichment (AHE) is a novel and highly efficient GPS for phylogenomics. Its probe design targets the capture of hundreds of loci with different proportions of conserved vs. non-conserved regions, therefore capturing loci that are potentially useful for a wide range of evolutionary depths. Although AHE was first tested in vertebrates, a newly-developed plant-capture probe set was recently applied to angiosperm lineages. The present work demonstrates the application of AHE to three highly-diverse and/or emblematic angiosperm Neotropical lineages: the genus *Epidendrum*, *Salvia* subgenus *Calosphace*, and *Aristolochia* subsection *Pentandrae*. Relationships within these plant groups have been difficult to resolve with conventional Sanger sequenced markers, making them excellent candidates for testing the effectiveness of AHE gene capture and sequencing, and estimating the performance of obtained loci in resolving and supporting phylogenetic relationships along a range of taxonomical depths. Because of their distribution across the angiosperm phylogenetic tree, together these plant lineages demonstrate the universality of the AHE angiosperm probe set.

O-04-PG03

Bayesian species tree estimation under the multispecies coalescent

Ziheng Yang[1]

[1]University College London (United Kingdom)

The multispecies coalescent (MSC) is a simple extension of the single-population coalescent to multiple species, and provides a natural framework for estimating species phylogenies using multi-locus genomic sequence data. It naturally accounts for the random fluctuations of the coalescent process and the resulting gene tree heterogeneity across the genome. Full-likelihood implementations of the MSC have mostly taken the Bayesian approach, using Markov chain Monte Carlo (MCMC) to average over the unknown gene trees and coalescent times, accommodating phylogenetic uncertainties. They are shown to have better statistical performance than two-step summary methods in simulations. However their implementations are challenging because simple Markov chain Monte Carlo algorithms do not scale well and suffer from mixing problems in large datasets of many loci. In this talk, I will discuss our recent efforts to design smart MCMC proposals to improve the mixing efficiency.

O-04-PG04

Genome-wide estimation of the evoutionary history of six baboon species with polymorphism-aware phylogenetic models

Carolin Kosiol[1, 2], Dominik Schrempf[1, 2]

[1]University of St Andrews (United Kingdom), [2]Vetmeduni Vienna (Austria)

The increased availability of sequenced genomes both from closely related species and from individuals of the same species, offers a great opportunity to study the speciation and evolutionary history of populations, provided we can properly model the process of sequence evolution using inter and intraspecific data together.

We have developed a new method called POlymorphisms-aware phylogenetic MOdel (PoMo). It extends any DNA substitution model and additionally accounts for polymorphisms in the present and in the ancestral population by expanding the state space to include polymorphic states. It is a selection-mutation model which separates the mutation process from the fixation process. Thereby, a Moran process is used to model genetic drift. PoMo naturally accounts for incomplete lineage sorting because ancestral populations can be in a polymorphic state.

Recently, we have implemented our method in IQ-Tree software package (Ngyuen et al., 2015) such that we can accurately and time-efficiently estimate the parameters describing evolutionary patterns for phylogenetic trees of any shape (species trees, population trees, or any combination of those for up to 100 taxa with multiple individuals). I will present what can be learned by applying these methods to genome-wide data sites of seven baboon populations about ancestral population history of this species such as new estimates of divergence times and mutation rates, and I will discuss consequences of these results for genome-wide molecular dating.

O-04-PG05

Phylogenomics of Pancrustacea

Andreas Zwick[1], Luisa Teasdale[1], Diana Hartley[1], Blanca Prado[2], Carmen Pozo[2], Hong Shen[3]

[1]CSIRO (Australia), [2]Unidad Chetumal (Mexico), [3]Nanjing Agricultural University (China)

The evolutionary relationships of arthropods have been studied extensively during the past century, but a consensus on higher relationships has remained elusive. Among the many debated hypotheses, the origin of Hexapoda within Pancrustacea ("Crustacea" + Hexapoda) is of particular interest. Advances in sequencing technology have spurred the study of arthropod relationships with ever larger molecular data sets for the last decade, yet, their taxon sampling has been very limited and biased. This is particularly true for transcriptome data sets, and results of independent studies are typically incongruent.

To address questions of pancrustacean evolutionary relationships, we sequenced transcriptomes of pancrustacean groups that were previously poorly or not at all represented, including the elusive Cephalocarida and most orders of Malacostraca. We present first results from Maximum Likelihood analyses of 173 taxa and compare them to other published studies. The assembly and analysis of such large transcriptome data sets pose particular challenges, e.g., the search for orthologous sequences, the alignment of multiple sequences and the detection of contaminants, which we discuss based on our own and published data. Using independent data and multiple analyses results in a single, strongly supported backbone topology, indicative of tracking evolutionary relationships. Furthermore, re-analysis of the published data from the most comprehensive, conflicting transcriptome study also results in this topology, and this new congruence strengthens our phylogenomic hypothesis and hints at analysis problems in the published study.

O-04-PG06

Genome-wide sequence information reveals multiple past hybridizations that shaped the ancestors of wheat

Nadine Bernhardt[1], Jonathan Brassac[1], Xue Dong[2], Eva-Maria Willing[2], Korbinian Schneeberger[2], Benjamin Kilian[3], Frank Blattner R.[1, 4]

[1]Leibniz Institute of Plant Genetics and Crop Plant Research (IPK) (Germany), [2]Max Planck Institute for Plant Breeding Research (Germany), [3]Global Crop Diversity Trust (Germany), [4]German Centre for Integrative Biodiversity Research (Germany)

Triticeae, the tribe of wheat grasses, harbors ~360 species in 20-30 genera including wheat (*Triticum aestivum*) and its wild relatives. Even after decades of research most of the phylogeny of Triticeae is still unresolved. This is due to low barriers against hybridization, which mostly result in allopolyploid hybrid taxa. Recently also homoploid hybrid speciation has been suggested for two species closely related to wheat.

We analyzed the relationships among all 13 diploid species closely related to wheat using two independent approaches designed for retrieving genome-wide sequence information: (1) a target-enrichment approach for 450 low-copy nuclear loci obtained through hybridization probes and (2) genotyping-by-sequencing (GBS). The loci were analysed with Bayesian, likelihood and coalescent-based methods. Pattersons D statistic was used to distinguish incomplete lineage sorting from hybridization as the source of conflicts among loci, and the results were contrasted with our recently published plastid phylogeny.

Our results show that many speciation events occurred contemporaneously between approximately 9 and 5 million years ago. We found that hybridization played a major role in shaping the genomes of the ancestral lineages resulting in several mosaic genomes detectable in the extant lineages. Although the existence of many hybrids complicates the inference of their ancestors, we identified a single lineage that played a major role in several independent introgression events. In contrast, species that diverged more recently (approximately 4 million years ago) do not seem to be affected by hybridizations anymore. During this talk a new scenario of hybrid speciation will be introduced.

O-04-PG07

Systematic error is ubiquitous, and frequently misleads phylogenetic inference

Robert Lanfear[1], Suha Naser[1], Wenqi Zhang[1]

[1]Australian National University (Australia)

Models of sequence evolution are, by necessity, drastic simplifications of reality. These simplifications can mislead phylogenetic inference when the mismatch between the model and the truth, known as systematic error, overwhelms the biological signal in the data. Despite this, it remains uncommon to test for systematic error in empirical studies. As a result we lack a clear understanding of the frequency and importance of systematic error in phylogenetics. We extended some existing tests and used them to measure the extent of systematic error in thousands of aligned loci from 35 published datasets. We show that a substantial proportion of the loci in almost every dataset significantly violate the assumptions of the models that were used to infer the tree. More importantly, we show that in about 25% of cases the inferred phylogeny changes significantly depending on whether the loci which violate the model assumptions are included or excluded. This suggests that we should pay more attention to systematic error if we want to infer accurate phylogenetic trees, particularly in the era of very large datasets. We hope that our publicly-available methods will facilitate this.

O-04-PG08

Stepwise Bayesian Inference of Phylogeny using RevBayes

Sebastian Hoehna[1]

[1]LMU Munich (Germany)

Bayesian inference of phylogeny can, in principle, jointly estimate a species tree, separate individual gene trees, substitution model parameters, divergence times, diversification rate parameters and any other model parameters. However, Bayesian inference under these hierarchical models is feasible computationally only for a small number of loci and/or species. Thus, a traditional hierarchical Bayesian inference approach is not applicable to genomic data. Here I will present a new stepwise Bayesian modelling approach, that divides the model into smaller sub-models (steps). Inference is then performed stepwise on each sub-model using the results of the previous step/analysis, i.e., the posterior distribution, as the data. For example, we first estimate the posterior distribution of unrooted gene trees, and then use this posterior distribution to estimate the posterior distribution of (relative) time-calibrated gene trees. Next, we estimate a species tree using a multi-species coalescent model and use the posterior distribution of time-calibrated gene trees as data. This stepwise approach, as I will show, has several advantages while being mathematically equivalent to a full joint inference: (1) the overall computation time is much lower due to fewer parameter dependencies; (2) easier checking of convergence in each analysis step; (3) faster model selection due to smaller models; (4) better repeatability of each analysis step because of shorter run times; (5) possibility of re-using previous analysis when new models are available; (6) embarrassingly easy ways to parallelize and speed-up each analysis step. I will show some results utilizing our implementation in RevBayes.

O-04-PG09

Effects of site selection and data partitioning schemes in the phylogenomic inference of the sub-family Delphininae (Cetacea), a recent radiation with an elusive topology and inter-lineage reticulation

Andre E Moura[1], Kypher Shreves[4, 1], Michael McGowen[3], Rus Hoelzel[2]

[1]University of Lincoln (United Kingdom), [2]Durham University (United Kingdom), [3]Smithsonian National Musem of Natural History (United States), [4]University of Aberdeen (United Kingdom)

Cetaceans are a particularly interesting model for the study of evolutionary biology and phylogenetic inference, as extant diversity likely results from a recent rapid radiation. Inference of phylogenetic relationships has been notoriously difficult for this group, particularly for dolphins of the sub-family Delphininae. In this study, we used a ddRADseq protocol to investigate the effects of site selection strategies and data partitioning schemes on the phylogenomics inference of this group. We included lineages differentiated at the genus, species, ecotype and population level, thus providing insight at various levels of evolutionary divergence. We tested the effects of different read processing options, sites under selection, GC-rich vs AT-rich regions, partitioned vs unpartitioned concatenated datasets, and different strategies of gene concordance based analyses. Finally we compare the usefulness of tree versus network based approaches in representing such shallow phylogenies. Our results suggest that maximizing variable sites at the expense of missing data/site provides robustness against sources of bias. In this context, removal of sites under selection is important, while partitioning of a concatenated tree improves accuracy mostly at the intra-specific level. Gene-concordance analyses results in a topology mostly consistent with inter-specific inference, while also being more consistent with network analyses and previous population level studies in those lineage that may experience ongoing gene flow. By using this experimental approach to RAD based phylogenomics inference, we produced a high resolution phylogenetic reconstruction of the sub-family Delphininae. Furthermore we reveal complex patterns of cross-lineage hybridization, and the potential adaptive nature of this fast radiation.

O-04-PG10

A new approach to model amino-acid compositional heterogeneity sheds new insight on the origin of mitochondria

Ding He[1, 2], Simon Whelan[2]

[1]Uppsala University (Sweden), [2]Uppsala University (Sweden)

Although the mitochondrial progenitor most likely originated from alphaproteobacteria, pinpointing its closest alphaproteobacterial relative remains a challenging phylogenetic problem. Amino acid compositional heterogeneity across sites (aaCHAS) is one of the most important sources of modeling error when not adequately accounted for. A popular approach for describing aaCHAS is through a mixture model, which assigns each site a probability of belonging to each of a set amino-acid profiles that have been estimated directly from the data (CAT model), or empirically derived from a set of structure-aware protein alignments (e.g. C60 model or the PMSF approximation). These approaches are computationally extremely demanding, which limits their application in the post-genomic era. In this study we present the novel FREDOM model (per-site amino-acid FREquency profiles of protein DOMains), which circumvents direct parameter estimation by using per-site amino-acid profiles based on the profile hidden Markov models of homologous sites on Pfam domains. We apply FREDOM on a carefully vetted phylogenomic dataset of 36 mitochondrial proteins and used information theoretic measures to compare a wide range of different models to investigate the alphaproteobacterial origin of mitochondria. We find that the FREDOM provides better model-fit under maximum-likelihood (ML) framework. Our results show different models recovered different ML trees supporting two alternative hypotheses (Rickettsiales or pan-alphaproteobacteria being the closest mitochondrial relatives), albeit with limited statistical supports. The FREDOM approach offers a new way to model aaCHAS in the phylogenomic analysis that has the potential to better describe protein evolution with negligible computation cost.

SY25: Post-transcriptional modifications: functions, diversity, pathogenesis and evolution (July 10, 16:00–18:30)

O-02-PT01

Antiviral APOBEC Enzymes Drive Tumor Evolution

Reuben s. Harris[1, 2]

[1]Masonic Cancer Center (United States), [2]University of Minnesota (United States)

APOBEC enzymes cause a large proportion of both dispersed and clustered cytosine mutations (kataegis) in many different human cancers including those of the breast, head/neck, cervix, bladder, and lung. The antiviral enzyme APOBEC3B accounts for the dominant fraction of these mutations, and the related enzyme APOBEC3H likely also contributes. APOBEC3B expression levels correlate with mutation loads and poor outcomes for patients with estrogen receptor positive breast cancer. While targeted therapies are available to treat these tumors, acquired drug resistance often develops. Clinical data strongly implicate APOBEC3B in tumor recurrence and resistance to tamoxifen therapy after surgical resection of primary tumors. These results are corroborated by experiments with murine xenograft models in which APOBEC3B knockdown and overexpression slow down and accelerate drug resistance, respectively. The clinical and experimental results combine to demonstrate that APOBEC3B drives resistance to targeted therapies in breast cancer. These results are likely to be broadly applicable to other cancer types and targeted therapies as APOBEC enzymes can be regarded as dominant and near-ubiquitous sources of mutagenic fuel for tumor evolution.

O-02-PT02

Post-transcriptional modifications: Adaptations or cellular errors?

Jianzhi Zhang[1]

[1]University of Michigan (United States)

The powerful next-generation sequencing has enabled the identification of many post-transcriptional modifications at the transcriptomic scale, allowing an investigation of their potential functions and adaptive values. Analyzing published data of A-to-I RNA editing, C-to-U RNA editing, m6A modification, and alternative polyadenylation from humans and other mammals, I show that most modification events likely manifest cellular errors and are non-adaptive. I also provide evidence that the previously reported positive selection for A-to-I RNA editing in coleoids (squids, octopuses, and cuttles) was to promote the restoration of ancestral functions instead of the generation of novel functions. Together, these findings require a paradigm shift in functional and evolutionary studies of post-transcriptional modifications.

O-02-PT03

The role of innate immune mechanisms in molecular evolution and pathogenesis of flaviviruses

Helen Piontkivska[1, 4], Noel-Marie Plonski[1, 4], Madeline Frederick[1], Dhani Prakoso[3], Kelli Lynn Barr[3], Michael Masao Miyamoto[2], Marta L. Wayne[2], Maureen Therese Long[3]

[1]Kent State University (United States), [2]University of Florida (United States), [3]University of Florida (United States), [4]Kent State University (United States)

Since 2015 - when the link between Zika Virus (ZIKV, Flavivirdae), microcephaly and fetal demise was discovered in Brazil - the interest in the evolutionary mechanisms that may be responsible for virus-mediated pathogenesis has significantly expanded. Of particular concern are neurodevelopmental abnormalities observed in the congenital Zika syndrome (CZS). However, neurological symptoms of Guillain-Barre syndrome (GBS), reported in adult ZIKV infections, also known to occur in West Nile Virus (WNV) and Dengue virus (DENV) infections. We recently showed that one arm of the innate immune response, RNA editing of adenosines (abbreviated as ADAR editing), which is also responsible for transcriptome diversification, left its footprints in ZIKV genomes. We observed that ADAR substitutions are underrepresented at the ADAR-resistant GA dinucleotides, with the particular enrichment of resistant GA dinucleotides on the positive (but not negative) strand, which indicates that the former is under stronger purifying selection than the latter. Moreover, the changes are spatially and temporally clustered (as expected of ADAR editing) for certain evolutionary lineages. ADAR mutagenesis can be linked to viral codon usage. We are now examining molecular evolutionary patterns in WNV and DENV genomes, to elucidate the influence of the host immune response on viral evolution. Viral-mediated ADAR editing influence on the host transcriptomes is also been considered, providing evolutionary insights into the host-pathogen interactions.

O-03-PR01

Evolutionary predictions from biophysical models

Michael Lassig[1]

[1]University of Cologne, Institute for Biological Physics (Germany)

Predictions of future evolutionary processes have recently been developed for a number of systems, including fast-evolving pathogens and cancer populations. Biophysical phenotypes of molecular host-pathogen interactions have proven to be informative for predictions. Such phenotypes can be learned from time-resolved sequence data and can be built into fitness models for predictions of pathogen evolution and host response. Conversely, successful predictions can improve our understanding of the underlying cell biology. From the case studies discussed in this talk, I will attempt to distill emerging concepts for predictive analysis of fast-evolving systems.

O-03-PR02

Prediction in optimal immune systems

Thierry Mora[1]

[1]CNRS and ENS Paris (France)

Immune systems protect us against a great diversity of largely unpredictable pathogens. What are the general design principles of immune systems that allow them to perform this complex
task efficiently and reliably, by making predictions about the future state of the pathogenic environment? I will explore this question by considering optimal immune designs at different scales of organisation. At the individual scale, how should the repertoire of specificities self-organise to best predict future infections? And at the population level, what are the optimal modes of immunity (adaptive, CRISPR, etc.) allowing maximal long-term growth, depending on the pathogen statistics?

O-03-PR03

Evolution of the adaptive immune system in response to cancer

Yuval Elhanati[1], Zachary Sethna[1], Curtis Callan[1]

[1]Princeton (United States)

The adaptive immune system protects against the vast diversity of pathogenic threats by maintaining a corresponding diversity of T-cells with different surface receptors, which bind to a pathogenic molecule thus mobilizing the immune system. The initial ensemble of cells is produced in a stochastic generation process, which is then subjected to selection pressures based on the functionality of the receptors. This amounts to a unique evolutionary process, termed somatic evolution, that can react to previously unknown threats by selecting cells with effective receptors from the generated diversity.

The dynamic nature of this process is essential for the immune reaction to cancer cells, which constantly produce new mutations acting as possible targets for T-cells. Today, we can follow the evolution of the T-cells ensemble using Next Generation Sequencing of blood samples from donors and patients. In my talk, I will show how we can analyze samples using statistical models, to infer and characterize the distribution of T-cells in different clinical situations. This in turn enable us to compare the immune system from different time points, patients and compartments, and model the responsive selection to the cancer. Eventually, following the mutations in the tumor at the same time, we can tie the evolution of the T-cell ensemble with that of the cancer, thus providing insights on immune reaction to cancer and possible treatments.

O-03-PR04

High-resolution lineage tracking of laboratory yeast populations over 1000 generations

Ivana Cvijovic[2, 1], Alex N Nguyen Ba[1, 2], Jose I Rojas Echenique[1, 2], Katherine R Lawrence[3], Artur Rego-Costa[2], Sasha F Levy[4], Michael M Desai[1, 2]

[1]Harvard University (United States), [2]Harvard University (United States), [3]MIT (United States), [4]Stony Brook University (United States)

Large, rapidly adapting microbial populations have remarkably rich evolutionary dynamics. A multitude of beneficial mutations occur simultaneously and compete in the population, the majority never rising to substantial frequencies. Yet the fates of those that do depend crucially on the benefits of the future mutations that arise in the population. However, the observation and monitoring of these dynamics has been challenging: traditional metagenomic sequencing methods allow us to observe only mutations that reach relatively high frequencies (>1%), whereas newer barcoding approaches are limited to the earliest phases of adaptation, before the first beneficial mutations rise to high frequencies. We have developed a new high-efficiency genomic re-barcoding technique that enables us to repeatedly introduce high-diversity DNA barcodes into yeast laboratory populations over the course of evolution. By sequencing the barcode locus over time, we can track the frequency of lineages founded by unique barcodes and their descendants, as they change due to drift and selection, and diversify through new mutations. The data reveal a remarkable Russian-doll-like pattern in the evolutionary dynamics, characterized by nested clonal interference events at a broad range of frequency scales. Using a combination of theoretical and statistical approaches, we can infer the fitnesses of individual lineages in a range of evolutionary conditions, in constant and in alternating environments. This enables us to analyze how the distribution of fitness within the population arises from clonality and changes over time, and to quantify how competition between high-fitness lineages determines the limits of predictability.

O-03-PR05

Clonal diversity accelerates the evolution of antimicrobial resistance

Ignacio Vazquez-Garcia[1, 2], Elisa Alonso-Perez[3], Jing Li[4], Johan Hallin[4], Magda Reis[5], Gianni Liti[4], Jonas Warringer[3, 6], Ville Mustonen[1, 7]

[1]Wellcome Trust Sanger Institute (United Kingdom), [2]University of Cambridge (United Kingdom), [3]University of Gothenburg (Sweden), [4]Universite Cote d'Azur (France), [5]University of Dundee (United Kingdom), [6]Norwegian University of Life Sciences (Norway), [7]University of Helsinki (Finland)

Clonal heterogeneity in microbial infections and in tumours is one of the major causes of treatment failure with antimicrobial or chemotherapy drugs. Here we study the clonal evolution of genetically heterogeneous and homogeneous populations following antimicrobial exposure, using budding yeast (*Saccharomyces cerevisiae*) as a model system. Using diverged strains, we generated a recombinant cross with unique haplotypes, and isolated single individuals from this cross. We carried out selective inhibition with antimicrobials in 5,760 populations over ~960 generations, under both constant and dynamically increasing dosage. We observed recurrent adaptive events across populations by tracking the dynamics of mutations using whole-genome sequencing, and phenotyping their fitness effects with real-time colony imaging. To discern driver mutations from hitchhiking passengers, we compared independent replicates that followed parallel adaptive paths to a given selection pressure. Replicate populations derived from the same founder systematically acquired parallel mutations in similar genes, suggesting that they explore common regions of the fitness landscape. The rate at which populations search the fitness space differs based on the ensemble diversity. On average, genetically diverse populations required fewer driver mutations to adapt than genetically homogeneous populations, though the number of drivers was highly variable between replicates. Conversely, genotypes in isolation typically became entrenched in a local fitness maximum and required similar numbers of drivers to adapt. These results imply that the predictability of molecular mechanisms of antimicrobial resistance depends on the degree of genetic diversity in a population, and suggest that genetic diversity may be a predictor of drug resistance.

O-03-PR06

Historical contingency shapes genome-wide diversity after antibiotic-driven bottlenecks

Alejandro Couce[1, 2], Andre Birgy[2], Coralie Pintard[2], Olivier Tenaillon[2]

[1]Imperial College London (United Kingdom), [2]Universite Paris Diderot (France)

Adapting microbial populations are often composed by transient ensembles of competing clones carrying multiple beneficial mutations. This phenomenon not only alters evolutionary dynamics but also implies that asexual populations harbour more genetic variation than previously considered. Despite its importance, little is still known about how to predict the amount of variation maintained in a population after an environmental change (e.g., antibiotic therapy). Here we address this question by tracking the diversity of spontaneous resistance mutations emerged as a by-product of long-term bacterial evolution to high temperature (42.2C). We first show that loss-of-function mutations in the glycerol transport-encoding gene glpT confer both resistance to fosfomycin and adaptive benefits in the drug-free environment in which the bacteria evolved. Moreover, by using targeted deep sequencing, we documented that multiple independent loss-of-function alleles arose recurrently on different competing backgrounds. We found lineages that retained high levels of diversity during almost 2,000 generations, while others are dominated by a single allele since very early on in the experiment; revealing that the degree of diversity at this specific locus is largely driven by historical contingency. Crucially, using whole-population whole-genome sequencing, we show that variation at the resistance-determining locus is positively correlated with variation elsewhere in the genome, irrespective of the effective population size of the resistant sub-population. This suggests that the capacity of a population surviving a drug-induced bottleneck to cope with future challenges (i.e. second-line therapy) is determined by past chance events, rendering pathogen evolution more unpredictable than previously anticipated.

SY27: Quantitative genetics of developmental evolution (July 11, 15:00–17:00)

O-03-QG01

Evolutionary Quantitative Genetics of Zebrafish Development

Katrina McGuigan[1]

[1]The University of Queensland (Australia)

Early-life events are expected to affect phenotypic variation in adults both through direct pleiotropic effects of alleles on temporally disjunct phenotypes, and through indirect, carry-over effects. Although there has long been interest in understanding the genetic relationship between embryo/juvenile and adult phenotypes, characterising early life phenotypes on large numbers of individuals is logistically challenging. Most information that we have is for size (or growth), the same recognisable phenotype expressed at different life stages. However, to understand adult phenotypic variation, and to understand the evolutionary dynamics of populations generally, we need better insight into how genetic variation in early developmental phenotypes affects adult fitness. Zebrafish are ideal for studying the quantitative genetics of development. Large sample sizes are obtainable, and, because embryos develop in transparent chorion, individuals can be directly observed from fertilisation. I will present results from a mutagenesis experiment in zebrafish. These data reveal that unobserved genetic variation in timing across the life cycle underlies variation in an ecologically important adult trait, swimming speed of adult males. Variation in adult swimming speed is genetically correlated with embryo viability, but the extent to which genetic variants with less extreme fitness effects affect phenotypic variation across life stages remains to be determined. Swimming speed is sometimes used as a proxy for fitness, and these data highlights the fact that a focus on causal associations with fitness at only the adult life-stage is likely to obscure genuine understanding of the processes acting to determine phenotypic variation of populations.

SY27: Quantitative genetics of developmental evolution (July 11, 15:00–17:00)

O-03-QG02

Characterising patterns of selection and the distribution of fitness effects using single-cell open chromatin data

David Garfield[1], Dermot Harnett[2]

[1]Humboldt University (Germany), [2]Berlin Institute of Health (Germany)

Selection acting on regulatory DNA plays an important role in evolution, but is difficult to quantify due to the challenges of delineating and annotating non-coding elements. Here, we make use of single-cell open chromatin profiling (via sci-ATAC-Seq) to construct a cell-type specific atlas of regulatory DNA at multiple stages of Drosophila development. By coupling this atlas to transcription factor occupancy data and population genetic analyses, we are able to quantify selective history and the fitness consequences of new mutations as a function of how regulatory DNA is used across tissues and developmental time points. We are also able to characterise the frequency of adaptive substitutions both within individual classes of regulatory element as well as sets of transcription factor binding sites. These analyses point to three present conclusions: 1) Adaptive substitutions are relatively common in early acting regulatory elements and in the binding sites of key developmental transcription factors (e.g. Zelda). 2) The frequency of substitutions in constitutively open gene promoters is vastly higher than would be predicted by the pleiotropic constraints that appear to influence the evolution of more distal regulatory elements. 3) Evolutionary rates in general appear to be higher in many classes of promoter than in distal regulatory elements, suggesting a lower frequency of adaptive substitution across the non-coding genome than previously estimated.

O-03-QG03

Somatic genome rearrangements in Oxytricha trifallax - a single-cell model for nuclear differentiation and development

Rafik Neme[1], Jaspreet S Khurana [1], Laura F Landweber [1, 2, 3]

[1]Columbia University (United States), [2]Columbia University (United States), [3]Columbia University (United States)

The ciliate *Oxytricha trifallax* undergoes extensive programmed DNA rearrangements to remodel a copy of its germline nucleus into a somatic nucleus during development, all within a single cell. This process requires the excision of more than 90% of its original content, starting with megabase-sized chromosomes and producing short gene-size chromosomes. In approximately 20% of somatic chromosomes, the precursor fragments are "scrambled", meaning that they have a different order or orientation relative to the final products.

Previous work from our group has described in detail the germline and somatic genomes (precursor and product, respectively), and probed the molecular mechanisms that guide the retention, rearrangement, and deletion of germline genome fragments. However, the temporal dynamics of genome rearrangements have only been studied for a few individual loci.

In this study, we used paired-end Illumina sequencing to capture snapshots of the rearranging genome across developing populations of *O. trifallax* during post-zygotic nuclear maturation. We present an overview of the parallel, developmental trajectories of thousands of differentiating germline loci, as they are excised, descrambled, and edited to produce mature, somatic chromosomes. Further, our study enables the direct quantification of errors at each stage, which in turn highlights which loci and timeframes are most sensitive or permissive during development, and when RNA-guided DNA repair is able to correct most of these errors. The temporal distribution of rearrangement pathways offers an unprecedented view into a highly orchestrated process of somatic nuclear development, that produces an end result of thousands of precisely-rearranged, telomere-capped somatic chromosomes.

O-03-QG04

Fishing for genes driving tooth evo-devo exclusive of the vertebrate jaw

Julia C Boughner[1], Gareth J Fraser[2], Nasim Rostampour[1], Cassy M Appelt[1], Alexandre Thiery[2], Aunum Abid[1]

[1]University of Saskatchewan (Canada), [2]University of Sheffield (United Kingdom)

Vertebrate teeth and jaws are incredibly diverse, but must still fit and function together for an animal to eat and survive. Yet, quixotically, teeth and jaws have distinct evolutionary and developmental (evo-devo) origins, while their formation requires many of the same genes. For this reason alone, the evo-devo genetics that regulate tooth macroevolution and morphogenesis independently of the jaw skeleton have remained unclear. Now, by using a toothless p63-null mouse model with normal jaw formation, we are teasing out the genes integral for tooth but not jaw development. Our microarray, in situ hybridization and immunohistochemistry studies done at embryonic days 10-14 discovered for the first time new genes, including Cbln1, Fermt1, Pltp and Prss8, expressed in mouse tooth epithelium. Next, we compared expression of these newly revealed genes in mouse teeth to gene expression in spotted catshark teeth, characterized using RNA-Seq, in situ hybridization and immunohistochemistry data. Several genes expressed in mouse, including Cbln1, Cldn23, Pltp and Prss8, were also expressed in catshark tooth organs, notably the successional lamina. These shark expression data implicate a p63-driven gene regulatory network in not only vertebrate tooth development but also tooth replacement. The expression of the same genes in mouse and shark teeth suggests that this p63-driven regulatory network is deeply conserved among living vertebrates, and has ancient origins at least as old as the first sharks. Lastly, this p63 gene regulatory network is also exciting because it may facilitate the macroevolution of the dentition without necessitating change in the jaw skeleton.

SY27: Quantitative genetics of developmental evolution (July 11, 15:00–17:00)

O-03-QG05

Genetic Basis of Natural Variations in Germline RNAi in C. elegans

Han Ting Chou[1], Annalise B. Paaby[1]

[1]Georgia Institute of Technology (United States)

The development of *C. elegans* is precise and stereotyped, including patterns of cell division in early embryogenesis. Nevertheless, natural genetic variation in wild-type isolates can cause dramatic differences in phenotype following single-gene perturbations, indicating that different wild-type genotypes harbor functional variation in critical gene networks. These same wild isolates also show extreme variations in the efficacy of germline RNAi. What are the genetic, molecular and cellular mechanisms that govern these differences? And how do they evolve when stabilizing selection ensures that phenotypic development remains stable and stereotyped?

Using the assembler DISCOVAR *de novo*, we assembled genomes of 7 wild isolates from libraries of Illumina 250 bp paired-end reads and MinION long reads. These data further unraveled hyper-variable genomic regions to enable investigation on the genetic background underlying important phenotypic variations. In order to identify candidate genes responsible for germline RNAi variations, we performed Bulked Segregant Analysis between the RNAi resistant strain QX1211 and the laboratory reference strain N2. Differences in germline RNAi sensitivity were measured using a semi-automated pipeline for high-throughput single-molecule FISH that we established to quantitatively score the gene expression during early embryogenesis. By characterizing the temporal and spatial heterogeneities of mRNA transcript numbers in the first few cell divisions, we can connect sub-cellular phenotypes to early embryonic pathway functions and germline RNAi.

SY27: Quantitative genetics of developmental evolution (July 11, 15:00–17:00)

O-03-QG06

The genetic basis of evolutionary transitions in early development

Christina Zakas[1], Matthew Rockman[1]

[1]New York University (United States)

Phenotypic evolution in animals is constrained by the mechanics of early development. Large-scale evolutionary changes are initially shaped by developmental program, where simple trade-offs can ultimately result in a vast spectrum of physiological, morphological, and ecological differences. How do these major transitions in development occur? The polychaete annelid *Streblospio benedicti* provides a unique opportunity to use forward genetics to experimentally dissect a major transition in animal development. *S. benedicti* produces two distinct offspring types that differ in egg size, early development, and larval morphology. Thus it is an ideal genetic model for the evolutionarily common transition between indirect and direct development. Using genetic crosses between these types, we constructed the first annelid genetic map, which reveals the distribution of genetic factors affecting a suite of genetically separable developmental phenotypes. Because early development is strongly influenced by maternal effects, our cross design disentangles maternal and zygotic genetic effects and shows that a transition from indirect to direct development requires contributions from both the zygotic and maternal genome; an increase in egg size alone is not sufficient to change development mode. By identifying the loci responsible for early developmental phenotypes, we begin to uncover how the transition from indirect to direct development proceeds.

SY28: Selection on complex traits: reuniting quantitative and population genetics (July 12, 12:30–15:30)

O-04-SC01

A population genetic interpretation of complex trait architecture in humans

Human genome-wide association studies (GWASs) are revealing the genetic architecture of anthropomorphic and biomedical traits, i.e., the frequencies and effect sizes of variants that contribute to heritable variation in a trait. To interpret these findings, we need to understand how genetic architectures are shaped by basic population genetic processes?notably, by mutation, natural selection, and genetic drift. Because many complex traits are subject to stabilizing selection and genetic variation that affects one trait often affects many others, we model the genetic architecture of a focal trait that arises under stabilizing selection in a multidimensional trait space. We solve the model at steady state, and find that the distribution of variances contributed by loci identified in GWASs should be well approximated by a simple functional form that depends on a single parameter: the expected contribution to genetic variance of a strongly selected site affecting the trait. This prediction fits the findings of GWASs for height and body mass index (BMI) well, allowing us to make inferences about the degree of pleiotropy and mutational target size for these traits. Our findings help to explain why the GWAS for height explains more of the heritable variance than the similarly sized GWAS for BMI and to predict the increase in explained heritability with study size. Considering the demographic history of European populations, in which these GWASs were performed, we further find that most of the associations they identified likely involve mutations that arose during the Out-of-Africa bottleneck at sites with selection coefficients around s=0.001.

SY28: Selection on complex traits: reuniting quantitative and population genetics (July 12, 12:30–15:30)

O-04-SC02

Anterior pituitary transcriptome suggests differences in adrenocorticotropic hormone release in tame and aggressive foxes

Anna V Kukekova[1], Hekman P Jessica[1, 2], Jennifer L Johnson[1], Whitney Edwards[3], Anastasiya V Vladimirova[4], Rimma G Gulevich[4], Anastasiya V Kharlamova[4], Yury Herbeck[4], Gregory M Acland[5], Lori T Raetzman[3], Lyudmila N Trut[4]

[1]University of Illinois in Urbana-Champaign (United States), [2]MIT (United States), [3]University of Illinois in Urbana-Champaign (United States), [4]Russian Academy of Sciences (Russian Federation), [5]Cornell University (United States)

Domesticated species exhibit a suite of behavioral, endocrinological, and morphological changes referred to as domestication syndrome. These changes may include a reduction in reactivity of the hypothalamic-pituitary-adrenal axis, specifically reduced adrenocorticotropic hormone release from the anterior pituitary. To investigate the biological mechanisms targeted during domestication, we investigated gene expression in the pituitaries of experimentally domesticated foxes, *Vulpes vulpes*. RNA was sequenced from the anterior pituitary of six foxes selectively bred for tameness and six foxes selectively bred for aggression. Expression, splicing, and network differences identified between the two lines indicated the importance of genes related to regulation of exocytosis, specifically mediated by cAMP, organization of pseudopodia, and cell motility. These findings provide new insights into biological mechanisms that may have been targeted when these fox lines were selected for behavior, suggesting importance of hormone release mechanisms in regulation of stress response and animal domestication.

O-04-SC03

New Insights into the Genetic Basis and Evolutionary History of Lactase Persistence in Africa

Alessia Ranciaro[1], Michael C. Campbell[2], Elizabeth Eyerman[1], Simon Thompson[1], William Beggs[1], Sununguko Wata Mpoloka[3], Gaonyadiwe George Mokone[4], Thomas Nyambo[5], Dawit Wolde Meskel[6], Gurja Belay[6], Meredith Yeager[7], Stephen Chanock[8], Sarah A Tishkoff[1, 9]

[1]University of Pennsylvania (United States), [2]Howard University (United States), [3]University of Botswana (Botswana), [4]University of Botswana School of Medicine (Botswana), [5]Muhimbili University of Health and Allied Sciences (Tanzania), [6]Addis Ababa University (Ethiopia), [7] National Cancer Institute (United States), [8]Leidos Biomedical Research Inc (United States), [9]University of Pennsylvania (United States)

Lactase persistence (LP) is a recent adaptive trait in humans that evolved mainly in populations practicing milk production. Prior studies of African populations have identified novel common variants in intron 13 of the *MCM6* gene (a known candidate enhancer region for *LCT*) associated with LP primarily in pastoralists. While these variants account for a large proportion of the phenotypic variance, there may be additional polymorphisms that contribute to the LP trait. To test this hypothesis, we resequenced 555 bp of intron 13 of *MCM6* in 1588 African individuals from diverse populations in Ethiopia, Tanzania and Botswana. We also genotyped these same individuals using the Illumina Omni 5M SNP array, and integrated our sequencing data with the 5M genotype data to detect signatures of selection. In addition, we performed the lactose tolerance test for a subset of 457 individuals. Based on our analyses, we identified for the first time the presence of the C-14010 and G-13915 variants in the Hadza hunter-gatherers from Tanzania. We also observed the G-13915 variant in the Burunge and Maasai populations from Tanzania, which has never previously been reported. Furthermore, we confirmed the presence of the G-13907 and the G-13915 alleles in Ethiopia. In addition, we detected a strong association between LP and genomic regions on several chromosomes in the Maasai pastoralists based on Fisher's Exact Test, suggesting candidate regions that might contribute to LP in this population. Overall, this genome-wide study offers new insights into the genetic basis and evolutionary history of LP in Africa.

SY28: Selection on complex traits: reuniting quantitative and population genetics (July 12, 12:30–15:30)

O-04-SC04

Paleo-population genetics: Illuminating the role of selection in shaping human diversity

Yassine Souilmi[1], Raymond Tobler[1], Fernando Racimo[2], Christian Huber[1], Angad Johar[1], Iain Mathieson[3], Murray Cox[5], Graham Gower[1], Matthew Williams[1], Matthew Robinson[4], Joao Teixeira[1], Alan Cooper[1]

[1]University of Adelaide (Australia), [2]University of Copenhagen (Denmark), [3]University of Pennsylvania (United States), [4]University of Lausanne (Switzerland), [5]Massey University (New Zealand)

Ancient human DNA studies have primarily focused on the questions surrounding ancestral migrations, admixture, phylogeography, and demography. To date, however, these studies have not fully explored the role of adaptation in shaping modern human diversity. With the number of publicly available ancient human genomes surpassing 1000 in 2018, it is now possible to utilise a range of population genetic approaches to elucidate the genetic history of multiple human phenotypes.

Using a comprehensive aDNA database, the Online Ancient Genome Repository (OAGR; oagr.org.au), we utilised ~300 ancient Eurasian genomes to explore human phenotypic evolution from the late Pleistocene to the present, a period covering the major socio-cultural transitions in human history. We performed a series of selection screens on Mendelian and polygenic traits, building a comprehensive spatiotemporal map of human adaptation over our recent history. Our research provides a unique window into the factors that have shaped modern human diversity and pathology.

SY28: Selection on complex traits: reuniting quantitative and population genetics (July 12, 12:30–15:30)

O-04-SC05

Identifying natural selection by constructing genome-wide genealogies

Leo Speidel[1], Simon Myers[1]

[1]University of Oxford (United Kingdom)

All observed genetic variation can be traced back to a genealogy, which records historical recombination and coalescence events. However, the reconstruction of genealogies has been very difficult due to the enormous number of possibilities and weak information for the correct genealogy. As a consequence, existing methods usually scale to no more than tens of samples. We have developed a method for estimating genome-wide genealogies for thousands of samples, and applied it to 2478 present-day humans from 26 populations in the 1000 Genomes Project. We estimated historical population sizes and identified split times between all pairs of populations. In agreement with previous studies, we have identified evidence of Neanderthal introgression in samples from Eurasia, as well as historical mutation rate fluctuations consistent with population structure. In this talk, we will focus on detecting signals of positive selection in the estimated genealogy, which can appear as bursts of coalescent events following an advantageous mutation. We define a p-value for selection evidence that quantifies how quickly a mutation has spread. We obtain a p-value for almost all biallelic SNPs in the 1000 Genomes Project dataset. We confirm evidence of strong positive selection in well-known targets, such as the Lactase region in Europeans and the EDAR region in East Asians. In addition, we find an enrichment of GWAS hits among SNPs with low selection p-value, suggesting a wide-reaching effect of positive selection in functional parts of the human genome. Finally, we identify multi-allelic traits likely to be under positive selection in different groups.

SY28: Selection on complex traits: reuniting quantitative and population genetics (July 12, 12:30–15:30)

O-04-SC06

Polygenic adaptation in response to a sudden change in the environment

Laura Katharine Hayward[1], Guy Sella[2, 3, 4]

[1]Columbia University (United States), [2]Columbia University (United States), [3]Columbia University (United States), [4]Columbia University (United States)

Multiple lines of evidence suggest that adaptive changes to the genome often involve traits that are highly polygenic. Yet the response to selection on polygenic traits is poorly understood, limiting our ability to find and interpret signatures of polygenic adaptation in population genetics data. To address this question, we study the phenotypic and genetic response to selection on a quantitative trait under stabilizing selection, after a change in environment induces a change in the trait's optimal value. We find that the phenotypic response displays one of two qualitative behaviors, depending on population genetic parameters. When mutations affecting the trait have small effect sizes and arise at high rates, the genetic variance in the population remains approximately constant and the mean phenotype approaches the new optimum at an exponential rate that depends on this variance (Lande, 1976). But when mutations have large effect sizes or arise at lower rates, the phenotypic response is more complex. In turn, changes in allele frequencies underlying the phenotypic response exhibit vastly different short and long-term behaviors. The rapid approach to the new optimum is mostly driven by alleles with large effect sizes. However, over longer time scales, the contribution of these alleles to the change in mean phenotype almost entirely disappears, and the shift in phenotype is taken over by alleles with relatively small effect sizes, some of which eventually fix. One implication of these dynamics is that fixations resulting from polygenic adaptation should have negligible effects on neutral diversity levels at linked sites.

SY28: Selection on complex traits: reuniting quantitative and population genetics (July 12, 12:30–15:30)

O-04-SC07

A dual field approach uncovers the complex genomics of local adaptation in Swedish A. thaliana.

Daniele Filiault[1], Benjamin Brachi[2, 5], Svante Holm[3], Caroline Dean[4], Joy Bergelson[2], Magnus Nordborg[1]

[1]Gregor Mendel Institute (Austria), [2]University of Chicago (United States), [3]Mid Sweden University (Sweden), [4]John Innes Centre (United Kingdom), [5]INRA Bordeaux (France)

While quantitative genetic approaches such as genome-wide association mapping (GWAS) are excellent tools for uncovering the genetics of complex traits in general, we need a better understanding of the adaptive value of this phenotypic and genetic variation in order to facilitate links between quantitative and population genetics. Therefore, we undertook two sets of field experiments in Sweden in order to explore the adaptive genomic landscape in *Arabidopsis thaliana*. In the first set of experiments, we carried out common garden experiments including 200 *A. thaliana* accessions grown at four sites over two years. The second set was a suite of multiyear select-and-resequence experiments using the same 200 accessions at sites in Northern and Southern Sweden. Both experiments uncovered clear evidence of local adaptation and demonstrated that fitness is a highly complex trait in field settings. A large number of variants were associated with fitness differences in both experiment types, and some of these variants could be linked to adaptive phenotypes using GWAS. However, many variants were unique to select-and-resequence studies, possibly reflecting the more natural growth conditions experienced by plants in these settings. Current work is focused on understanding the population genetics and evolutionary history of the variants that we have identified in order to explore the origins and distribution of adaptive variation. This set of experiments therefore represents a uniquely comprehensive effort to link phenotypes, genotypes, and local adaptation on a genome-wide scale.

SY28: Selection on complex traits: reuniting quantitative and population genetics (July 12, 12:30–15:30)

O-04-SC08

The effects of demography and genetic architecture on the neutral distribution of quantitative traits

Evan Koch[1]

[1]University of Chicago (United States)

Neutral models for quantitative trait evolution have been useful for identifying phenotypes under selection in natural populations through goodness-of-fit tests. Such models nearly always start by assuming phenotypes are normally distributed, but traits such as gene expression levels are likely violate this assumption in part due to sparse genetic architectures. We develop and analyze a neutral model for quantitative traits that makes minimal assumptions about the genetic architecture or the distribution of coalescent times and hence the structure or demography of the population. This is done from a coalescent perspective by extending a model developed by Schraiber and Landis (2015) whose central result, the characteristic function for the distribution of trait values is shown here to be a special case of a more general theory. We then demonstrate how deviations from the normal distribution depend on both demography and the genetic architecture. In particular, population growth exacerbates deviations, while bottlenecks reduce them. The normal distribution itself arises in an infinitesimal limit as the number of loci gets large and the effect size of mutations becomes small. Finally, we investigate the impact of sparse genetic architectures on quantitative genetic tests for selection based on population differentiation (i.e. the Qst vs Fst framework) and the possibility of inferring genetic architecture from trait values and genomic data.

SY28: Selection on complex traits: reuniting quantitative and population genetics (July 12, 12:30–15:30)

O-04-SC09

Evidence for stabilizing selection at pleiotropic loci for human complex traits

Emily S Wong[1], Steve Chenoweth[1], Mark Blows[1], Joseph Powell[2, 3]

[1]The University of Queensland (Australia), [2]The University of Queensland (Australia), [3]The University of Queensland (Australia)

How genetic variation contributes to phenotypic variation is a central question in genetics. Association signals for a complex trait are found throughout the majority of the genome suggesting much of the genome is under some degree of genetic constraint. Here, we develop a intraspecific population genetics approach to define a measure of population structure for each single nucleotide polymorphism (SNP). Using this approach, we test for evidence of stabilizing selection at complex traits and pleiotropic loci arising from the evolutionary history of 47 complex traits and common diseases. Our approach allowed us to identify traits and regions under stabilizing selection towards both global and subpopulation optima. Strongest depletion of allelic diversity was found at disease loci, indicating stabilizing selection has acted on these phenotypes in all subpopulations. Pleiotropic loci predominantly displayed evidence of stabilizing selection, often contributed to multiple disease risks, and sometimes also affected non-disease traits such as height. Risk alleles at pleiotropic disease loci were more concordant than expected by chance suggesting that stabilizing selection acting on pleiotropic loci is amplified through multiple disease phenotypes.

O-04-SC10

Demography drives differences in the distribution of gene expression

Arun Durvasula[1], Kirk E Lohmueller[1, 2]

[1]University of California, Los Angeles (United States), [2]University of California, Los Angeles (United States)

Understanding how non-equilibrium population dynamics influence the efficacy of selection remains challenging. While recent empirical work has focused on contrasting patterns of deleterious variation in populations with distinct demographics, theoretical work (Lande 1976) also points to population size being an important factor in shaping the phenotypic distribution of complex traits under stabilizing selection. Here we use gene expression as a phenotype to examine how demography affects stabilizing selection. We develop a model of gene expression and an inference framework to test for differences in the distribution of expression between populations. Applying our approach to RNA sequencing data from Yoruba and European individuals from the GEUVADIS dataset shows that Yoruba individuals have larger expression variance across individuals, suggesting that expression in Yoruba individuals is closer to the optimum. Further, we find greater variance in the mean expression across genes in the European population, indicating that selection is less effective in the bottlenecked European population. Additionally, considering the distribution of variance in expression across all genes, Yoruba individuals have a larger median variance in gene expression, reflecting the increased genetic diversity compared to European individuals. We recapitulate the empirical results using population genetic simulations under a realistic human demography and find that demographic effects on the efficacy of stabilizing selection are predicted on timescales and population sizes relevant for human evolution. Our work provides a new perspective on the efficacy of selection in complex traits and suggests that demography is an important determinant in the evolution of complex traits.

SY29: Somatic mutation and the evolution of multicellularity (July 9, 16:00–18:00)

O-01-SM01

Mutation and Selection Within an Individual

Maria E Orive[1]

[1]University of Kansas (United States)

Somatic mutations potentially play greatly different roles in organisms depending on their form of reproduction. In organisms that reproduce gametically but do not have a separate germline, somatic mutations may eventually be represented in gametes and thus be exposed to between-individual selection. In organisms with agametic, clonal reproduction in addition to sexual reproduction, new clonal individuals may also carry somatic mutations that can be passed to subsequent generations. Thus somatic mutation can affect important evolutionary characteristics, such as mutation load, the generation of genetic diversity, and the rate of evolution. Mathematical analysis shows that within-organism selection between cell lineages can substantially reduce the deleterious mutation rate observed among offspring as well as the mutation load within a population by allowing cells, rather than individuals, to provide the "selective deaths" generated by load. This effect is made more pronounced by plastic development and by the production of multicellular, clonal offspring. A model considering the fate of particular mutations allows a comparison of the relative strengths of somatic and meiotic mutations on allele frequencies at gamete formation, and thus on the generation of between-individual genetic diversity. Recent studies utilizing within-individual sequencing demonstrate possible mechanisms for both an increased opportunity for within-organism selection, and a reduction in between-genotype competition for access to reproduction, underscoring the importance of considering life history details when predicting the eventual result of mutation and selection within individuals.

SY29: Somatic mutation and the evolution of multicellularity (July 9, 16:00–18:00)

O-01-SM02

Somatic evolution in cancer and healthy tissues

Inigo Martincorena[1]

[1]Wellcome Trust Sanger Institute (United Kingdom)

Cancer develops as a result of somatic mutation and clonal selection, but quantitative measures of selection in cancer evolution are lacking. In the first part of my talk, I will introduce our recent work adapting dN/dS models to cancer evolution and I will describe our work to describe the patterns of selection across ~8,000 tumours from 29 different cancer types. Cancers display a distinct and universal pattern of selection rarely seen in species evolution, with positive selection outweighing negative selection during cancer development. On average, less than 1 coding base substitution/tumor is lost through negative selection, with purifying selection almost absent outside homozygous loss of essential genes. In the second part of my talk, I will present our work describing a hidden world of cellular competition and somatic evolution in human healthy tissues, with implications for our understanding of cancer and ageing.

SY29: Somatic mutation and the evolution of multicellularity (July 9, 16:00−18:00)

O-01-SM03

Somatic mutation in Sitka spruce

Vincent Hanlon[1], Sarah P Otto[2], Sally N Aitken[1]

[1]University Of British Columbia (Canada), [2]University Of British Columbia (Canada)

Early studies predicted that the accumulation of heritable somatic mutations with cell divisions or time-dependent DNA damage should result in a high mutation rate in trees. We estimate the somatic base substitution rate in 20 Sitka spruce (*Picea sitchensis*) trees averaging 76 m tall and 300-500 years old. We argue that the result is high for per-generation evolutionary processes, such as the accumulation of deleterious mutations that contribute to genetic load, but low for per-year processes, such as adaptation to fast changes in climate or coevolution with short-lived insect herbivores. The genetic variation that somatic mutation produces may facilitate local adaptation and perhaps enable selection within individual trees.

SY29: Somatic mutation and the evolution of multicellularity (July 9, 16:00–18:00)

O-01-SM04

Somatic Mutations, Development, and Phylogenomics

Reed Cartwright[1]

[1]Arizona State University (United States)

Somatic mutations are mutations which affect somatic cells in multicellular organisms. Understanding somatic mutation is important to understanding the evolution and maintenance of multicellularity, including the pathology of cancer. Most of the research in somatic mutations has been focused on animal systems, while plants, algae, fungi, and protists have received less attention. However, with modern advances in whole-genome sequencing and bioinformatics, biologists are beginning to understand somatic mutation across the tree of life. This in turn will allow molecular evolutionists to answer questions about selection and diversity in soma, including how this impacts on long-term patterns of molecular evolution in the germline.

Properly studying somatic mutation processes using whole-genome sequencing is a difficult challenge. Somatic mutations are rare and often technical errors which occur during sequencing can mask the biological signal. Furthermore, because the developmental relationship between tissues is often not fully understood, the estimation of somatic mutations is actually a phylogenomics problem, where the developmental tree connecting tissue samples must be inferred along with somatic mutations. In this presentation, I will describe (1) new phylogenomic methods in DeNovoGear (https://github.com/denovogear/denovogear/) to study somatic mutations and (2) the application of these methods to somatic mutations in 3 species: (1) Whole genome sequencing of 24 samples taken from a single *Eucalyptus melliodora* tree; (2) Exome sequencing of 7868 human samples taken from 550 donors; (3) Whole genome sequencing of 25 tissue cultures created from single cells taken from two mice.

SY29: Somatic mutation and the evolution of multicellularity (July 9, 16:00–18:00)

O-01-SM05

Accumulation of mutations and natural selection in experimental evolution of basidiomycete fungus Schizophyllum commune

Aleksandra Bezmenova[1, 3], Alexey Penin[3, 2], Elena Zvyagina[5], Artem Kasianov[3, 2], Tatiana Neretina[3], Georgii Bazykin[1, 2], Alexey Kondrashov[4]

[1]Skolkovo Institute of Science and Technology (Russian Federation), [2]Institute for Information Transmission Problems of RAS (Russian Federation), [3]Lomonosov Moscow State University (Russian Federation), [4]University of Michigan (United States), [5]State Nature Reserve Yuganskiy (Russian Federation)

Basidiomycete fungus *Schizophyllum commune* is a unique organism with the highest genetic diversity among studied species, and a high mutation rate of $2.0*10^{-8}$substitutions per nucleotide per generation. We have designed an experimental system which allows us to study the accumulation of somatic mutations in the process of vegetative growth of *S. commune* mycelium, and, in future, to study natural selection acting in the mycelium. We have been cultivating 13 derived lines of 4 original haploid *S. commune* cultures in very thin tubes in order to strictly minimize the effective population size of growing hyphae. Such system is aimed to imitate mutation accumulation line with minimum effect of natural selection, where the mutation rate can be measured. These derived lines have grown over ~45 cm, which corresponds to ~6000 mitoses. We identified 105 *de novo* single mutations accumulated among all lines, which corresponds to the mutation rate of $3*10^{-11}$ substitutions per nucleotide per cell division. We saw no evidence of selection acting in studied cultures; in order to investigate natural selection in *S. commune* mycelium, we plan to study *S. commune*cultures cultivated in tubes 4 mm in diameter, where the size of population of growing hyphae is large.

SY29: Somatic mutation and the evolution of multicellularity (July 9, 16:00–18:00)

O-01-SM06

Multicellular organisms face an inescapable double bind between cell senescence and cancer

Paul G Nelson[1], Joanna Masel[1]

[1]University of Arizona (United States)

Aging is a failure to maintain soma. It is not obvious, however, that somatic mutation and other heritable forms of degradation should necessarily result in systemic accumulation of damaged cells in organisms that are capable of growth and regeneration. Just as competition between organisms can purge deleterious alleles, competition between cells within an organism can purge senescent cells, raising the possibility of an immortal soma. However, not all somatic mutations can be purged through intercellular competition. Some mutations allow a cell to proliferate while disrupting its function within the organism, bestowing higher cell fitness at a cost to the organism. We present a very general mathematical model that incorporates both somatic changes that increase cell fitness, and somatic changes that decrease cell fitness. Using our model of somatic evolution, we find that a mix of senescent and cancer somatic mutations produces a double bind that results in inevitable organismal degeneration over time. The degree of intercellular competition mediates the balance; if intercellular competition is limited, senescent cells accumulate, while if intercellular competition is prevalent, cancerous cells proliferate. The double bind makes aging inevitable for all multicellular organisms with a sequestered germline.

O-02-TM01

Six impossible things before breakfast: assumptions, models and belief in molecular dating

Lindell Bromham[1]

[1]Australian National University (Australia)

Confidence in molecular dating analyses have grown with the increasing sophistication of the methods, which now allow rates of molecular evolution to vary across the phylogeny. Previously problematic case studies where the molecular dates were in disagreement with palaeontological estimates, such as the radiation of placental mammals and the Cambrian explosion of animal phyla, appear to have been resolved with a growing agreement between molecular dates and palaeontological estimates. But we cannot relax just yet. The growing analytical sophistication of molecular dating methods relies on an increasingly large number of assumptions about evolutionary history and processes. Many of these assumptions are based on statistical tractability rather than being informed by increasing knowledge of molecular evolution, yet changing the prior assumptions can dramatically change the date estimates. How can we tell if the answers we get are driven more by the assumptions we make than by the molecular data being analysed?

O-02-TM02

Undersampling genomes has biased time and rate estimates

Blair Hedges[1], Julie Marin[1, 2]

[1]Temple University (United States), [2]National Museum of Natural History (France)

Genomic data drive evolutionary research on the relationships and timescale of life but the genomes of most species remain poorly sampled. Phylogenetic trees can be reconstructed reliably using small data sets and the same has been assumed for the estimation of divergence time with molecular clocks. However, we show here that undersampling of molecular data results in a bias expressed as disproportionately shorter branch lengths and younger divergence times in the youngest nodes and branches leading to increasing speciation and diversification rates towards the present. Any evolutionary analyses derived from these biased branch lengths and speciation rates will be similarly biased. The widely-used timetrees of the major, species-rich studies of tetrapod vertebrates are all data-poor and show upswings in diversification rate suggesting that their results were biased by undersampling. Our results show that greater sampling of genomes is needed for accurate time and rate estimation, which are basic data used in ecological and evolutionary research.

O-02-TM03

Phylogenetic incongruences - opportunities to improve the reconstruction of a dated tree of life

Dominik Schrempf[1], Gergely Szollosi[1, 2]

[1]Eotvos Lorand University, Budapest (Hungary), [2]Hungarian Academy of Sciences (Hungary)

The increasing amount of sequence data that has recently become available has not brought relief to many questions in evolutionary biology but instead has highlighted the complex and ambivalent histories of genes. It is long known that gene trees are not species trees, but the effect of processes such as Incomplete Lineage Sorting (ILS), gene Duplication and Loss (DL) and Horizontal Gene Transfer (HGT) on phylogenetic inference have been underestimated.

We focus our attention on these processes causing incongruence. After all, the strength of ILS is proportional to the population size; the timing of DL can provide information about the evolution of gene functions or pathways; and HGT only occurs between co-existing species and, hence, gives indication about the order of speciation events and the root position.

We plan to perform joint inference in the presence of ILS, DL and HGT using a hierarchical model. On the top level, species form and decay according to a linear birth and death process. The resulting species tree which includes extinct species acts as a constraint for genes that develop in populations, are duplicated, lost or transferred. For this combined model, we aim to derive the likelihood of specific parameters given data and, consequently, develop an MCMC sampler to perform Bayesian parameter inference.

We hope that modeling incongruences helps resolve, e.g., the timing of microbial evolution and its relationship to Earth history, where dating methods are limited by the paucity of fossils; or the position of eukaryotes among archaea.

O-02-TM04

Pervasive correlation of molecular evolutionary rates in the tree of life

Qiqing Tao[1, 2], Koichiro Tamura[3, 4], Fabia Battistuzzi[5], Sudhir Kumar[1, 2]

[1]Temple University (United States), [2]Temple University (United States), [3]Tokyo Metropolitan University (Japan), [4]Tokyo Metropolitan University (Japan), [5]Oakland University (United States)

New species arise from pre-existing species and inherit similar genome biology and environment. This predicts greater similarity of mutation rates and tempo of molecular evolution between direct ancestors and descendants, which will cause autocorrelation of evolutionary rates in the tree of life. Surprisingly, molecular sequence data have not confirmed this expectation, possibly because available methods lack power to detect autocorrelated rates. Here we present an accurate machine learning method to detect autocorrelation of rates in large phylogenies. By applying this method to multigene and genome-scale sequence alignments from mammals, birds, reptiles, insects, metazoans, plants, fungi, and prokaryotes, we discover pervasive and strong autocorrelation in molecular evolutionary rates throughout the tree of life in both DNA and protein sequences. These findings show concordance between molecular and non-molecular evolutionary patterns and will foster unbiased and precise dating of the tree of life.

O-02-TM05

Global rate variation in bony vertebrates

Naoko Takezaki[1], Hidenori Nishihara[2]

[1]Kagawa University (Japan), [2]Tokyo Institute of Technology (Japan)

In this study with amino acid sequence data of 625 genes for 25 bony vertebrates we investigated the substitution rate differences using three calibration points, divergences between lobe-finned vertebrates and ray-finned fish, between mammals and sauropsids, and between holosteans (gar and bowfin) and teleost fish. In our result the substitution rate was two to three times higher in the stem branches of lobe-vertebrates before the mammal-sauropsid divergence than the rate in amniotes. The rate in the stem branch of ray-finned fish before the holostean-teleost fish divergence was also a few times higher than the holostean rate, while it was similar to or somewhat slower than the teleost fish rate. Taking into accounts the high rate in the stem branches, the interval of divergences of coelacanth and lungfish was estimated as 9-13 million years. The times for ordinal divergences within eutherian mammals tend to be smaller than those in previous studies that used the calibration points within the lineage, with deeper divergences before the KPg boundary and shallower ones after the boundary. In contrast those within birds were estimated to be larger, with the divergence between Galliformes and Anseriformes about 80 million years ago (Ma) and that between Galloanseriae and Passeriformes 110 Ma.

O-02-TM06

Recalcitrance of avian divergence times and phylogenetic topology may be related to selection for reduced body size across the K-Pg boundary

Jacob Berv[1], Daniel Field[2]

[1]Cornell University (United States), [2]University of Bath (United Kingdom)

Survivorship following major mass extinctions may be associated with a decrease in body size-a phenomenon called the Lilliput Effect. Body size is a strong predictor of many life history traits (LHTs), and is known to influence demography and intrinsic biological processes. Pronounced changes in organismal size throughout Earth history are therefore likely to be associated with deterministic changes in evolutionary rates. Here, we report on pronounced heterogeneity in rates of molecular evolution (varying up to ~20-fold) across an exon-rich avian data set, and show that nucleotide substitution rates have potentially independent associations with body size and metabolic rate. We also identify potential body size reductions associated with the Cretaceous-Paleogene (K-Pg) transition, consistent with a Lilliput Effect in the wake of that mass extinction event. We posit that selection for reduced body size across the K-Pg extinction horizon may have resulted in transient increases in substitution rate along the deepest branches of the extant avian tree of life. This "hidden" rate acceleration, invisible to both strict and relaxed molecular clocks, may bias age estimates of the avian crown group through the relationship between life history and demographic parameters that scale with molecular substitution rate. Such a scenario may also imply a unifying hypothesis to explain topological conflict across avian phylogenomic datasets related to GC-biased gene conversion. If reductions in body size are a common property of lineages surviving mass extinctions, this phenomenon may help us better understand persistent phylogenetic debates across the tree of life.

SY31: Tracing back bacterial pathogen evolution from ancient and modern genomics (July 9, 10:30–12:30)

O-01-TB01

The genetic history of plague: From the Stone Age to the 18th century.

Johannes Krause[1]

[1]Max Planck Institute for the Science of Human History (Germany)

High throughput DNA sequencing has revolutionized the field of archaeogenetics in the past decade, providing a better understanding of human genetic history, past population dynamics and host pathogen interactions through time. Targeted DNA capture approaches have allowed reconstructing complete ancient bacterial genomes providing direct insights into the evolution and origin of some of the most infamous pathogens known to humans such as *Yersinia pestis, Mycobacterium tuberculosis* or *Mycobacterium leprae*. Here we discuss the potential of ancient pathogen genomics using *Yersinia pestis* as a model organism. Phylogenetic comparisons of modern and ancient *Y.pestis* strains spanning over 5000 years of human history from the Stone Age to modern times are discussed. They provide direct evidence for the timing and emergence of major virulence factors essential for the transmission of bacteria by fleas. We furthermore present the oldest reconstructed genomes of *Y.pestis* that are fully capable of causing the bubonic form of plague from the Eastern European Bronze Age. We suggest that the emergence of this form of the disease happened more than 1000 years earlier than previously thought. Temporal studies of pathogens might thus throw new light on the origin of human diseases and potentially allow predicting and preventing further transmissions and disseminations in the future

O-01-TB02

Adoption of the pathogenic niche

Caitlin Pepperell[1]

[1]University of Wisconsin-Madison (United States)

The vast majority of bacteria exhibit limited pathogenic potential, with most disease attributable to a tiny minority of species. A central question in infectious diseases is how virulence evolves. Concepts of virulence and pathogenicity have themselves evolved, and recent work in phylogenomics has revealed a striking amount of diversity in the adaptive paths to virulence among microbes. Pathogenicity is also increasingly appreciated as an emergent property arising from interactions between microbes and hosts, which complicates and enriches our understanding of infectious disease evolution. In this talk I will discuss recent research in the population genomics of bacterial pathogens, and its applicability to these developing models of disease emergence.

SY31: Tracing back bacterial pathogen evolution from ancient and modern genomics (July 9, 10:30–12:30)

O-01-TB03

A 2,200 year old Mycobacterium leprae genome from an Egyptian mummy

Judith Neukamm[1, 2, 3], Saskia Pfrengle[1], Martyna Molak[4], Michael Francken[1, 5], Beatrix Welte[1], Katerina Harvati[1, 5, 6], Kay Nieselt[2], Johannes Krause[7, 1, 5], Verena J. Schuenemann[1, 3, 5]

[1]University of Tuebingen (Germany), [2]University of Tuebingen (Germany), [3]University of Zurich (Switzerland), [4]Polish Academy of Sciences (Poland), [5]University of Tuebingen (Germany), [6]University of Tuebingen (Germany), [7]Max Planck Institute for the Science of Human History (Germany)

Analyzing ancient DNA opens the possibility to study the evolutionary history of pathogens, such as *Mycobacterium leprae*, one of the oldest recorded diseases in human history. *M. leprae* is the main causative agent of leprosy, which was a common disease in Europe until the 16th century, but is still endemic in many parts of the world. Recent studies on ancient and modern *M .leprae* genomes resulted in a phylogeny that separates five main lineages. Until now, ancient *M. leprae* genomes were recovered from skeletal remains, mostly teeth and bones, of infected individuals with characteristic lesions, while other sources have not been used.

Advanced sequencing technologies and improved analysis methods allowed us recently to establish Egyptian mummies as a genetic source to study ancient human history. Here we present a metagenomic study of various soft tissue and skeletal samples from 93 mummified individuals from Egypt with a focus on the general bacterial composition. For one individual we were able to reconstruct a complete *M. leprae*genome from the Ptolemaic period dated to cal BC 342-117. The so far oldest *M. leprae* genome clusters with modern strains from West Africa and Brazil and is located close to the node of lineage 4.

Our results demonstrate the possibility to use Egyptian mummies as a source for metagenomic analysis. Furthermore, for one particular pathogen, *M. leprae*, our data allow us to refine the current phylogeographic models of leprosy's past distribution and spread.

O-01-TB04

Salmonella enterica genomes from victims of a major sixteenth-century epidemic in Mexico

Alexander Herbig[1, 2], Åshild J Vågene[1, 2], Michael G Campana[6], Nelly M Robles García[4], Christina Warinner[1], Susanna Sabin[1], Maria A Spyrou[1, 2], Aida Andrades Valtueña[1], Daniel H Huson[5], Noreen Tuross[3], Kirsten I Bos[1, 2], Johannes Krause[1, 2]

[1]Max Planck Institute for the Science of Human History, Jena (Germany), [2]University of Tuebingen (Germany), [3]Harvard University (United States), [4]National Institute of Anthropology and History (INAH) (Mexico), [5]University of Tuebingen (Germany), [6]Center for Conservation Genomics, Washington DC (United States)

Indigenous peoples of the New World experienced high mortality rates during the early contact period as a result of infectious diseases, many of which were introduced by Europeans. However, the identification of the pathogenic agents that caused these outbreaks based solely on archaeological and historical evidence is challenging.

Using the novel high-throughput DNA sequence alignment and taxonomic assignment tool MALT (MEGAN ALignment Tool) we were able to identify traces of Salmonella enterica DNA in individuals buried in an early contact era epidemic cemetery at Teposcolula-Yucundaa, Oaxaca in Southern Mexico. This cemetery is linked to the 1545-1550 CE epidemic that affected large parts of Mexico, the pathogenic cause of which has been debated for more than a century. After application of a specifically designed targeted DNA enrichment procedure we generated genome-wide data from ten individuals for *Salmonella enterica* subsp. *enterica* serovar Paratyphi C, a bacterial cause of enteric fever.

In order to provide insights into the evolutionary history of this pathogen we comparatively analyzed our ancient genomes in the context of modern *S. enterica* strain variation for the assessment of phylogenetic relationships as well as presence and absence of virulence factors.

We propose *S.* Paratyphi C as a strong candidate for the epidemic population decline during the 1545 outbreak at Teposcolula-Yucundaa and provide evidence that it was likely introduced by Europeans.

O-01-TB05

A Single, >2.2 Ga Old Event of Host Adaptation in the Order Legionellales

Eric Hugoson[1], Daniel Leenheer[1, 2], Tea Ammunet[1], Helena Ishak[1], Tiscar Graells[1, 3, 4], Madeleine Larsson[1], Anaisa Moreno[1], Lionel Guy[1]

[1]Uppsala University (Sweden), [2]University of Tsukuba (Japan), [3]Uppsala University (Sweden), [4]Universitat Autonoma de Barcelona (Spain)

Host-adapted intracellular bacteria are of critical importance to human health, ecology and economy, but obtaining whole genomes for these fastidious organisms is challenging. The gammaproteobacterial order *Legionellales* has representatives at various stages of host-adaptation, from facultative intracellular like Legionella to obligate mutual symbionts, and is thus an excellent model to understand the evolution of host-adaptation.

Twenty-seven novel metagenome-assembled genomes (MAGs) belonging to *Legionellales* were assembled or retrieved from databases; two whole genomes were sequenced. A phylogenomic tree inferred from 109 conserved marker genes was used to reconstruct the gene content of key ancestors of the order. A divergence time estimation was calibrated with the last common ancestor of purple sulfur bacteria, which are phylogenetically close to *Legionellales*, and which geological records can confidently date to no later than 1.64 Ga. It places the *Legionellales* last common ancestor (LLCA) at 2.4 Ga (95% credible interval: 2.2-2.7 Ga), well before the last eukaryotic common ancestor (LECA) (1.0-1.9 Ga).

These analyses confirm that the order is more diverse than previously thought, with most of the new MAGs representing novel genera and species, and one group possibly representing a novel family. We infer that the Type IV Secretion System and many key eukaryotic-like effector proteins were already present in LLCA, confirming that one single event of host-adaptation occurred in the order. The time estimate for LLCA is compatible with *Legionellales* colonizing the new eukaryotic niche during early eukaryogenesis.

O-01-TB06

A high-quality 17th century Mycobacterium tuberculosis genome adds resolution to mycobacterial dating and phylogeography

Susanna Sabin[1], Alexander Herbig[1], Ashild J. Vagene[1], Torbjorn Ahlstrom[2], Gracijela Bozovic[3], Elizabeth A. Nelson[1], Caroline Arcini[4], Kirsten I. Bos[1]

[1]Max Planck Institute for the Science of Human History (Germany), [2]Lund University (Sweden), [3]Skane University Hospital and Lund University (Sweden), [4]National Historical Museum (Sweden)

Tuberculosis occupies an intense focus of medical and microbiological research, and questions persist regarding its origin. Based on the phylogeography of modern *Mycobacterium tuberculosis* complex (MTBC) genomes, one hypothesis suggests their most recent common ancestor (MRCA) followed human Pleistocene migrations out of Africa ~70,000 years before present. However, studies using ancient genomes to calibrate the molecular clock have indicated much younger MRCA dates of less than 6,000 years. Here we aim to resolve this discrepancy by analyzing a calcified lung nodule from the mummified remains of Bishop Peder Winstrup of Lund (born 1605 - died 1679), which offers a unique opportunity to generate a calibration point for the historical phylogenetic reconstruction of the MTBC and account for potential substitution rate heterogeneity between different lineages over time. We were able to identify DNA belonging to the MTBC with a metagenomic approach. Enrichment of the genetic material was accomplished with a custom-designed in-solution capture method, which allowed us to reconstruct a 141-fold coverage genome of *Mycobacterium tuberculosis* from a library for which only 0.045 percent of the metagenomic DNA had been identified as belonging to members of the MTBC. Phylogenetic analysis revealed this sample to be part of the recently defined L4.10/PGG3 sublineage of the MTBC. This high-quality, high-coverage 17th century *Mycobacterium tuberculosis* genome provides an opportunity to apply temporally diverse data to a phylogeographic analysis of the L4.10/PGG3 sublineage, and adds a reliable calibration point for dating the MTBC via Bayesian methods.

SY32: Transposable elements in gene regulation and genome evolution

(July 10, 16:00–18:30)

O-02-TE01

Transposable elements as catalysts of convergent evolution

Cedric Feschotte[1]

[1]Cornell University (United States)

n this talk I will argue that the cooption of transposable element sequences is a common path for the emergence of similar molecular innovations in different organismal lineages. This mode of convergent evolution has been fueled by both coding and noncoding components of transposable elements and has led to the remodeling of a wide range of biological processes. I will highlight specific cases of convergent cooption that we have recently contributed to uncover in several organisms. One involves the regulation of the innate immune response of mammals; and another concerns a trans-cellular RNA signaling system that modulates synaptic plasticity in flies and vertebrates.

O-02-TE02

Chromatin sinks and mutational burdens in males

Lauren Gibilisco[1], Doris Bachtrog[1]

[1]University of California, Berkeley (United States)

Large portions of eukaryotic genomes consist of repetitive elements, and the establishment of transcription-repressing heterochromatin during early development safeguards genome integrity. Heterochromatin formation coincides with genome-wide activation of zygotic expression. Males often contain substantially more heterochromatic DNA than females, due to the presence of a large, repeat-rich Y chromosome. This may cause a delay in heterochromatin establishment at repetitive sequences in males compared to females, since more repressive chromatin marks need to be generated in the repeat-rich male genome. We performed chromatin immunoprecipitation experiments in single embryos to study the dynamics of heterochromatin formation during early development. We find that maternally deposited piRNAs target silencing of repeats, and we show that heterochromatin formation is indeed delayed in early male Drosophila embryos. Most interestingly, we find that this delay in establishing transcription-repressing heterochromatin coincides with an increase in repeat expression in male versus female embryos. Moreover, we also find more de novo insertions of repeats in males. Thus, the Y chromosome may indirectly create a mutational burden in males by postponing the establishment of silencing chromatin marks. These results suggest that the adverse effect of the Y chromosome as it relates to transposable element mobilization is not just limited to aging flies, but in fact may begin as soon as zygotic transcription starts.

O-02-TE03

The contribution of transposable elements to regulation of genes underlying symbiosis in *Epichloe*

David Winter[1]

[1]Massey University (New Zealand)

Epichloe species comprise a genus of ascomycete fungi that form symbiotic relationships with particular grass species. This symbiosis can have profound effects on the grass, providing resistance to pathogens and herbivores, and tolerance of drought conditions. We recently generated a complete and gapless assembly for *E. festucae*, a model *Epichloe* species. Our results revealed a 'patchwork' genome, in which long gene-rich segments are interspersed by AT-rich regions made almost entirely of inactivated transposable elements (TEs). Using chromatin conformation capture data, we show that these TE-rich regions have a considerable impact on the three-dimensional structure of the genome and frequently act to isolate neighbouring gene-rich regions from each other within the nucleus. Genes that are differentially expressed *in planta* or known to be important for maintaining the plant-fungus interaction are associated with the boundaries of these TE-rich tracts, suggesting these regions may play a key role in the regulation of symbiosis-associated genes.

In this talk, I will discuss our ongoing work to test the hypothesis that TEs contribute to the regulatory evolution in *Epichloe*. We have generated high-quality assemblies and RNA-seq data from two additional species, each of which are adapted to a different host grass species. We are using this data to test whether novel TE-insertions of genomic rearrangements associated with TEs can explain species-specific gene expression patterns that may underpin adaptation between specific *Epichloe* and grass species.

O-02-TE04

Pervasive epigenetic effects of euchromatic transposable elements impact genome evolution

Grace Yuh Chwen Lee[1, 2], Gary H Karpen[1, 2]

[1]Lawrence Berkeley National Laboratory (United States), [2]University of California, Berkeley (United States)

Transposable elements (TEs) are ubiquitous genomic parasites, and their evolution has remained a critical question in evolutionary genomics. Here, we study the relatively unexplored functional and evolutionary consequences of the *epigenetic* effects of TEs. Euchromatic TEs can be epigenetically silenced via host-directed enrichment of repressive heterochromatic marks, which can spread *in cis* to and influence the function of adjacent sequences. We quantified the epigenetic effects of TEs in multiple wild-derived *D. melanogaster* and *D. simulans* genomes. Over half of the euchromatic TEs show spread of repressive epigenetic marks to nearby DNA, which can extend up to 20kb. Such effects result in differential epigenetic states and transcript levels of genic alleles with and without neighboring TEs, which in return leading to selection against TEs. Interestingly, compared to *D. melanogaster*, TEs in *D. simulans* show stronger epigenetic effects and should be more strongly selected against, which could explain the lower TE content in the species. We provide evidence that this between-species difference in TE's epigenetic effects can be contributed by variation in host genetic factors known to modulate epigenetic silencing. In addition to *cis*-spreading of heterochromatic marks, we found that TE-induced enrichment of heterochromatic marks could also lead to spatial interactions between TE-flanking euchromatic regions and heterochromatin domain, which results in even more extensive *cis*-spreading as well as epigenetic silencing on the homologous chromosome (*trans* silencing). Our study demonstrates that the epigenetic effects of TEs play a critical role in the evolution of TEs and, more broadly, host genomes.

O-02-TE05

Poxvirus capture of host genes by retrotransposition in infected cells

Della Fixsen[1]

[1]University Of Utah (United States)

Horizontal gene transfer (HGT) provides a major source of genetic variation in virus evolution. Many viruses encode host proteins that have diverged and gained new functions altering cellular processes or subverting host defenses. While host gene acquisition is a common means of viral adaptation, mechanisms and dynamics of the process are not well understood. We devised a system to induce and select for HGT events in poxvirus populations, using vaccinia as a model. Mutant viruses lacking the essential K3L gene were propagated in complementing cells with K3L chromosomally integrated. We then screened for strains that overcame the antiviral PKR pathway to replicate in non-complementing cells. We recovered six virus clones from these experiments. Each strain independently acquired K3L via HGT from the complementing cells. Genomic analysis revealed target site duplications, poly(A) tails, and a lack of introns from the integrated construct, which all support a mechanism where K3L was retrotransposed into the viral gene through LINE-1 activity in infected cells. We also found corroborating evidence of LINE-1 mediated HGT in published genomes of diverse viruses. To test whether viruses influence LINE-1 activity to favor HGT, we used a GFP reporter of LINE-1 activity in infected versus uninfected cells. Surprisingly, we find that human cells show a reduction in LINE-1 retrotransposition upon infection with vaccinia and we are testing whether poxviruses redirect LINE-1 proteins to virus factories in infected cells. These results reveal that LINE-1 mobilization of host genes to virus genomes might represent a primary mechanism in virus evolution.

O-02-TE06

The birth of new genome defense genes by transposable elements

Richard McLaughlin[1], Harmit Malik[2], Michael Emerman[3]

[1]Pacific Northwest Research Institute (United States), [2]Fred Hutchinson Cancer Research Center (United States), [3]Fred Hutchinson Cancer Research Center (United States)

The replication of mobile genetic elements imposes fitness costs on hosts. Consequently, host genomes evolved an array of restriction factors to block the replication and deleterious effects of these elements. However, the composition of a host's pathogen repertoire changes over evolutionary time, as there is always an advantage to be gained by pathogens in evading host restriction. The APOBEC3 (A3) family of cytidine deaminase genes provides restriction of infectious and endogenous pathogens. The dynamic expansion and contraction of this gene family likely represents host adaptation to keep apace of the dynamic evolution of pathogens. The birth of new A3 genes within the A3 locus comes from segmental duplications. Here, we describe a dramatic expansion of the A3 gene family in New World monkeys, with up to nine additional copies of A3G in some species; however, this expansion occurred via retroposition, in which an endogenous retroelement duplicates a host gene via its transcribed RNA. We find these duplications to be pervasive in New World monkeys and other primates. Further, we see that some A3 retrogenes are ancient and some have been born very recently. Finally, we show that some of these gene copies encode important genome defense functions including retrovirus restriction. We suggest that retroposition of host defense factors represents an unappreciated mechanism of rapid adaptation to block pathogens. In this way, the host genome co-opts retroelement activity to block the deleterious consequences of pathogens. In this way, retroelements create new host genes which block the deleterious consequences of pathogens.

O-02-TE07

The tuatara genome - a detailed look into early amniote genome evolution and a smorgasbord of recently active transposons

Alexander Suh[1], Valentina Peona[1], Claire Peart[1, 2], Vera Warmuth[1, 2], Neil Gemmell[3]

[1]Uppsala University (Sweden), [2]Ludwig-Maximilian University of Munich (Germany), [3]University of Otago (New Zealand)

The tuatara (*Sphenodon punctatus*) is the only living member of the archaic taxon Rynchocephalia and shared a common ancestor with squamates (snakes and lizards) some 220-250 Mya. Due to its phylogenetic position and large genome size, the latter suggestive of a slow rate of transposable element (TE) turnover, the tuatara provides a unique window into early amniote genome evolution. The large collaborative efforts of the Tuatara Genome Consortium generated a high-quality tuatara genome assembly (~4.6 Gb), one of the largest and most repetitive vertebrate genomes thus far assembled. Here, we undertook in-depth annotation of the repetitive fraction of the tuatara genome (~50% of the genome) with a special focus on DNA transposons and short interspersed elements (SINEs). As expected, the tuatara genome contains a substantial diversity of ancient TEs, which, by comparison with the similarly slowly evolving genomes of crocodilians and turtles, permits a detailed reconstruction of TE accumulation during early amniote genome evolution. Surprisingly, we also identified a high diversity of very recently active TEs. At least 16 SINE subfamilies and 30 DNA transposon subfamilies have hundreds or thousands of identical copies, respectively, including many MIR elements with insulator motifs. Altogether, this suggests that the tuatara genome is likely the most dynamic known amniote genome in terms of putatively ongoing (retro)transposition. The presence of intact copies from TE families which were previously assumed to be long extinct is promising for future mechanistic studies.

O-02-TE08

The role of transposable-element mediated rewiring of regulatory networks: dosage compensation in Drosophila

Doris Bachtrog[1]

[1]University of California, Berkeley (United States)

Evolutionary innovations and adaptations often require rapid and concerted changes in regulation of gene expression at many loci. Transposable-elements (TEs) constitute the most dynamic part of eukaryotic genomes, and insertions of TEs can influence the expression of surrounding genes by donating new regulatory elements. A longstanding hypothesis, first proposed by Barbara McClintock, postulates that the dispersal of TEs may allow for the same regulatory motif to be recruited at many genomic locations, thereby drawing multiple genes into the same regulatory network. Empirical evidence for this model is however scarce, and most putative examples of TE-mediated rewiring of regulatory networks rely on a statistical association between remnants of a TE at a subset of genes or genomic regions. We recently provided direct functional evidence of an active TE rewiring a regulatory network by showing that the acquisition of novel binding sites for the dosage compensation complex at young neo-sex chromosomes in Drosophila was driven by dispersal of a domesticated TE. Our detailed understanding of how dosage compensation in Drosophila works at the molecular level makes it an ideal model system to study the rewiring of regulatory networks. Here we quantify the involvement of TEs vs. acquisition of regulatory sites by other mutations in rewiring regulatory networks, by systematically studying the evolution of dosage compensation binding sites at multiple independently formed young neo-sex chromosomes in Drosophila. We show that the involvement of TEs in rewiring regulatory networks is contingent on the repetitive genome landscape, and is thus expected to differ among taxa.

O-04-TT01

Experimentally evolving molecular Rube Goldberg machines

Anthony Poole[1]

[1]University of Auckland (New Zealand)

In assessing the function of molecular complexes and processes it is routine to examine the function of each and every contributing component. However, it is plausible that some components are 'essential' by historical accident, or only if organized in a particular manner. There may therefore be a difference between the 'function' of a particular component that performs a specific step in a process, and whether the step itself is necessary for the process. I argue that some processes or components (which when knocked-out exhibit an 'essential' phenotype) are in fact not necessary at all. To illustrate this, I will present evolution experiments from our lab that show how seemingly essential components of bacterial translation initiation can be dispensed with under certain conditions, and that RNA editing-type processes may emerge and become necessary without contributing to function. Our results suggest that some 'function' may be better explained through the addition of needless steps or processes (sensu Rube Goldberg machines) than by functional improvement or enhancement.

O-04-TT02

Inteins and Homing Endonucleases: long term survival and constructive neutral evolution

Johann Peter Gogarten[1], Shannon M Soucy[1], Joshua Skydel[1], Artemis Dyanna Louyakis [1]

[1]University of Connecticut (United States)

Inteins, aka protein introns, are selfish genetic elements that remove themselves from a host protein through a self-splicing reaction, thereby restoring the host protein to its normal conformation and reducing the fitness cost to the host organism. Intein often harbor a central homing endonuclease domain, which cleaves non-invaded target sites, facilitating the invasion of populations through the intein. The presentation will discuss possible reasons for target site preference; the fitness cost of the intein to the host organism; different models for the long-term survival of inteins with functioning homing endonuclease in populations; and examples for inteins acquiring functions that became necessary for the host (they are under purifying selection), in some instances without initially providing an increase in fitness to the host (constructive neutral evolution).

O-04-TT03

Can the human genome be 100% functional? An argument based on the concept of mutational load

Dan Graur[1]

[1]University of Houston (United States)

For the human population to maintain a constant size from generation to generation, an increase in fecundity must compensate for the reduction in the mean fitness of the population caused by deleterious mutations. The required increase depends on the deleterious mutation rate and the number of sites in the genome that are functional. These dependencies and the fact that fecundity cannot be arbitrarily large, allow us to estimate an upper limit for the fraction of the human genome that can be functional. By estimating the fraction of deleterious mutation out of all mutations in known functional regions, we conclude that the fraction of the human genome that can be functional cannot exceed 25%, and is almost certainly much lower.

O-04-TT04

Selfish modifiers of recombination and mutation, their response to stress, and implications for evolvability

Lilach Hadany[1]

[1]Tel Aviv University (Israel)

I will present models for the evolution of selfish elements that increase host recombination and mutation rates in response to stress (for example through transposition). I will discuss the implications of the evolved patterns of variation for further evolution of the host genome. Specifically, I will show that such stress induced variation, occurring due to its selfish advantage, may facilitate the evolution of the host, especially when the fitness landscape is rugged and complex adaptation is required.

O-04-TT05

An evolutionary lock-in event facilitates the persistence of self-splicing introns in S. cerevisiae

Tobias Warnecke[1, 2]

[1]London Institute of Medical Sciences (United Kingdom), [2]Imperial College London (United Kingdom)

Self-splicing introns are mobile elements that have invaded a number of highly conserved genes in prokaryotic and organellar genomes. They are considered to be classic genetic parasites, not known to benefit host fitness. Naively, then, one would expect that removing these introns from the genome would be beneficial or, at worst, make no difference to the host. In contrast to this expectation, we show that wholesale deletion of all 13 self-splicing introns from the *Saccharomyces cerevisiae*mitochondrial genome is stressful to the host. A strain without mitochondrial introns displays hallmarks of the retrograde response, with altered mitochondrial morphology, gene expression and metabolism impacting growth and lifespan. We show that, without introns, transcript maturation becomes too efficient: fewer transcripts mis-splice and are degraded, leading to deleterious excess levels of mature *cob* and *cox1* host mRNA. Curiously, inefficient splicing has become an integral part of normal mitochondrial gene expression. We propose that the persistence of *S. cerevisiae* self-splicing introns has been facilitated by an evolutionary lock-in event, where the host genome adapted to primordial invasion in a way that incidentally rendered subsequent intron loss deleterious. Following compensatory evolution that enabled the host to cope with intron invasion, removing the invader has suddenly become maladaptive. Our study suggests that, counterintuitively, the long-term persistence of mobile elements can be facilitated by early adaptive changes in the host genome that ameliorate deleterious effects of mobile element activity.

O-04-TT06

Functional Shifts in Duplicated Genes via Specialization of Interacting Partners

Ashley Irene Teufel[1], Claus Wilke[1]

[1]The University of Texas at Austin (United States)

Gene duplication is believed to be a major source of structural and functional divergence in genome evolution. However, the molecular mechanisms that underline how duplicated proteins gain new functions or retain ancestral ones is poorly understood. Here we investigate how duplicated proteins evolve under a number of different selective scenarios with the use of evolutionary simulations. We find that some selective scenarios are more conducive to the preservation of ancestral function than others, even when this function is not being selected for. Further, when examining the case where just one member of a pair of interacting proteins is duplicated we uncover a previously unrecognized process that may influence how protein interfaces evolve and how duplicate genes diverge. Here, we show that a non-duplicated binding partner can be instrumental in shaping the fate of the duplicated gene. For example, if a duplicated protein is under selective pressure not to bind its original binding partner, this can lead to specialization of the binding interface to exclude the duplicate. Interestingly, this finding implies that the presences of a duplicate can impact how the original function operates. Though it is conventionally assumed that how duplicate genes acquire novel functionality depends on changes to the gene itself, here we demonstrate that these ends can be facilitated by modifications to the original functionality. This observation suggests that the fate of duplicate genes may not only be controlled by their own ability to accumulate mutations, but also how non-duplicated interacting partners cope with their presences.

O-02-VG01

Patterns of genomic diversification in natural communities of marine bacteriophages

Marcia F Marston[1]

[1]Roger Williams University (United States)

Viruses are one of the most abundant entities in the ocean. Interactions between viruses and their hosts can influence the composition and evolution of both viral and bacterial populations; yet the long-term genomic consequences of these interactions in natural communities are not well understood. In this study, we examined the genomic structure, diversification, and ecological-cohesion within natural communities of marine bacteriophages. We analyzed the genome sequences of over 250 *Synechococcus*-infecting myovirus and podovirus isolates collected from the coastal waters of Southern New England, USA over 18 years. Combining bacteriophage genomic data, targeted-PCR analysis of an additional 800 isolates, and long-term ecological abundance data, we observe that bacteriophages form distinct genomic clusters and subclusters that remain stable for a decade or longer. Further, these clusters and subclusters exhibit distinct spatial and/or temporal patterns of abundance, suggesting that these units represent distinct viral ecotypes, similar to those observed in bacteria and archaea. We also examined the genomic locations and types of genetic changes that lead to the diversification of bacteriophages into distinct ecotypes. Interestingly, the patterns of genomic diversification differ between myoviruses and podoviruses suggesting that these two families of viruses may rely on different processes to coexist with their hosts and persist in natural environments.

O-02-VG02

The rise and fall of “cheater” viruses in experimentally evolved MS2 phage populations

Adi Stern[1]

[1]Tel Aviv University (Israel)

Moran Meir, Danielle Miller, Noam Harel, Maoz Gelbart, Avigdor Eldar, Uri Gophna and Adi Stern

RNA viruses provide several advantages for studying evolution in action due to their large populations, short generation times, small genomes and high mutation rates. In this study, we evolved the MS2 RNA bacteriophage for twenty serial passages at multiplicity of infection of one. Unexpectedly, following time-series next generation sequencing, we observed rapid increase in population frequency of a deletion mutation ($f¢1764$) that eliminates two of the four reading frames of the phage. In three of four experimental lines $f¢1764$ reached a population frequency of around 50%, which remained stable for several passages and was then followed by a dramatic decline down to a population frequency of almost zero. Surprisingly, a similar but unidentical pattern occurred for several other variant mutations located proximal to $f¢1764$. We hypothesized that $f¢1764$ as well as other variants behave as “cheater” phages, able to replicate only in the presence of wild-type (WT) phage and subverting the infected cell to favor cheater genomes over WT genomes. We substantiated our hypothesis for $f¢1764$ using plaque assays and real-time PCR. We next developed a mathematical model of phage replication that suggests differential fitness of the cheater phage that depends both on the ratio of cheater to WT co-infection, and on the density of the host population. While cheaters are thought to typically replicate their genomes faster than WT genomes, we show this is not the case in our experiment. We further show that cheater phages exhibit delayed host lysis. Finally, the continued evolution of the phage for an additional ten passages led to a plethora of novel and old mutations that rise and fall at different tempo, revealing that the dynamics of the phage are not deterministic and suggesting a strong stochastic component in the evolution of phages during co-infection.

O-02-VG03

Intra-patient evolutionary dynamics of HIV drug resistance evolution

Alison Feder[1], Soo-Yon Rhee[5], Susan Holmes[3], Robert Shafer[5], Zandrea Ambrose[4], Joachim Hermisson[6], Pleuni Pennings[2], Dmitri Petrov[1]

[1]Stanford University (United States), [2]San Francisco State University (United States), [3]Stanford University (United States), [4]University of Pittsburgh (United States), [5]Stanford University (United States), [6]University of Vienna (Austria)

At the beginning of the HIV epidemic, drug resistance to treatment evolved quickly and predictably across all patients. Now we treat HIV with combination therapies of three drugs so any single HIV mutation is insufficient for replication. While drug resistance evolution has plummeted in response, a minority of viral populations become resistant nonetheless. Why and how do certain populations overcome efficacious combination therapy? In answer, we analyze how the mode of drug resistance evolution changed within patients throughout the epidemic using historical HIV sequences. We find evolution has shifted from multiple origins of drug resistance ("soft sweeps") to single origins ("hard sweeps") as treatments improved. This suggests that while drug resistance was once inevitable, now patients that fail combination therapy are merely unlucky, and not predestined to fail due to factors like poor adherence. However, questions remain about how hard sweeps of drug resistance occur at all under combination therapy. Theory suggests that spatial structure of the intra-patient population drives multidrug resistance through creating pockets of spatial monotherapy whereby mutations can be acquired sequentially instead of simultaneously (Moreno-Gamez et al, 2015). However, within-body population structure remains unknown. To bridge this gap, we analyze Simian-HIV-infected macaques sampled spatially (across tissues) and temporally during drug resistance evolution. We find that HIV exists in an understudied population genetic regime where selection, migration and mutation are all strong forces, and build a framework to do parameter estimation under these conditions. These projects demonstrate how evolutionary approaches can advance our understanding of medically-important problems.

O-02-VG04

The origin of Vertebrate RNA Viruses

Mang Shi[1, 2], Edward Holmes[1, 2], Yong-zhen Zhang[2]

[1]The University of Sydney (Australia), [2]China CDC (China)

Our understanding of the diversity and evolution of vertebrate RNA viruses is largely limited to those found in mammalian and avian hosts and associated with overt disease. Here we performed a large scale meta-transcriptomics survey of potential vertebrate-associated RNA viruses in more than 186 host species representing the extensive diversity within the phylum Chordata, which includes reptiles, amphibians, lungfish, ray-finned fish, cartilaginous fish, jawless fish, and lancelet. We discovered 214 vertebrate-associated viruses in these hosts and they appeared in every family/genus of RNA viruses associated with vertebrate infection, including those containing human pathogens such as Influenza viruses, the Hantaviridae, the Arenaviridae, and the Filoviridae, and with branching orders that broadly reflect the phylogenetic history of their hosts. A long evolutionary history was established for most groups of vertebrate RNA virus, and was supported by evaluating evolutionary time-scales using dated orthologous endogenous virus elements. We also identified a diverse range of genome architectures which reveals a more complete evolutionary history of virus genomes than previously depicted. In summary, this study shows that RNA viruses in vertebrates tend to broadly follow the evolutionary history of their hosts that began in the ocean and extends for hundreds of millions of years.

O-02-VG05

Host switches and intrahost speciations play a central role in herpesvirus evolution

Anderson Brito[1], John Pinney[1]

[1]Imperial College London (United Kingdom)

Herpesviruses (HVs) are classified in three subfamilies (Alpha-, Beta-, and Gammaherpesvirinae), which infect mammals, birds and reptiles. Cospeciation has been suggested as the main mechanism underlying herpesvirus-host evolution, with intrahost speciation, extinction and host switch deemed to play minor roles. So far no studies have tried to reconcile time-calibrated phylogenies to assess the prevalence of such mechanisms. In our study we performed such an analysis, producing chronologically consistent reconciliations.

Our results reveal an intricate series of events shaping the evolution of herpesviruses over the past 400 million of years. Two processes played a central role: (i) early intrahost speciations gave rise to ancestors of the three herpesviral subfamilies before the split between proto-mammals (Synapids) and bird/reptile ancestors (Sauropsida) in the Carboniferous; and (ii) mammalian alphaHVs most likely originated from viruses transferred from birds in the early Cretaceous. As host and viral phylogenies mostly disagree both topologically and chronologically, cospeciation is shown to be rare. Losses, on the other hand, were the most frequent events, followed by intrahost speciations, with similar frequencies in all subfamilies, and host switches, which have been shown to be particularly common among alphaHVs, probably as a result of their larger number of sampled taxa.

Our results suggest that as more taxa are included in reconciliation analysis, more intrahost speciations and host switches are evidenced. If such hypotheses are consistent, more attention would be required as to the potential impacts of herpesvirus host switches on human health and on domestic animals of economic importance.

O-02-VG06

Norovirus pandemics emerge from hidden reservoirs and are not driven by the acquisition of viral genetic changes

Christopher Ruis[1], Judith Breuer[1], Richard Allen Goldstein[1]

[1]UCL (United Kingdom)

The norovirus genotype GII.4 is a leading cause of human gastroenteritis and causes frequent pandemics. It has previously been proposed that pandemics are driven by the acquisition of viral genetic substitutions that enable evasion of population immunity. By reconstructing the temporal evolutionary history and phylogeography of GII.4 we demonstrate instead that pandemic strains diverge and circulate cryptically over a wide geographical area for years prior to pandemic emergence. These strains diversify during this period, with multiple lineages emerging nearly simultaneously. This indicates that the antigenic characteristics required for pandemic emergence occurs substantially prior to the start of the pandemic. We conclude that the driver of new norovirus pandemics cannot be the acquisition of new viral genetic changes; rather we hypothesise that changes in host population immunity drive pandemic emergence of a pre-adapted strain that has been circulating undetected.

SY35: Weak forces in genome evolution (July 9, 16:00–18:00)

O-01-WF01

The dark side of recombination: biased gene conversion in the tree of life

Laurent Duret[1]

[1]CNRS, Universite de Lyon (France)

Recombination is typically thought as a symmetrical process resulting in large-scale reciprocal genetic exchanges between homologous chromosomes. Recombination events, however, are also accompanied by short-scale, unidirectional exchanges in the neighborhood of the initiating double-strand break: gene conversion. A large body of evidence suggests that gene conversion is GC-biased in mammals. AT/GC heterozygotes produce a larger amount of GC- than AT-gametes, thus conferring a population advantage to GC-alleles in high-recombining regions. This apparently unimportant feature of our molecular machinery has strong evolutionary consequences. Structurally, GC-biased gene conversion (gBGC) explains the spatial distribution of GC-content in mammalian genomes - the so-called isochore structure. Functionally, gBGC promotes the "undesired" segregation and fixation of deleterious AT GC mutations, thus increasing our genomic mutation load. I will present an analysis of the signatures of gBGC in the tree of life. Our results show that gBGC is probably a universal process, present not only in many eukaryotic taxa but also in bacteria.

O-01-WF02

Genomic disintegration in woolly mammoths on Wrangel island

Rebekah L Rogers[1], Montgomery Slatkin[2]

[1]UNC Charlotte (United States), [2]UC Berkeley (United States)

Woolly mammoths (Mammuthus primigenius) populated Siberia, Beringia, and North America during the Pleistocene and early Holocene. Recent breakthroughs in ancient DNA sequencing have allowed for complete genome sequencing for two specimens of woolly mammoths (Palkopoulou et al. 2015). One mammoth specimen is from a mainland population 45,000 years ago when mammoths were plentiful. The second, a 4300 yr old specimen, is derived from an isolated population on Wrangel island where mammoths subsisted with small effective population size more than 43 fold lower than previous populations. These extreme differences in effective population size offer a rare opportunity to test nearly neutral models of genome architecture evolution within a single species. Using these previously published mammoth sequences, we identify deletions, retrogenes, and nonfunctionalizing point mutations. In the Wrangel island mammoth, we identify a greater number of deletions, a larger proportion of deletions effecting gene sequences, a greater number of candidate retrogenes, and an increased number of premature stop codons. This accumulation of detrimental mutations is consistent with genomic meltdown in response to low effective population sizes in the dwindling mammoth population on Wrangel island. In addition, we observe high rates of loss of olfactory receptors, either because these loci are non essential or because they were favored by divergent selective pressures in island environments. Finally, at the locus of FOXQ1 we observe two independent loss of function mutations, which would confer a satin coat phenotype in this island woolly mammoth.

O-01-WF03

The role of genetic variation in the evolution of enzyme functions

Nobuhiko Tokuriki[1]

[1]University of British Columbia (Canada)

Genetic variation across orthologous proteins is the outcome of genetic drifts in respect to their native, physiological function within each host. How such genetic variation can cause variation in phenotypic levels, variation is adaptive, evolutionary potential, i.e., evolvability, of proteins is an important concept to understand protein evolution. Such phenotypic variation may underlie evolutionary contingency in nature, i.e., certain population may promptly adapt to new environment while others may stagnated. More fundamentally, what is the molecule causes for such variation in phenotype and evolvability?

We address this question using several enzyme model systems to explore variation in phenotypic and evolvability among natural and laboratory occurred homologous enzymes. First, I present how natural orthologs of antibiotic resistance enzymes, metallo-beta-lactamases, showed substantial phenotypic variations in particular host organisms, while their catalytic efficiency (Kcat/KM) for degrading beta-lactam antibiotics. I show that proteostasis, in particular, the translocation of the enzyme to periplasm causes substantial (>100-fold) difference in the levels of antibiotic resistances. Second, I present our several our laboratory evolution experiments to discuss how that seemingly neutral genetic variation between the orthologous enzymes can cause significant difference in evolutionary outcomes. Our experimental evolution revealed that the evolvability of homologous enzymes differ more than 100-fold in parallel experimental evolution. I also discuss molecular basis underlying such variation in evolvability.

Finally, I discuss implications of genetic variation for evolution in nature and protein engineering strategies to generate new functions in the laboratory.

O-01-WF04

Contrasting the influence of gBGC on adaptive statistics in primates and birds

Marjolaine Rousselle[1], Alexandre Laverre[1], Emeric Figuet[1], Benoit Nabholz[1], Nicolas Galtier[1]

[1]University of Montpellier (France)

GC-biased gene conversion (gBGC) is a recombination-linked mechanism that increases the probability of fixation of GC alleles around recombination hotspots, independently of their fitness effects. As such, it is thought to mimic the effects of positive selection and to potentially cause errors in the estimation of quantities like dN/dS and the adaptive substitution rate[1,2]. However, the determinants and patterns of variation over time of the recombination landscape vary across taxonomic groups, making the impact of gBGC on molecular evolution not easy to predict. Mammals and birds are two taxa known to be affected by gBGC[2,3] while showing very different patterns of hotspot location in space and time. Here, we explore how gBGC differently affects adaptive statistics in these two taxa using genome-wide data of six species of fowls and six species of primates. In each species we computed dN/dS, site frequency spectra and estimated the proportion of adaptive substitutions according to GC-content. We did this for AT->GC, GC->AT and G<->C/A<->T mutations to disentangle the effects of gBGC from linkage effects. We show that gBGC strongly decreases dN/dS when the recombination rate increases through an increase in dS in birds, whereas it increases dN/dS in primates. Additionally, we reveal an unexpected effect of GC-content on GC-conservative substitution rates, in agreement with the hypothesis of a mutagenic effect of recombination.

[1]Figuet et al. 2016. Mol.Biol.Evol. 33:1517-1527.

[2]Ratnakumar et al. 2010. Phil.Trans.R.Soc.B. 365:2571-2580.

[3]Bolivar et al. 2016. Mol.Biol.Evol. 33:216-227.

O-01-WF05

Distinguishing among evolutionary forces acting on genome-wide base composition: Computer simulation analysis of approximate methods for inferring site frequency spectra of derived mutation in recombining regions

Tomotaka Matsumoto[1], Hiroshi Akashi[1]

[1]National Institute of Genetics (Japan)

Within-species DNA variation (polymorphism) data allows quantification of fundamental evolutionary forces including mutation, recombination, and fixation biases (i.e., natural selection and biased repair) and advances in technology provide such data on a genome-scale. We employ simulated data to test the reliability of ancestral inference methods with special consideration of genomic regions that undergo recombination and in which base composition is biased; a substantial fraction of eukaryotic genomes fall within this category. A heuristic ancestral state reconstruction approach that incorporates population frequency information appears to provide accurate and robust inference of ancestral and derived states for population genetic analysis. This method employs the "bifurcating tree (BT)" approach but weights the estimated probabilities of ancestral state by an expected site frequency spectrum. Analyses of simulated data show that polarization by the "bifurcating tree with weighting (BTW)" method allows accurate estimation of unfolded SFS in several evolutionary scenarios of base composition, including different fixation biases among mutation categories, heterogeneous selection pressure among lineages and population size expansion/reduction.

We apply the BTW method to genome polymorphism data of *Drosophila melanogaster*and *D. simulans* to study the forces acting on the evolution of intron base composition and codon usage bias.

O-01-WF06

Direct measurement of fitness effects of natural variation through precision genome editing

Shi-An A Chen[1], Eilon Sharon[2, 1, 3], Neil M Khosla[1], Jonathan K Pritchard[2, 1, 3], Hunter B Fraser[1]

[1]Stanford University (United States), [2]Stanford University (United States), [3]Stanford University (United States)

How natural genetic variation contribute to individual phenotypes is a central question in genetics and evolution. Current methods for mapping of causal variants often has insufficient resolution, limiting our ability to pinpoint causal variants. We used a novel high-throughput, precision gene editing methods to map causal variants genome-wide, at single-nucleotide resolution. With our new method, tens of thousands of genetic variants between two diverged yeast strains were functionally assayed in a single experiment. Surprisingly, we observe widespread fitness effects for synonymous variants. Moreover, synonymous variants were as likely to affect growth fitness as missense variations in our screen. These results challenge the classical assumption that synonymous variants are neutral. We further apply this framework to detect condition-specific causal variants, which would provide insight into how synonymous variation interacts with environmental factors.