POB-001

Evolutionary Dynamics of the CDPK Signalling Network in Plants

Bernhard Wurzinger[1], Ingo Ebersberger[2], Katrin Schrader[3], Markus Teige[1]

[1]University of Vienna (Austria), [2]Goethe University Frankfurt am Main (Germany), [3]University of cologne (Germany)

Plants are exposed to constantly changing external and internal stimuli. In order to thrive within these dynamic conditions the stimuli have to be perceived and decoded properly. To this end, they heavily rely on Ca^{2+}-signals and Ca^{2+}-dependent protein kinases (CDPKs) for decoding of those signals. Studying the molecular events downstream these CDPKs remains challenging due to their redundancy and current limitations in phospho-proteomics. In order to obtain a broader view on the substrates of CDPKs we performed an evolutionary guided approach by tracing an experimentally determined CDPK substrate consensus motive through the proteomes of 38 plant species ranging from higher plants to green algae. The experimental validation of this approach revealed that substrate specificity of the chosen CDPKs is retained within the plant kingdom including green algae despite a substantial increase in CDPK gene number throughout evolution from the last common ancestor of green algae and higher land plants. Furthermore, only a minority of CDPK consensus sites are found to be conserved on orthologous proteins throughout the plant kingdom. Importantly, we observe a non-negative selection for the CDPK consensus motive in all land plants and the majority of green algae. Considering that experimentally verified and functionally relevant CDPK consensus motives are restricted to distinct clades within the plant kingdom we conclude that the CDPK signalling network is highly dynamic on the substrate level, which allows for rapid rewiring of CDPK signalling cascades on evolutionary timescales.

POB-002

Deciphering evolutionary history with multifaceted evidence: an example from one drosophila group in East Asia

Jinming Lu[1], Huiluo Cao[2], Yuan Zhang[1], Hongwei Chen[1]

[1]South China Agricultural University (China), [2]The University of Hong Kong (China)

Continuous change of phenotypic characters might be correlated with adaptive process of organisms. In versus, studies on continuous morphological features could show hints into the phylogenetics of organisms. A wealth of information from molecular phylogeny and phenotypic character transformation integrated with biogeographic evidence should convince the systematic research. Here, we employ the *shirozui* group that is the largest one within the subgenus *Steganina* as an example to perform a comprehensive study on reconstructing credible phylogeny. The *shirozui* species group includes 22 described species, which are mostly distributed from the tropical to subfrigid zones in East Asia. The molecular phylogeny of the *shirozui* group was constructed using mitochondrial (*COI* and *ND2*) and nuclear (*28S rRNA*) data. According to the phylogeny, three clades were identified. Biogeographical analyses suggest that the *shirozui* group probably originate from Southwest Chin. Clade I of the *shirozui* group was proposed to disperse into Central China and South China, with subsequently vicariant events leading to the diversification in Southwest China and Central China, even further to northern China. Ancestors of Clade II and III were predicted to be mainly distributed in Southwest China, while the descendants radiated were retained in ancestral distribution and diversified. Several phenotypic characters, in particular, the male terminalia underlying the speciation process were identified with continuous changes within the *shirozui* group, consistent with biogeographical distribution of different lineages identified on the basis of the molecular phylogeny. In summary, multidisciplinary evidence resolves the phylogenetic and evolutionary puzzles of the *shirozui* group.

POB-004

Massive gene amplification on a recently formed Drosophila Y chromosome

Shivani Mahajan[1], Doris Bachtrog[1]

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

Sex chromosomes are derived from a pair of ordinary autosomes. However, after recombination is suppressed between the X and the Y, the two chromosomes evolve independently. Y chromosomes typically undergo massive gene loss and degeneration. X chromosomes, in contrast, retain most of their ancestral genes. Old Y chromosomes generally have very few genes and are highly repetitive. However, evolutionary studies involving sex chromosomes, and in particular Y chromosomes have proven challenging due to the lack of high quality reference assemblies. *Drosophila miranda* has served as a model to study sex chromosome differentiation, but its young neo-Y chromosome, which is repeat rich but also retains significant homology to its neo-X chromosomes proved especially difficult to assemble using traditional short read sequencing methods. In this study, we combined single-molecule sequencing together, Hi-C analysis and BAC clone sequencing to assemble the neo-Y chromosome of *D. miranda*. We show that the Y chromosome in *D. miranda* has increased almost 3-fold in size, driven mostly by widespread accumulation of repeats, but also by gene amplification. Although widespread loss of genes is considered a hallmark of Y chromosome evolution, we show that the initial stages of Y evolution are driven by massive amplification of testis-specific genes. We show that amplified genes are enriched for meiosis functions, are often co-amplified on the X, and are often targeted by piRNAs. This suggests that their amplification is likely driven by X vs. Y antagonism, and implicates meiotic drive as an important force in sex chromosome evolution.

POB-005

Template switching in genome instability and evolution

Sumita Omer[1, 2], Bar Lavi[1], Piotr A Mieczkowski[3], Shay Covo[2], Einat Hazkani-Covo[1]

[1]The Open University of Israel (Israel), [2]Hebrew University (Israel), [3]University of North Carolina at Chapel Hill (United States)

Okazaki fragments that are formed during lagging strand DNA synthesis include an initiating primer consisting of both RNA and DNA. The RNA fragment must be removed before the fragments are joined. In Saccharomyces cerevisiae, a key player in this process is the structure-specific flap endonuclease, Rad27p. To obtain a genomic view of the mutational consequence of loss of RAD27, a S. cerevisiae *rad27-null* strain was subcultured for 25 generations and sequenced using Illumina paired-end sequencing. We were able to detected a previously neglected class of template-switching events. These events were presumably generated by quasi-palindrome to palindrome correction, as well as palindrome elongation. The formation of these events is best explained by folding back of the stalled nascent strand and resumption of DNA synthesis using the same nascent strand as a template. Evidence of quasi-palindrome to palindrome correction that could be generated by template switching appears also in yeast genome evolution.

POB-007

Dynamics of cis-regulatory regions for introducing the divergent transcription factor motifs in the human genome

Jia-Hsin Huang[1], Ryan Shun-Yuen Kwan[1], Zing Tsung-Yeh Tsai[2], Huai-Kuang Tsai[1]

[1]Academia Sinica (Taiwan), [2]University of Michigan, Ann Arbor (United States)

Divergence in transcription factor (TF) repertoires such as altering binding specificities (motifs) provides major sources to shape the gene regulatory evolution in eukaryotes. However, it is currently unclear how dynamic changes of cis-regulatory regions by introducing novel TF binding motifs in the genome over evolutionary time. According to that the primary DNA binding specificities of TFs with similar DNA binding domain sequences are generally conserved between distantly related species, we proposed a novel metric index to estimate the evolutionary divergence level of TF binding motifs in humans. Our results of *in silico* motif scan and empirical TF-ChIP (chromatin immunoprecipitation) show that the divergent motifs tend to be introduced at the borders of the cis-regulatory regions and likely affiliate the expansion of TF binding regions. We suspect that an expanded mechanism by incorporating more divergent motifs at the borders of cis-regulatory regions serves as common evolutionary intermediates in rewiring regulatory transcription circuits.

POB-008

From higher-order organisms to microbes: a novel quantitative species identification method based on ancient DNA

Evangelos Antonios Dimopoulos[1], Irina Velsko[1, 3], Evan Irving-Pease[1], Laurent Frantz[1, 2], Greger Larson[1]

[1]The University of Oxford (United Kingdom), [2]Queen Mary University of London (United Kingdom), [3]Clemson University (United States)

Genetic species identification of archaeological specimens is often difficult due to low DNA content and degradation. Yet specific and accurate identification of samples is essential not only for taxonomic identification, but also for elucidating community membership of ancient microbial metagenomic samples. Therefore, we created a method for identifying both higher-order organisms and microbes via aDNA, by investigating the presence of specific species in archaeological material, and quantifying the confidence of the performed identification. We present two case studies to highlight the utility of our pipeline in archaeological studies.

Lively debate surrounds the introduction of non-indigenous domestic livestock to southern Africa, and it has been hypothesized that the frequency of domestic stock in faunal assemblages has been overestimated by zooarchaeologists, with recent genetic studies discrediting previous taxonomic assignments. After analyzing samples from this area, and trying to quantify how well supported a genetic taxonomic identification was, it was suggested that more DNA data need to be generated from them, to confidently reject the morphological identifications.

By using ancient dental calculus samples, we have attempted to robustly identify bacterial species within the sequences that have been generated from the substrate, and estimate how reliable a species assignment is. Thus far, we have detected a positive signature of *Corynebacterium diphtheriae*, the causative agent of diphtheria and a transient member of the oral biofilm. Our findings highlight the potential of ancient dental calculus to act as a reservoir for respiratory pathogens, and the value of screening this substrate for species of epidemiological interest.

POB-009

CATANA: Comprehensive Alternative Transcript Atlas based oN Annotation

Cheng-Kai Shiau[1, 2, 3], Jia-Hsin Huang[1], Huai-Kuang Tsai[1]

[1]Academia Sinica (Taiwan), [2]Academia Sinica (Taiwan), [3]National Yang-Ming University (Taiwan)

In higher eukaryotes, the generation of alternative gene products from a single gene locus is a prevalent mechanism to increase the functional and regulatory diversity of the genome. The transcriptome diversity in the genome comprises the use of alternative transcriptional initiation/termination (AT) and alternative splicing (AS) through selectively removal and reconnection of exons. However, no tool is available to identify all AT and AS events. Here, we develop CATANA, a tool for comprehensive alternative transcripts annotation, by extracting either genome annotation files from annotation authorities and transcript annotations generated from transcriptome analysis tools, such as Cufflinks or StringTie. CATANA extracts AT/AS events by the following steps: (1) flattening gene model and calculating numbers of exon-intron junctions; (2) generating exon usage pattern; and (3) calling AT/AS events. CATANA could provide the alterative transcript annotations comprehensively to facilitate the study of function and regulation on alternative exon usage from RNA-seq data in both annotated and unannotated genomes.

POB-010

The mayfly Cloeon dipterum as a platform to study evolutionary innovations: novel sexually dimorphic organs and the origin of insect wings

Isabel Almudi[1, 5], Isabel Garcia[1, 5], Carlos Martin-Blanco[1, 5], Kristofer Davie[2], Ferdinand Marletaz[3], Fernando Cruz[4], Tyler Alioto[4], Stein Aerts[2], Fernando Casares[1, 5]

[1]Andalusian Centre for Developmental Biology (Spain), [2]University of Leuven (Belgium), [3]Okinawa Institute of Science and Technology (Japan), [4]CNAG-CRG (Spain), [5]Decision Making in Cell Collectives/DMC2 (Spain)

Evolutionary innovations are biological revolutions: new organs are critically associated with the emergence of new species and their exploitation of new niches, however, the origin of morphological novelties is a long-standing question in Evolutionary Biology. The mayfly species Cloeon dipterum is a privileged model to study evolutionary innovations. First, these mayflies display one of the most striking examples of a sexually dimorphic novel structure. Cloeon males develop, in addition to the compound eyes, an extra pair of dorsal, turban-shaped eyes. Furthermore, as ancestral mayflies were the first insects in which wings evolved, they are key to understand the origin of insect wings, which led to the conquest of the sky and the adaptation to a huge diversity of new ecological niches. To answer these questions, we have successfully established C. dipterum as a model species with a continuous culture in the lab. We have sequenced and assembled a high quality reference genome and generated a wide transcriptomic profiling of Cloeon tissues and developmental stages. Using these datasets, we compare sex-specific gene expression in nymphal heads, with a special focus on genes of the highly conserved Retinal Determination Network (RDN), to show how RDN elements could have evolved to play a role in the origin of this dimorphic organ. Finally, we provide the first analyses of the genomic bases underlying the origin of insect wings, providing an unprecedented opportunity to test the different proposed hypotheses about their origin and paving the way for a better understanding of this key evolutionary innovation

POB-011

Genetic mechanism of achiasmy in *Drosophila albomicans*

Arika Inuyama[1], Yosuke Seto[1], Yoshitaka Ogawa[1], Masafumi Nozawa[1, 2], Koichiro Tamura[1, 2]

[1]Tokyo Metropolitan University (Japan), [2]Tokyo Metropolitan University (Japan)

The rate of meiotic recombination is often different between sexes. This phenomenon is called heterochiasmy. In the most extreme case, heterogametic sex completely lacks meiotic recombination, which is called achiasmy. Achiasmy is believed to be quite important at the initial stage of sex chromosome evolution, because sex determination would be violated by recombination between sex chromosomes on the way of divergence. Once sex chromosomes are diverged enough, however, achiasmy is unlikely to be important because recombination between sex chromosomes no longer occur due to their high sequence divergence. Achiasmy and heterochiasmy are widely observed in many different groups of organisms but their underlying genetic mechanisms largely remain unknown. In this context, two *Drosophila* species belonging to the *Drosophila nasuta* subgroup are ideal materials to tackle this question, because *D. albomicans* do not exhibit meiotic recombination in males, whereas its sibling species, *D. nasuta*, do it. It has been reported that male recombination seems to have also occurred in the common ancestor of these two species. We therefore investigated whether *D. albomicans* acquired a mechanism of recombination suppression or lost a mechanism of recombination. To test these hypotheses, we generated a hybrid of these two species and examined whether the hybrid males exhibit recombination based on extensive sequence analyses. The results showed that the hybrid males do not exhibit recombination, indicating that "no recombination" is a dominant state. We tentatively conclude that *D. albomicans* acquired the mechanism of recombination suppression.

POB-012

Novel experimental Ribo-seq measurements demonstrate genomic adaptation to decrease ribosome queuing

Alon Diament[1], Anna Feldman[1], Elisheva Schochet[2], Martin Kupiec[3], Yoav Arava[4], Tamir Tuller[1, 5]

[1]Tel Aviv University (Israel), [2]Tel Aviv University (Israel), [3]Tel Aviv University (Israel), [4]Technion-Israel Institute of Technology (Israel), [5]Tel Aviv University (Israel)

Ribosome queuing is a fundamental phenomenon suggested to be related to topics such as genome evolution, synthetic biology, gene expression regulation, intracellular biophysics, and more. However, this phenomenon hasn't been quantified yet at a genomic level. Current methodologies for studying translation (e.g. ribosome profiling) are usually calibrated to capture only single ribosome protected footprints (mRPFs) and thus limited in their ability to detect ribosome queuing. Here we present an experimental-computational approach for studying ribosome queuing based on sequencing of RNA footprints extracted from pairs of ribosomes (dRPFs) using a modified ribosome profiling protocol. We combine our approach with traditional ribosome profiling to generate a detailed profile of ribosome traffic, and analyze the data using computational models of translation dynamics.

Data collected from *S. cerevisiae* show that ribosome queuing is more frequent than previously thought, suggesting that at least one to five translating ribosomes is in a traffic jam; these queued ribosomes cannot be captured by traditional methods. While queuing is related to higher density of ribosomes on the transcript (characteristic of highly translated genes), we report cases where traffic jams are relatively more severe in lowly expressed genes and possibly selected for. Our analysis also demonstrates that higher adaptation of the coding region to tRNA levels is associated with lower queuing. Our analysis also suggests that the *S. cerevisiae* transcriptome undergoes selection for eliminating traffic jams. Thus, our proposed approach is an essential tool for high resolution analysis of ribosome traffic during mRNA translation and understanding its evolution.

POB-013

Experimental evolution of Escherichia coli mutators in a complex environment

Wei-Chin Ho[1], Megan G. Behringer[1], Samuel F. Miller[1], Michael Lynch[1]

[1]Arizona State University (United States)

How mutation rates affect adaptation is equivocal. While an elevated mutation rate may decrease the waiting time of beneficial mutations, the accompanying mutation load can inhibit adaptive evolution by inducing strong clonal interference in bacteria. This paradox is reinforced in complex environments, due to the interactions between a higher number of beneficial mutations and a more rugged fitness landscape. To study the impact of mutation rates on bacteria evolution in complex environments, we evolved *Escherichia coli* in Luria-Bertani media with or without the background deletion of *mutL*, which causes a ~100-fold difference in mutation rates. Three population-genetic environments were created by different daily transfer volumes. Using pooled-sequencing performed every 100 days for up to 600 days (~2,000-14,000 generations), we found that mutations accumulated linearly in general. The *ΔmutL* lines accumulated mutations faster than wild-type only by 5-12 folds, suggesting that their mutations experienced more purifying selection or less beneficial selection. Using synonymous mutations as a control, we found that *ΔmutL* lines accumulated nonsynonymous mutations significantly faster in the highest transfer volume but slower in the lowest transfer volume. On the contrary, they accumulated structural mutations at a slower rate in the highest transfer volume (largest population size). Moreover, the mutational spectra of evolved lines are highly similar to the corresponding mutational inputs for wild-type lines but not for *ΔmutL* lines, suggesting weaker mutational constraints in mutators. Together, these results demonstrate that the mutation-rate effect varies depending on the population-genetic environment in a complex environment.

POB-014

Understanding the genomic basis of convergent evolution and evolutionary innovation across the Metazoa Tree of Life: an all-phyla approach.

Rosa Fernandez[1], Toni Gabaldon[1]

[1]Center for Genomic Regulation (Spain)

Animals are fascinating creatures that represent one of the main groups of eukaryotes and that have dwelled on Earth over 600 million years. Modern animals can be grouped by the arrangement of their body plans into more than 30 phyla, each one exhibiting different morphology, anatomy and physiology. While some phyla are charismatic and well-known, many of them have been rarely included in comparative studies. Consequently, we lack sufficient understanding of the genetic basis of the evolution of body plans across all phyla. To shed new light on the evolutionary innovations and convergent evolution across the Animal Tree of Life, we undertook a comparative genomics approach. We compared 200 transcriptomes including representatives of all animal phyla. Moreover, we interrogated multiple genomes representing all major animal lineages through the lens of comparative genomics. Our results, based on more than 3.8 million transcripts and 260,000 gene trees, showed a bipolar distribution of gene gain across all nodes in all lineages: the largest proportion of the genome in every animal clade was formed by lineage-specific and orphan genes, followed by genes that were recruited in the first earliest-splitting nodes of animals and their closest unicellular relatives. Little addition to gene content occurred in swallower nodes. Moreover, we found that lineage-specific gene expansions were enriched in functions related to the evolutionary novelties in each lineage, such as the neural system in Ctenophora and Cnidaria or the ecdysis-related machinery in Ecdysozoa. We will discuss these results from a systematic and evolutionary perspective.

POB-015

Visual adaptation of sea snake

Takashi Seiko[1], Takushi Kishida[2], Mamoru Toda[3], Mina Toyama[4], Takahiko Hariyama[4], Takashi Okitsu[5], Akimori Wada[5]

[1]SOKENDAI (The Graduate University For Advanced Studies) (Japan), [2]Kyoto University (Japan), [3]University of the Ryukyus (Japan), [4]Hamamatsu University School of Medicine (Japan), [5]Kobe Pharmaceutical University (Japan)

Sea snake in a *Elapidae* family have completely adapted to the aquatic environment from their terrestrial ancestors, although the environments between terrestrial and sea are largely different especially in the light environment. In this research, I attempted to clarify the visual adaptation of sea snakes by comparison of snake visual pigment and the light environment in sea snake habitats through measurement of the absorption wavelength.

The shift of absorption wavelength of the visual pigments is caused by amino acid sequence changes in opsin protein and difference in the retinal types (Yokoyama, 2000). Previously, I compared opsin sequences among semi-aquatic, full-aquatic, and terrestrial species closely related to aquatic species. As a result, I found amino acid substitutions at the positions where the effect on absorption wavelength shift was reported in both *LWS* and *RH1* opsin genes.

Visual pigment consists of opsin protein and retinal. The type of retinal used in the snake eyes is still unknown. So we also checked which type of retinal the sea snake is using with HPLC. As a result, sea snakes mainly use A1-retina.

Then I measured the absorption wavelength of opsin protein. To measure the absorption wavelength, I cloned two opsin genes, *LWS* and *RH1*, into pCMV plasmid vector for protein expression. I reconstituted the visual pigments using the expressed protein and retinal, and measured the visual pigments to investigate the difference in absorption wavelengths. The results of opsin protein measurement are similar to predicted shift by amino acid substitution.

POB-016

Recalibrating Quality Scores for Low-Depth Data of Ancient Genetic Material

Ilektra Schulz[1, 2], Vivian Link[1, 2], Zuzana Hofmanova[1, 2], Athanasios Kousathanas[3, 2], Christoph Leuenberger[4], Daniel Wegmann[1, 2]

[1]University of Fribourg (Switzerland), [2]Swiss Institute of Bioinformatics (Switzerland), [3]University of Lausanne (Switzerland), [4]University of Fribourg (Switzerland)

Whole genome sequencing offers an unbiased way to learn about genetic diversity and to identify past evolutionary events. Yet sequencing genomes at high depth remains costly, particularly if samples are degraded (e.g. ancient DNA) or if a large number of individuals are studied. An emerging, cost effective alternative is low-depth sequencing, which does not allow for accurate genotype calls, but is readily used to infer genetic diversity, evolutionary histories or genotype-phenotype associations if the uncertainty in genotypes is properly accounted for. Many such methods have recently been developed that integrate out genotype uncertainty through the use of genotype-likelihoods. However, all these methods rely on an accurate quantification of sequencing errors through base quality scores, and as we show here through simulations and by downsampling high depth data, even slightly distorted quality scores may strongly bias the inference of genetic diversity or the Allele Frequency Specturm (AFS) from genotype likelihoods.

We thus introduce a novel approach to recalibrate quality scores that does not rely on reference data but exploits (pseudo) homozygous regions of the genome such as mtDNA, sex-chromosomes in heterogametic individuals, or ultra-conserved elements. Our model is readily extended to account for various covariates such as the machine-provided quality score, the position within a read, the sequencing context or the total read length, to name a few. As we again show with both simulations and by downsampling high-depth samples, accurate inference of genetic diversity is indeed possible from ultra low-depth data once quality scores are accurately recalibrated.

POB-017

Multi-stage disease expression profiles revealed dynamics in connectivity of biological networks

Wen-Yu Chung[1], Ru-Feng Peng[1]

[1]National Kaohsiung University of Applied Sciences (Taiwan)

Studies in network biology have shown that the links among genes and proteins are important in evolution, development and disease models. Moreover, complex disease is often not affected by one single gene, but rather groups of related genes and their interactions. Thus, researchers often apply gene or pathway enrichment tests which aim to determine whether a set of genes is statistically significant in a particular functional group or pathway. We are, on the other hand, interested in identifying the important dynamics of connections between two or more biological networks built from multiple disease phases. First, the gene co-expression networks were constructed in all available time points. Next, we applied our scoring function to give each gene pair a predicted disease-association value by comparing the network topologies. The scoring function tested whether the alteration of links between the gene pairs in all biological networks is statistically significant. Top-scoring genes are those harbored large modifications of connections (either gain or lose) and are then provided for further analyses. The initial examination on liver cancer datasets showed that the highest scoring genes are of related functions, indicating the utility of our methods. This is the first step toward a comprehensive understanding of associated genes and critical interactions in common and complex disease.

POB-018

TIPars: Taxa Insertion Using Parsimony and Ancestral Reconstructed Sequences

Tommy T.Y. Lam[1]

[1]The University of Hong Kong (China)

The rapidly increasing genetic sequence data led to a growing demand of updating a reference phylogenetic tree with new sequences without computing the entire tree from scratch. We proposed a parsimony-based algorithm ('TIPars') to insert new taxa into a phylogenetic tree with the aid of pre-computed ancestral sequences. Compared with the maximum likelihood based methods, TIPars showed robust performance in most of the cases, and has advantage of simplicity and efficiency. This method will be useful for the phylogenetic typing of organisms in metagenomics studies especially with the aim of classification accuracy at strain level e.g. virus genomes with diverse circulating strains. TIPars is freely available at https://github.com/id-bioinfo/TIPars

POB-019

Conservation phylogenetics and computational species delimitation of Neotropical primates

Beatriz Mello[1]

[1]Federal University of Rio de Janeiro (Brazil)

Accurate definition of species boundaries plays a crucial role in the management and conservation of biological diversity. Although the science of genetics has been largely used to unveil cryptic biodiversity, phylogenetic theory has not been fully explored to address this issue. Recently, single-locus approaches to delimit evolutionary significant units (ESUs) have been widely used in several groups of organisms, presenting a valuable tool to test the uniqueness of biological clades. In this context, this study aimed to test the performance of coalescent-based delimitation methods using Neotropical primates, which is a lineage that includes several endangered species. This was achieved by evaluating the congruence between inferred ESUs and the nominal species recognized by taxonomists based on single-locus data. The results suggested that the molecular marker that best replicates standard taxonomy is the COI gene. GMYC presented the best overall performance at preventing lumping profiles, and the methodological combination that led to the lowest lumping frequency was the ST-GMYC using the timescale inferred on RelTime. The highest matching frequencies were produced by the ST-GMYC based on a BEAST timescale inferred with a Yule prior for the branching process. Overall, CSD approaches produced higher frequencies of matches and splits than lumps. The implementation automated the delimitation of evolutionary units in a circularity-free system, which is valuable for decision-making policies, enabling the refinement of an entirely taxonomy-based procedure. The results also highlighted the importance of phylogeny-based analyses for the discipline of conservation.

POB-020

Metabolic Capabilities of the Early Eukaryotes

Matthew Moore[1], David Newman[1], Martin Rusilowicz[1], James McInerney[1]

[1]University of Manchester (United Kingdom)

With the increasing number and diversity of eukaryotic genomes it's now possible to assess features that must have been present in the last eukaryotic common ancestor (LECA). Similarly, how genomes were shaped during diversification into the known eukaryotic supergroups may be inferred by comparison and modelling of extant genomes. Genome annotation and subsequent metabolic model reconstruction allows us to now assess the ancestral metabolic capabilities across the eukaryotes, providing an insight into the lifestyles and potential drivers of evolution of the earliest cells. Genome wide metabolic models were reconstructed from 318 diverse eukaryotic species genomes, based on PRIAM enzyme profiles and interproscan annotation. Widely distributed orthologs were selected (n = 247) and concatenated into a taxon sampled supermatrix. The phylogeny was approximated using rate and composition heterogeneous models. With this dataset it was possible to reconstruct the ancestral metabolic capabilities of early eukaryotic cells using MCMC multistate independent models. By modelling enzymatic and pathway evolution across the tree it has been possible to determine bursts of metabolic innovation and compare metabolic capabilities. This workflow is readily reproducible so that understanding of early metabolism can be refine as yet more diverse and deep branching eukaryotic genomes are sequenced.

POB-022

Bursts of amino acid replacements in protein evolution

Anastasia Stolyarova[1], Georgii Bazykin[1, 4], Tatyana Neretina[2, 3], Alexey Kondrashov[3, 5]

[1]Skolkovo Institute of Science and Technology (Russian Federation), [2]Lomonosov Moscow State University (Russian Federation), [3]Lomonosov Moscow State University (Russian Federation), [4] Institute for Information Transmission Problems (Kharkevich Institute) of the Russian Academy of Sciences (Russian Federation), [5]University of Michigan (United States)

Evolution can occur both gradually and through alternating episodes of stasis and rapid changes. However, the prevalence and magnitude of fluctuations of the rate of evolution remains obscure. Detecting a rapid burst of changes requires a detailed record of past evolution, so that events that occurred within a short time interval can be identified. Here, we use the phylogenies of the Baikal Lake amphipods and of Catarrhini, which contain very short internal edges facilitating this task. We detect 6 radical bursts of evolution of individual proteins during such short time periods, each involving between 6 and 38 amino acid substitutions. The "background" rate of evolution of the genes containing such bursts on the whole phylogenetic tree excluding bursts-carrying edges is sufficiently low. The bursts we detected were extremely unlikely to have occurred neutrally, and were apparently caused by positive selection. On average, in the course of a time interval required for one synonymous substitution per site, a protein undergoes a strong burst of rapid evolution with probability at least ~0.01.

POB-024

Using comparative transcriptomics and ribosome profiling to identify de novo genes in yeast

William Robert Blevins[1, 2], Xavier Messaguer[4], Jorge Ruiz-Orera[1], Jose Luis Villanueva Canas[1], Bernat Blasco-Moreno[3], Laura Avino-Esteban[2], Guillem Lopez-Grado Salinas[2], Lorena Espinar[2], Juana Diez[3], Lucas Carey[2], M. Mar Alba[1]

[1]Hospital del Mar Research Institute (Spain), [2]Universitat Pompeu Fabra (Spain), [3]Universitat Pompeu Fabra (Spain), [4]Universitat Politecnica de Catalunya (Spain)

De novo gene birth can occur when a segment of noncoding DNA undergoes a series of changes which enables transcription; this process, albeit infrequently, can lead to a new protein with a novel function. Young *de novo* genes have no homology with other genes and do not appear to be under the same selective constraints as other genes. To study the origins of *de novo* genes, we have performed deep RNA sequencing on 11 species of yeast from the phylum of Ascomycota in both rich media and oxidative stress conditions. We assembled *de novo* transcriptomes for each species to compliment the reference transcriptomes; on average, we identified 700 unannotated transcripts in each species. We also performed ribosome profiling experiments in both conditions with *S. cerevisiae* to detect which transcripts had signals of translation.

We combined BLAST homology searches with synteny-based comparisons to determine the conservation of genes at different depths in the yeast phylogeny. Of the 6987 transcripts we analyzed in *S. cerevisiae*, 257 were species-specific (188 novel+69 annotated). We identified 72 putative recently originated *de novo* genes in sensu stricto; these genes are taxonomically restricted, do not overlap transcripts in syntenic regions of closely-related species, and have signals of translations from our ribosome profiling experiments. We evaluated differential expression due to oxidative stress and found that there was no enrichment in putative *de novo* genes. Initial results from knock-out experiments on these putative *de novo* genes show that several genes confer a slight fitness advantage in normal conditions.

POB-027

Predicting Disease Causality of Mutations in Human Beta Globin gene

Sangeetha Udani Ratnayake[1]

[1]Hokkaido University (Japan)

A single amino acid change can result in disease which may destabilize the structure of a protein. The structural nature of such mutations has been studied in terms of evolutionary conservation as a key to understand the effects towards the disease likelihood. However, there is a doubt whether the level of evolutionary conservation of amino acids is always applicable to predict the patterns of disease causality in proteins. To answer the question, we investigated the relationship between amino acid exchangeability and disease and non-disease causality in terms of both evolutionary and structural information. Human beta globin is a subunit of hemoglobin which is widely studied in the globin family. Beta thalassemia is one of the major Mendelian genetic diseases in human beta globin gene: impair the functions of iron metabolism with an imbalance structure in whole hemoglobin with mutated beta subunits. In this study, we inferred the relationship between physicochemical properties and disease causality of human beta globin gene. Other than the physicochemical properties, we considered other factors, including evolutionary information, and structural information. The mutations in human beta globin genes showed complex but a consistent pattern of disease causality that correlates with the classical amino acids grouping of physicochemical properties: hydrophobicity and volume. The relationship between the disease likelihood of missense mutations and amino acid physicochemical properties could be applied to improve the modern medical investigations to understand the genesis of hereditary diseases.

POB-028

Comparative transcriptome analyses for cold acclimation response in *Drosophila albomicans*

Tomohiko Kimura[1], Yosuke Seto[1], Yoshitaka Ogawa[1], Msafumi Nozawa[1, 2], Koichiro Tamura[1, 2]

[1]Tokyo Metropolitan University (Japan), [2]Tokyo Metropolitan University (Japan)

In the recent few decades, *Drosophila albomicans* has expanded its distribution from the tropical to the temperate zone. It has been demonstrated that cold tolerance enhanced by cold acclimation is a main factor of this distribution expansion. In this study, we examine the differences in transcriptome caused by cold acclimation in *D. albomicans*, using RNA-seq of cold-acclimated and non-acclimated flies from three *D. albomicans* strains, whose response to cold acclimation vary. Comparing the differences in transcriptome caused by cold acclimation among the strains, we found that *CG7214*, *CG18609* and *CG3264* homologs were differently expressed between the acclimated and non-acclimated flies in all the three strains, among which *CG18609* is homologous to a gene of ELO gene family, which synthesizes fatty acids over C_{16} from C_{16} fatty acids. The suggested down-regulation of *CG18609* via cold acclimation is expected to cause an accumulation of C_{16} fatty acids. In addition, *FASN3* was detected to be a differentially expressed gene in WL1 strain, which showed the maximum enhancement of cold tolerance via cold acclimation. *FASN3* encodes an enzyme involved in the synthesis of C_{16} fatty acids. The suggested up-regulation of *FASN3* via cold acclimation is expected to induce the synthesis of C_{16} fatty acids. Since increase in C_{16} fatty acids in cell membrane is responsible for the enhancement of cold tolerance in *Drosophila* species distributing in the temperate zone, it is speculated that the induced C_{16} fatty acid synthesis is attributable to the enhancement of cold tolerance via cold acclimation in *D. albomicans*.

POB-029

From FECA to LECA; gauging complexity in the First, to the Last, Eukaryote Common Ancestor.

David Newman[1], James McInerney[1, 2], Fiona Whelan[1], Martin Rusilowicz[1], Matthew Moore[1]

[1]University of Manchester (United Kingdom), [2]University of Nottingham (United Kingdom)

The modern Eukaryotic cell is a structure much more complex than the typical prokaryote. The process through which this complexity arose is a major question as yet unresolved in evolutionary biology. We have created a partial reconstruction of the genome of LECA through identifying what genetic elements are conserved in contemporary eukaryote genomes across a majority of the major taxa in the eukaryotic tree of life. Our analysis suggests LECA likely possessed much if not all of this complexity and would have been little different from a modern protist. LECA appears to have had chromosomes, ribosomes, tubulin and actin cytoskeleton elements, mitochondria, a nuclear envelope and nuclear pore system, Golgi apparatus, proteasome, ER, spliceosome, and sophisticated membrane biosynthesis and trafficking. LECA seems to have potentially also been capable of complex behaviours such as mitosis, meiosis and phagocytosis. We are currently disentangling the elements that we have identified as being in LECA to identify what elements were present prior to the formation of eukaryotes and are informative about the nature of FECA and what genes and functions were novel innovations of eukaryotes proper.

POB-030

Molecular evolution of RuBisCO subunits in angiosperms

Kana Yamada[1, 2], Nicolas Salamin[1, 2], Iakov Davydov[1, 2, 4], Romain Studer[3]

[1]University of Lausanne (Switzerland), [2]Swiss Institute of Bioinformatics (Switzerland), [3]The European Bioinformatics Institute (United Kingdom), [4]University of Lausanne (Switzerland)

Multigene families play significant roles in adaptive evolution. Involvement of multigene families in evolution of photosynthesis has already been discussed. However, that of *rbcS* multigene family which encodes the small subunits of Ribulose-1,5-bisphosphate carboxylase (RuBisCO) has been studied little. Therefore, we studied the evolution of the *rbcS* multigene family and investigated the characteristics of each copy. First, the reconstruction of the phylogenetic tree in angiosperms resulted in a similar topology as the species tree. Indeed, similarities between paralogous gene copies were higher than that of orthologous gene copies. A signal of gene conversion between gene copies was not detected. We suppose that the *rbcS* gene tree may have been organized by Birth-Death evolution. Second, coevolution with the gene encoding the large subunit of RuBisCO, and the stabilities of modeled RuBisCO structures with different *rbcS* copies suggested that all the gene copies of *rbcS* may have a common function. Third, the expression levels of gene copies within species are different. Each copy changes their expression levels in different environmental conditions and tissues. Each member of the *rbcS* gene family may have adapted to the environment in different manners by changing their gene expression levels. The comparison of expression between species suggested that multiple-*rbcS* copies may not exist for dosage compensation of single-*rbcS* copy. Finally, positive selection was detected in some but not all the *rbcS* gene copies of C4 species. The transition of photosynthetic type may not be the factor of selective pressure in *rbcS* evolution.

POB-031

Extent Mitochondrial plastid combined insertions into the nuclear genome

Shir Portugez[1, 2]

[1]Tel Aviv University (Israel), [2]The Open University of Israel, Ra'anana (Israel)

Mitochondrial and plastid DNA is continually transferred to the nuclear genomes of eukaryotes giving rise to numts and nupts, respectively. Both numts and nupts were shown to integrate into double strand breaks in the nuclear genome through a DNA repair mechanism known as nonhomologous end joining mechanism (NHEJ). Short homology (microhomology of two to ten nucleotides) in the site of integration is a known hallmark of NHEJ. Sporadic cases of loci containing both mitochondrial and plastid fragments were previously reported, but the extent and the molecular mechanism of these mix insertions are unclear.

To characterize such mosaic insertions, we identified all organelle fragments in the nuclear genome of nine plants. These events were classified as mitochondrial (numts), plastid (nupts) or mosaic insertions, in which homology to both organelles was found at the same nuclear loci. Suprisingly, our analysis shows that mosaic insertions are frequent (3-10% of insertion events) and appear in all analyzed genomes.

We next compared the junctions of complex insertions from single organelle origin to junctions of dual organelle origin in mosaic insertions. All Junctions were found to be characterized by the same pattern of microhomology as the pattern of complex events from a single organelle origin. Therefore, despite their dual origin, mosaic insertions are integrated into the nuclear genome in the same NHEJ mechanism as numts and nupts. Our results suggest that the meeting of mitochondria and plastid fragments is necessary before integration of mosaic insertions; however, it is unclear where and when this meeting occurs.

POB-032

The Caenorhabditis Genomes Project

Lewis Stevens[1], Mark Blaxter[1]

[1]University of Edinburgh (United Kingdom)

The completion of the *Caenorhabditis elegans* genome in 1998 remains one of the milestones of modern biology, and it has since become an essential backdrop to the vast body of work on this key model organism. However, the *C. elegans* genome is merely a single example of how a genomic system generates a complex organism, and a complete understanding of its structure and function will only be achieved when studied in an evolutionary context. As part of an international collaboration, we are generating that context - by sequencing the genomes of all 50 known species of *Caenorhabditis*. Our data are enabling the study of the evolution of genome size, and its relationship with reproductive mode and the proliferation of transposable elements. We have also studied the evolution of gene structure, and revealed rapid and extensive intron loss in many species, including in *C. elegans*. We are analysing the evolutionary forces shaping protein-coding gene content, including gene birth/death and gene family expansion/contraction. By employing long-read sequencing to generate assemblies with chromosome-scale contiguity, we are now able to study the evolution of large-scale chromosome organisation. Our data, all made publicly available prior to publication at caenorhabditis.org, have become a key resource for those biologists who seek to understand the evolutionary forces that have shaped the *C. elegans* genome.

POB-033

Amino acid exchangeabilities vary across the tree of life

Zhengting Zou[1], Jianzhi Zhang[1]

[1]University of Michigan (United States)

The relative exchangeability (RE) between a pair of amino acids is the fixation probability of mutations converting between these amino acids, relative to the average fixation probability of all amino acid-altering mutations. Different amino acid pairs have drastically different REs, and accounting for this variation is an important and common practice in identifying homologs, aligning sequences, reconstructing phylogenies, and testing Darwinian selection, among other things. In all such endeavors, however, REs have been considered invariant across the tree of life, an assumption that has yet to be scrutinized. We utilized a maximum likelihood (ML) method to estimate REs from the concatenated protein sequence alignment of two taxa, and conducted extensive computer simulation to demonstrate the reliability of the ML method. We then applied this method to 136 genome sequences and estimated REs from 15, 1, and 52 clades in Eukarya, Archaea, and Bacteria, respectively. Contradictory to common belief, substantial and statistically significant differences in REs are found between clades. Interestingly, REs are more different for one-to-one orthologous proteins between clades than for unrelated proteins within clades, suggesting genome-wide, clade-specific factors underlying RE differences. We propose that among-species variations in nucleotide, codon, and amino acid usages (due to mutation bias and/or selection) cause different species to be differentially sensitive to changes in certain amino acid properties such as hydrophobicity, polarity, and size, impacting REs. The revelation of among-species RE variations cautions against blindly assuming constant REs in various analyses and suggests a higher-than-expected diversity in the mode of proteome evolution.

POB-034

Comparative plastomics of yews reveals a genome structural polymorphism and effective super-barcode

Chung-Shien Wu[1], Lian-Ming Gao[2], Chao-Nan Fu[2], Yu-Wen Chang[1], Shu-Miaw Chaw[1]

[1]Academia Sinica (Taiwan), [2]Kunming Institute of Botany, Chinese Academy of Sciences (China)

Taxus (yew) is the species-richest genus in the family Taxaceae. Because of the morphological similarity of *Taxus* entities, it is difficult to distinguish the yew species. In this study, we sequenced plastid genomes (plastomes) from three to four geographically remote individuals for each of the sixteen *Taxus* species, including two cryptic ones. This comprehensive sampling comprises all of the known *Taxus* species around the world. Our comparative analyses show several gene loss events independently occurred in *Taxus*, resulting in its plastid gene number smaller than other genera in Taxaceae. In *Taxus*, we found two plastome forms A and B that differ by the orientation of a 35 kb fragment flanked by trnQ-IRs. PCR and read mapping results reveal coexistence of A and B forms within each individual. Moreover, the relative abundance between these two forms has repeatedly shifted at both intra- and inter-specific levels, indicating a plastome structural polymorphism within *Taxus* species. Utility of the whole plastome sequence as a super-barcode gave a fully successful rate for discrimination of *Taxus* species. Such a full rate was also observed in three genes *accD*, *ycf1*, *ycf2*, and one intergenic spacer *rrn16-rrn23*, suggesting that they be promising specific DNA barcodes for yews. However, only two of these four single loci contain a barcoding gaps between intra- and inter-specific pairwise distances. As a result, we propose that using barcoding gaps as a criterion to explore barcoding loci may overlook some potential candidates.

POB-036

Determining the genomic architecture of complex traits in multiple ethnic human populations

Michael C Turchin[1, 2], Sohini Ramachandran[1, 2]

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

Genome-wide association studies (GWAS) have identified thousands of significant genetic associations in humans across a number of complex traits, including height, BMI, and type 2 diabetes. However, the vast majority of these studies have been conducted in datasets of predominantly European ancestry. It has generally been thought that complex trait genetic architecture should be transferable across populations of different ancestries; but recent work has shown a number of differences between ethnic groups, including heterogeneity in both the causal variants being discovered and in the effect size estimates for many overlapping variants. Here, we work to precisely identify and compare the genetic architecture of complex traits in multiple human ethnic groups.

Specifically, we use recently published pathway, shrinkage, and multivariate methods to compare height and BMI effect size estimates from different ethnic groups in the UKBioBank dataset. By using these approaches we are able to overcome the imprecision inherent to the smaller sample sizes currently available in multi-ethnic cohorts. We find further evidence that for highly polygenic traits there are indeed different causal variants between ethnic groups, and even for overlapping variants, variation between effect size estimates. By comparing different sample size subsets of the European cohort, we show how these methods can incorporate and improve upon imprecise estimates to consistently reveal complex trait architecture. This work sheds new light on the importance and feasibility of utilizing different ethnic groups in evaluating the full scope of human polygenic architecture.

POB-038

Exploring the genetic basis of environmental sensitivity

Amanda J Lea[1, 2], Julien F Ayroles[1, 2]

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

Both evolutionary theory and empirical work suggest that environmental challenges can create genotype-by-environment (GxE) interactions, such that not all genotypes respond identically to novel or stressful conditions. These context-dependent alleles are thought to be important in explaining inter-individual variation in environmentally-induced disease risk, but few studies are designed to detect GxE effects and their phenotypic contribution remains poorly understood. Here, we use an experimental, in vitro system to identify and characterize context-dependent alleles across a number of controlled environmental exposures. Specifically, we treated panels of immortalized lymphoblastoid cells from genetically diverse HapMap populations to seven different stimuli (ranging from immune signaling molecules to environmental contaminants), and measured genome-wide gene expression. We identified many genes with evidence for GxE interactions in one treatment condition, as well as cases in which genetic variation predisposed individuals toward transcriptional sensitivity or robustness across multiple environmental exposures. Both environmentally sensitive genes (i.e., those that were differentially expressed following any of the seven treatments) and genes with evidence for GxE interactions were more likely to overlap previous GWAS hits, supporting their contribution to disease and complex trait variation. Further, we find that variants that control the magnitude of an individual's environmental response often carry signatures of recent positive selection. Together, our results build on previous studies (which have focused on either a limited number of exposures or a limited number of genotypes) to improve our understanding of how genetic background and environmental perturbations combine to influence phenotypic diversity.

POB-039

Patterns of robustness and deregulation in gene expression networks under dietary stress

Luisa F Pallares[1], Anett Schmittfull[1], Serge Picard[1], Julien F Ayroles[1]

[1]Princeton University (United States)

At the functional level, differences in sensitivity to environmental stress likely emerge from the disruption of regulatory systems, where more sensitive individuals have decreased transcriptional robustness. However, the degree to which robustness is under genetic control remains unclear. To overcome the limitations of previous studies addressing this question, we developed two resources: First, a synthetic outbred population of Drosophila melanogaster that allows us to move away from inbred lines and skewed allele frequency spectrums, and allow us to assay individual outbred flies. Second, we have developed a new method for 3 TagSeq to generate gene expression data from single fly heads at a fraction of the cost of existing methods; this allowed us to screen an unprecedented number of flies. Here, we explore how gene expression in the brain responds to dietary stress by exposing flies to a high-sugar diet compared to the standard diet. We obtained RNA-seq data for thousands of individual fly heads in each environmental condition. The scale of this study allowed us to identify not only condition-dependent eQTL for individual genes, but also condition-dependent co-expression networks. Interestingly, the response of the heads transcriptional profile to high-sugar diet, points towards the regulation of lipid metabolism. This suggests that the fly brain might be controlling the processing and storage of lipids in the body when exposed to stress. By using a systems genetics approach, we explored the general landscape of stress response providing a comprehensive picture of transcriptional robustness in Drosophila melanogaster

POB-040

Spatio-temporal biases in the dynamics of horizontal gene transfer

Alexander Esin[1, 2], Tobias Warnecke[1]

[1]MRC London Institute Of Medical Sciences (LMS) (United Kingdom), [2]Imperial College London (United Kingdom)

Horizontal gene transfer (HGT) is a key driver of genome evolution in bacteria. Importantly, this process is not random. For example, gene product function and donor-recipient genome similarity both affect the probability of HGT. In contrast, we know surprisingly little about where in the host genome horizontally transferred genes are likely to integrate on arrival and whether they are more likely to survive in some locations than others. Understanding the interplay between HGT and genome topology would not only improve our understanding of bacterial evolution but might also serve as a practical guide for genome engineering. Here, we use a stringent phylogenetic and model-based approach to predict HGTs from all assembled prokaryotic genomes into a single bacterial group, *Geobacillus*. Focusing on HGT into a model group allowed us to extensively explore parameter space and predict gene history with much greater confidence than would be possible otherwise. We find a strikingly non-random spatial distribution of transferred genes: HGT-derived genes are enriched near the terminus and, surprisingly, near the origin of replication (although strongly depleted in its direct vicinity). In contrast, they are broadly depleted in the replichore flanks. We find that recent HGTs are more promiscuous in location, indicating that there is no initial spatial barrier to acquisition but rather that selection maintains HGT-derived genes in distinct domains in the longer term. We further show that horizontally transferred genes with different functions have distinct spatial biases and suggest that characterizing such spatio-functional biases in HGT can facilitate intelligent genome engineering.

POB-041

Gene duplication and subfunctionalization of the centromeric histone gene in Drosophila virilis

Lisa Kursel[1, 2], Harmit Malik[2]

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

Gene duplication is an abundant source of new genes and new functions. I leveraged my unexpected discovery of gene duplications of the centromeric histone, *Cid*, to explore its multiple, potentially antagonistic functions. Although Cid is essential for chromosome segregation, it evolves rapidly, possibly due to its role as a meiotic drive suppressor. Cid's rapid evolution could be highly deleterious for its mitotic function. I reasoned that gene duplication could resolve this incompatibility if one paralog functions as a meiotic drive suppressor, while the other carries out its canonical, mitotic centromeric role. I found that *Cid* underwent four independent gene duplications across Drosophila evolution. All Drosophila *Cid* paralogs localize to centromeres, yet they encode distinct N-terminal tails and evolve under different evolutionary constraints. I tested my sub-functionalization hypothesis using genetics and cytology to functionally interrogate two Cid paralogs (Cid1 and Cid5) in *D. virilis*. *Cid1* is ubiquitously expressed whereas *Cid5* is only expressed in the male germline. I found that Cid1 and Cid5 have distinct localization patterns even within the male germline. Cid1 localizes exclusively to centromeres in mitotic cells but is replaced by Cid5 in meiotic cells and developing sperm. To further test this specialization, I have generated transgenic D. virilis to knockdown *Cid1* and *Cid5*. My findings reveal a dramatic example of subfunctionalization, in which Cid1 performs mitotic functions exclusively, whereas Cid5 functions in meiosis and trans-generational inheritance. My findings reveal the unexpected tension between multiple centromeric functions on single *Cid* genes, which are resolved via gene duplication.

POB-043

Genetic ancestry in cancer cell lines

Michael D. Kessler[1], Nicholas Bateman[2], Julie C. Dunning Hotopp[1, 3, 4], Timothy D. O'Connor[1, 4, 5]

[1]University of Maryland School of Medicine (United States), [2]Henry Jackson Foundation (United States), [3]University of Maryland School of Medicine (United States), [4]University of Maryland School of Medicine (United States), [5]University of Maryland School of Medicine (United States)

Cell lines are an essential resource for studying cancer biology and the genomic variation that drives oncological disease as well as developing cancer treatments, including pharmaceuticals. However, many of these cell lines are of unknown ancestral origin which impacts the interpretation of different hypotheses and analyses. The Cell Line Project of COSMIC has collected and curated over 1000 cancer cell lines and produced genome-wide chip data for these samples. We estimated the first comprehensive ancestral annotation of these cancer cell lines by combining these samples with the 1000 genomes project and a Native American cohort with chip data. In 11 pairs of cell lines, we identified closer than average relationships, usually indicative of first cousin kinship. After removing one from each of these pairs and performing principal component and admixture analyses, we find that the correct ethnicity is reported for most of the 312 cell lines with a described ancestry, but the majority of cell lines did not have a described ethnicity and of those 215 are East Asian, 30 are African or African American, and 453 are European. We also identified low, but universal levels, of contamination or batch effects across the cell lines. We find that these cell lines are predominantly of European ancestry, although there is a substantial proportion that are East Asian in origin. As investigators consider which cell lines to use for their respective experiments, it is critical that they have accurate ancestry information that can facilitate appropriate comparisons and study designs.

POB-044

Using Homology Search to Identify De Novo Genes

Caroline M. Weisman[2], Andrew W. Murray[2], Sean R. Eddy[1, 2]

[1]Harvard University, Howard Hughes Medical Institute (United States), [2]Harvard University (United States)

Genes that have emerged from previously non-genic sequence, or de novo genes, are an intriguing potential source of molecular innovation during evolution, but much about them, including how common they are, how they emerge, and what they do, remains poorly understood. A clear first step to answering these questions is to identify a set of genes that are good candidates for having originated de novo.

One common method for identifying de novo genes is to use homology search algorithms like BLAST: if a gene lacks an identifiable homolog outside of a clade, it is inferred to have originated de novo along the founding lineage. However, recent work has shown that this method may be prone to errors in which a homolog is actually present outside the clade but is not detected, resulting in the gene being falsely called de novo.

We present a tool to determine how susceptible a given gene is to this form of false de novo classification: the probability that, even if its homolog were present outside of a clade, it would not be detected. Previous work has addressed this question using simulation-based approaches, but has also demonstrated that results are sensitive to simulation parameters, the optimal choice of which remains unknown. We use an alternative and faster approach, and also present a method to quantitatively assess result accuracy, reducing concern about parameter choice. Our work decreases the risk of this mode of false de novo classification, enabling more robust homology search-based de novo gene identification.

POB-048

Phylogenetically young genes and their function in Anopheles mosquitoes.

Elzbieta Krzywinska[1], Jaroslaw Krzywinski[1]

[1]The Pirbright Institute (United Kingdom)

We analyzed male and female transcriptomes of the African malaria mosquito, *Anopheles gambiae*, sampled at 20 time points, encompassing all developmental stages, to identify over 500 testis-specific genes. Phylogenomic comparisons across 20 mosquito and 5 other dipteran species revealed that at least 24 of those genes are phylogenetically young and limited to mosquito lineages. RNA *in situ* hybridization revealed that they are all transcribed in the primary spermatocytes. We are currently studying the *in vivo* function of these recently emerged genes. To date, the results of mutagenesis through TALENs- or CRISPR/Cas9-mediated knock-in of a fluorescent marker expression cassette, to disrupt the reading frame, resulted in no detectable phenotypes. Thus, these genes must perform non-essential reproductive functions, similar to many recently emerged genes in *Drosophila*.

POB-049

Layout of coexpressed gene modules mirrors ontogeny and organ system classification of animal tissues

Ben-Yang Liao[1], Ting-Yan Chang[1]

[1]National Health Research Institutes (Taiwan)

Modules in coexpression gene networks have been used to group genes with related functions for candidate gene prioritization and to understand the evolutionary conservation of genetic programs. However, the biological significance of the layout of modules in a coexpression gene network has not been characterized and described. Based on the expression profile of adult mouse tissues, we constructed the weighted coexpression network of mouse genes, identified the modules within, and assigned the associated tissue/organ functions of each gene module by overrepresented tissue/organ-level phenotypes derived from functional assays of the member genes. For each module pairs of different assigned tissue functions, we calculated the network distance between tissues/organs derived from the averaged coexpression distance between the member genes of the module pairs examined (Tmod). Tissue/organ pairs of the same organ system generally showed smaller Tmod in comparison with those of different organ system. In addition, we observed a significant positive correlation between Tmod and the distance between tissues in ontogeny (Tdev). Further analysis showed that the relationships between Tmod versus organ system classification and Tmod versus Tdev are independent. The correspondences of coexpressed gene modules to organ system formation and to developmental cell lineages revealed the significant functional and developmental constraints that are imposed on animal transcriptomes.

POB-051

Testing the Expensive Germline Hypothesis

Hwei-yen Chen[1], Cecile Jolly[1], Simone Immler[2]

[1]Uppsala University, Evolutionary Biology Center (Sweden), [2]University of East Anglia (United Kingdom)

The evolution of the germline is a consequence of one of the major transitions in evolution - the evolution of multicellularity. While it was long believed that the germline and the soma are separated by a barrier preventing the exchange of information between the two, evidence for a tight linkage between the germline and the soma is rapidly rising. In fact, recent findings suggest an actual trade-off between the two with the idea that the germline is expensive to maintain because it contains the foundation for the next generation. In contrast, the soma is disposable and if energy is limited, it will be channeled towards the germline. We tested the expensive germline hypothesis using the zebrafish as a model species. Using morpholinos, we created germline-free male zebrafish if some of the eggs in a clutch and reared them together with two controls, untreated embryos and buffer-injected embryos to adulthood for further testing. We exposed each of the three groups to one of two treatments, irradiation or no irradiation and tested for signs of a trade-off between the germline and the soma.

POB-053

Evolution of feeding preferences in Calliphoridae (Diptera)

Tatiana Teixeira Torres[1], Gisele Antoniazzi Cardoso[1]

[1]University of Sao Paulo (Brazil)

Flies of the Calliphoridae family are characterized by the ability of their larvae to develop in animal flesh. Members of this family can be divided into three groups based on their larval feeding habits: saprophagy, facultative ectoparasitism, and obligate parasitism. The range of life-history strategies and the appearance of obligate parasitism in at least three independent occasions in Calliphoridae make it an ideal model system for the study of the convergent evolution of phenotypic traits. We tested the hypothesis that variation in the gene expression of orthologous genes underlie the evolution of feeding preferences.

We compared gene expression profiles among five Calliphoridae species. In a candidate gene approach, we identified a single gene, Malvolio, with expression differences among flies with different feeding preferences. Using genome-wide RNA-seq data and a phylogenetic approach, we found more than 5k ortholog transcripts among Calliphoridae species. Among them, 38 and 242 were differently expressed between feeding preferences, in larvae and adult females, respectively. These genes are involved in immune response, proteolysis, and olfactory responses. In addition, we searched for evidence of positive selection in coding regions. We found a strong action of negative selection. Nevertheless, we detected 50 genes with evidence of positive selection. These genes are mainly involved in olfactory responses, metabolism of toxic substances, immune response, proteolysis, and signal transduction. The comparative analyses are still underway and will allow us to test if the convergent evolution of feeding preferences involves the same genes, or similar metabolic pathways and molecular functions.

POB-054

Evolutionary and Structural Analysis of Galectin Genes Involved in Human Immunity and Pregnancy

Zackery A Ely[1], Antonis Rokas[1], John Capra[1], Amandeep Sangha[2], Michelle Moon[1], Xingxing Shen[1], Jens Meiler[2]

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

Galectins are a family of carbohydrate binding proteins present in organisms throughout the animal kingdom. As the most widely expressed and most ancient members of the lectin superfamily, galectins are multifunctional, influencing a spectrum of biological processes. In humans, seven galectin genes are expressed in the placenta and play crucial roles in the establishment and maintenance of pregnancy. Several are also implicated in immunity related diseases, such as Crohns disease and multiple sclerosis. Though these genes are well studied in terms of their involvement in disease, we know far less about their evolution and genetic variation among human populations, and how such variation may influence protein structure. To address this, we performed an evolutionary analysis of these genes, aiming to detect signatures of ancient selection in mammalian lineages, as well as signatures of recent selection in human populations. Using molecular modeling software, we also analyzed the structural and energetic consequences of high frequency missense variants observed in different human populations.

Findings

Galectin 3 experienced accelerated evolution in placental mammals, and then experienced an even higher evolutionary rate in primates,

Galectin 8 experienced accelerated evolution specifically in primates,

Missense variants in several galectin genes are consistent with the action of recent positive selection. In the case of galectin 8, four missense variants comprise two major haplotypes,

Molecular modeling predicts that the non reference haplotype produces large structural differences in galectin 8 protein conformations that were based on the reference sequence.

POB-056

Functional co-evolution in metabolic pathways

Mathias Bockwoldt[1], Toni I. Gossmann[3], Mathias Ziegler[2], Ines Heiland[1]

[1]UiT Arctic University of Norway (Norway), [2]University of Bergen (Norway), [3]University of Sheffield (United Kingdom)

In a pathway that is composed of several enzymes, the overall performance of the pathway is dependent on all enzymes, even though some enzymes might have more control than others. Especially flux distribution at branch points is dependent on the substrate affinity and turnover of all competing enzymes. Thus, we should be able to identify co-evolutionary features in enzymes of the same pathway, in particular for those that compete for the same substrate.

As an example, we analyzed the NAD-biosynthesis and consumption pathway - a central metabolic pathway determining the concentration of the major co-factor for redox reactions, NAD(P). Based on a dynamic pathway model, we indeed find that competing enzymes are likely to have altered evolutionary pressure on each other. Using sequence based analyses, we identify corresponding co-evolutionary pattern between several enzymes at branching points of the pathway and furthermore revealed evolutionary events in the development of the whole pathway that can be well explained by dependencies in the pathway dynamics.

POB-060

Forces shaping the distribution of rice genetic diversity

Rafal Marek Gutaker[1], Michael Purugganan[1, 2]

[1]New York University (United States), [2]New York University Abu Dhabi (United Arab Emirates)

Understanding the forces shaping the distribution of crops in the world is a key to adapting agriculture to new environments. In this project, we employ rice with its vast geographic area of cultivation and availability of large genomic dataset to investigate the forces behind the distribution of crop genetic diversity. Cultivated rice encompasses two genetically distinct subspecies; indica with long grains, and japonica with short and sticky grains. These subspecies spread in Asia independently, however, became sympatric through common cultivation. Our analyses suggest that populations of the two rice subspecies are distributed broadly following the isolation by distance scenario and are geographically structured. Both rice groups have consistent migration barriers, which can be attributed to geographic barriers (seas and mountains) and borders of climatic zones (subtropical and continental). Despite that, two rice subspecies exhibit different spatial patterns of diversity and distinct demographic histories. This reflects different dynamics of introduction in Asian regions moderated by anthropogenic factors.

POB-061

The impact of DNA repair status on germline mutation rate and spectra in mice

Mizuki Ohno[1], Noriko Takano[1], Kunihiko Sakumi[2], Teruhisa Tuzuki[3]

[1]Kyushu University (Japan), [2]Kyushu University (Japan), [3]Fukuoka Dental College (Japan)

The human de novo germline mutation (dGM) rate has been estimated to be around 1.20 $x10^{-8}$ /bp/generation. Surprisingly, most of mutations came from paternal genome, and the mutation frequency increase with the father's age. These data suggested that the dGM rate differs between individuals, and the factor influence dGM rate per generation is sequence-independent, sex-dependent, age-dependent.

We hypothesized that the cause of dGM would be endogenously and spontaneously generated in the germline cells. Not only DNA replication errors, but also endogenous DNA lesions are strong candidates for the source of spontaneous dGMs. Disruption of DNA repair pathways would be manifested the effects due to damaged DNAs by endogenous mutagenic factor. We thus attempted to detect and analyze dGMs experimentally by using the mice defecting DNA repair enzymes. First, we focused on the oxidative DNA damages and its repair pathways on the dGM rate and spectra. We found the 18-times higher dGM rate in the oxidative DNA repair-deficient mice compared with the wild-type mice. Moreover, we found decreased litter size, increased infant mortality, increased tumor incidence and increased abnormal phenotypes (e.g. hydrocephalus, coat color change, absent eye) in this mice line. These results indicated that spontaneously generated oxidative DNA damage was one of the cause of germline mutations, and the repair system was effectively working in the germline cell in the wildtype mice to reduce mutation rate. We will also report the results from the mismatch repair-deficient mice.

POB-063

How human chromosome 19 got its clusters of duplicated genes

Juan Felipe Ortiz[1], Antonis Rokas[1]

[1]Vanderbilt University (United States)

Clusters of duplicated genes (CTDGs) are great contributors to the diversity exhibited by many phenotypes, like snake venom, animal body plans, and olfaction. To systematically study CTDGs, we developed CTDGFinder, a homology-based algorithm that statistically takes into account the size and spacing of duplicate gene locations in a genome and has been shown to identify several well-known mammalian CTDGs (e.g., the Hox, globin, and protocadherin CTDGs). To gain insights into the genome-wide distribution of CTDGs, we comprehensively characterized the identity and distribution of CTDGs in the human genome. Remarkably, we found that more than 20% of human genes belong to CTDGs and that CTDGs span 7% of the total length of the human genome. Moreover, the distribution of CTDGs is not uniform across chromosomes, with chromosome 19 being the most significantly enriched in CTDGs. To investigate how the remarkable density of CTDGs in human chromosome 19 evolved, we compared human chromosome 19 to its orthologous chromosomes in other mammalian genomes. We found that chromosome 19 evolved from a chromosome fusion event prior to the split of the primate and rodent lineages. Interestingly, only one of the two chromosome 19 ancestors shows a history of conserved high CTDG density across mammals. Furthermore, genomic alignments of chromosome 19 orthologous regions show that its ancestral chromosomes experienced contraction, increasing chromosome 19 CTDG density. These results highlight the importance of CTDGs in chromosomal evolution, and open avenues of research towards a deeper understanding of the interactions between CTDGs and molecular evolutionary processes.

POB-064

RhesusBase: Understanding Human Biology in the Genomic Framework of Rhesus Macaque

Jiguang Peng[1], Shi-Jian Zhang[1], Chu-Jun Liu[1], Xiao-Ming Zhong[1], Qing Sunny Shen[1], Chen-Qu Wang[1], Xiu-Qin Zhang[1], Chuan-Yun Li[1]

[1]Peking University (China)

While rhesus macaque is a unique model for evolutionary and translational studies of human biology, several unresolved issues have limited the current use of this model ---- inadequate functional genomics annotations, error-prone genome assembly and gene structures, and lack of a platform for visualizing and assessing high-throughput data. RhesusBase was thus developed to de novo assemble the macaque genome, to refine genome-wide macaque gene structures, to integrate macaque functional annotations, and to provide a 'one-stop' knowledgebase for the primate research community. Briefly, the macaque genome was de novo assembled using a combinational approach with PacBio, BioNano and Illumina sequencing. A comprehensive population genetics profile in 31 macaque animals was identified on the basis of whole genome sequencing (*Molecular Biology and Evolution*, 2016). ~30% of the macaque transcripts were convincingly revised on the basis of PacBio Iso-seq and RNA-seq in multiple macaque tissues (*Nucleic Acids Research*, 2013; *Molecular Biology and Evolution*, 2017). 7.6 billion functional genomics annotations from 60 categories of public and in-house resources were integrated, and multiple interfaces were developed to update, retrieve, visualize, and download the meta-data (*Molecular Biology and Evolution*, 2014). Taken together, RhesusBase provides a big data-driven, information-rich framework that will broadly benefit genomics research in general and serves as an important resource for in-depth evolutionary studies of human biology.

POB-065

Loss of Conserved Protein Ubiquitylation Sites during Human Evolution

Dongbin Park[1], Chul Jun Goh[1], Ji Seok Lee[1], Yoonsoo Hahn[1]

[1]Chung-Ang University (Republic of Korea)

Ubiquitylation of lysine residues in proteins plays a pivotal role in the efficient removal of misfolded or unused proteins and in the control of various regulatory pathways by monitoring protein activity that may lead to protein degradation. The loss of ubiquitylated lysines may affect the ubiquitin-mediated regulatory network and result in the emergence of novel phenotypes. We analyzed mouse ubiquitylation data and orthologous proteins from 62 mammals to identify 194 conserved ubiquitylation sites from 170 proteins that were lost in the Euarchonta lineage leading to humans. Nine proteins, including BHMT, CNNM3, RRBP1, and SLC37A4, lost 1 conserved lysine residue, which is ubiquitylated in the mouse ortholog, after the human-chimpanzee divergence. Seventeen of the lost ubiquitylated lysines are also known to be modified by acetylation and/or succinylation in mice. In 8 cases, a new lysine evolved at positions flanking the lost conserved lysine residues, possibly as a compensation. We propose that loss of ubiquitylation sites during evolution may lead to the development of advantageous phenotypes which are then fixed by selection. The ancestral ubiquitylation sites identified in this study may be a useful resource for investigating the relationship between loss of ubiquitylation sites and the emergence of novel phenotypes during evolution towards the human form.

POB-066

Acceleration of Olfactory Receptor Gene Loss in Primate Evolution: Possible Link to Anatomical Change in Sensory Systems and Dietary Transition

Yoshihito Niimura[1, 2], Atsushi Matsui[1, 2], Kazushige Touhara[1, 2]

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

Primates have traditionally been regarded as vision-oriented animals with low olfactory ability, though this "microsmatic primates" view has been challenged recently. To clarify when and how degeneration of the olfactory system occurred and to specify the relevant factors during primate evolution, we here examined the olfactory receptor (OR) genes from 24 phylogenetically and ecologically diverse primate species. The results revealed that strepsirrhines with curved noses had functional OR gene repertoires that were nearly twice as large as those for haplorhines with simple noses. Neither activity pattern (nocturnal/diurnal) nor color vision system showed significant correlation with the number of functional OR genes while phylogeny and nose structure (haplorhine/strepsirrhine) are statistically controlled, but extent of folivory did. We traced the evolutionary fates of individual OR genes by identifying orthologous gene groups, demonstrating that the rates of OR gene losses were accelerated at the ancestral branch of haplorhines, which coincided with the acquisition of acute vision. The highest rate of OR gene loss was observed at the ancestral branch of leaf-eating colobines; this reduction is possibly linked with the dietary transition from frugivory to folivory because odor information is essential for fruit foraging but less so for leaf foraging. Intriguingly, we found accelerations of OR gene losses in an external branch to every hominoid species examined. These findings suggest that the current OR gene repertoire in each species has been shaped by a complex interplay of phylogeny, anatomy, and habitat; therefore, multiple factors may contribute to the olfactory degeneration in primates.

POB-067

Gain and loss of functions in the taste receptors of primates

HIROO IMAI[1], Emiko Nishi[1], Laurentia Purba[2], Nami Suzuki-Hashido[1], Kanthi Widayati[2], Takashi Hayakawa[1], Bambang Suryobroto[2]

[1]Kyoto University (Japan), [2]Bogor Agricultural University (Japan)

Most animals avoid bitter compounds and prefer sweet compounds. However, some primates ingest food items which are bitter to humans and/or not sweet to humans, suggesting a species-specific sense of taste. To reveal the mechanism of specific taste phenotypes, we conducted genetic and functional analysis of taste receptors in cultured cells along with behavioral analysis. We found species-specific gain and loss in sensitivities of taste receptors, some of which are supported by the behavioral test. For example, Japanese macaques are less sensitive to salicin, a bitter compound in the bark of willow tree, than are humans, due to the change in some amino acid residues that are situated in the putative ligand binding and intracellular regions of TAS2R16 (1,2). On the other hand, sweet taste receptor, TAS1R2/TAS1R3 showed increased sensitivity in Japanese macaques relative to humans for natural sweet compound, maltose (3). Colobines show very low sensitivity to Phenylthiocarbamide (PTC) due to the change in some amino acid changes in TAS2R38 (4). These differences in receptor sensitivities highlight the relevant tastes of compounds in the habitat of primates and contribute to their survival and adaptation.

(1) H. Imai et al., Biology Letters 8, 652-656 (2012)

(2) H. Imai et al., Biophysics and Physicobiology 13, 165-171 (2016)

(3) E. Nishi et al., Scientific Reports 6, 39352 (2016)

(4) L. H. P. Purba et al., Biology Letters 13, 20160834 (2017)

POB-068

Isoform evolution in primates through independent combination of alternative RNA processing events

Xuke Luan[1, 2, 3], Shi-Jian Zhang[1, 4], Chenqu Wang[1, 2, 3], Chuan-Yun Li[1]

[1]Peking University (China), [2]Peking University (China), [3]Peking University (China), [4]China Agricultural University (China)

Recent RNA-seq technology revealed thousands of splicing events that are under rapid evolution in primates, whereas the reliability of these events, as well as their combination on the isoform level, have not been adequately addressed due to its limited sequencing length. Here, we performed comparative transcriptome analyses in human and rhesus macaque cerebellum using single molecule long-read sequencing (Iso-seq) and matched RNA-seq. With Iso-seq data, we substantially expanded the repertoire of alternative RNA processing events in primates, and found that intron retention and alternative polyadenylation are surprisingly more prevalent in primates than previously estimated. We then investigated the combinatorial mode of these alternative events at the whole transcript level, and found that the combination of these events is largely independent along the transcript, leading to thousands of novel isoforms missed by current annotations. Notably, these novel isoforms are selectively constrained in general, and 1,119 isoforms have even higher expression than the previously annotated major isoforms in human, indicating that the complexity of the human transcriptome is still significantly underestimated. Comparative transcriptome analysis further revealed 502 genes encoding selectively constrained, lineage-specific isoforms in human but not in rhesus macaque, linking them to some lineage-specific functions. Overall, we propose that the independent combination of alternative RNA processing events has contributed to complex isoform evolution in primates, which provides a new foundation for the study of phenotypic difference among primates.

POB-069

Natural selection on the genetic and functional variations of bitter taste receptors (TAS2Rs) in wild chimpanzees

Takashi Hayakawa[1, 2], Yasuka Toda[3], Eiji Inoue[4], Hodaka Matsuo[1], Naruki Morimura[5], Miho Inoue-Murayama[5], Chie Hashimoto[1], Takumi Misaka[6], Hajime Ohigashi[7], Tetsuro Matsuzawa[1, 2, 8], Hiroo Imai[1]

[1]Kyoto University (Japan), [2]Japan Monkey Centre (Japan), [3]Meiji University (Japan), [4]Toho University (Japan), [5]Kyoto University (Japan), [6]The University of Tokyo (Japan), [7] Fukui Prefectural University (Japan), [8]Kyoto University (Japan)

Bitter taste receptors (TAS2Rs), one of the GPCR families, function as a detector of poisonous compounds in the oral cavity. Nucleotide sequences of 28 *TAS2R* genes in chimpanzee are evolutionarily diversified among subspecies, suggesting that each subspecies may genetically adapt to region-specific poisonous foods (Hayakawa et al. 2012 *PLOS ONE*). In this study, we focused on wild chimpanzees to elucidate ecological factors of such *TAS2R* diversification. We determined genetic diversity of two selected *TAS2R*s, *TAS2R38* and *TAS2R46*, in a wild West African chimpanzee (*P. t. verus*) population in Bossou (Guinea) and two wild East African chimpanzee (*P. t. schweinfurthii*) populations in Mahale (Tanzania) and Kalinzu (Uganda), using DNA from feces. As a result, nucleotide differentiation under the possible natural selection was observed in both *TAS2R38* and *TAS2R46* of the chimpanzee populations. Cellular assay confirmed that chimpanzee TAS2R46 is a functional receptor, using candidate ligands to which human TAS2R46 responds; therefore, genetic diversity of chimpanzee *TAS2R46* may affect bitter plant ingestion in wild chimpanzees. We also collected plant food items of chimpanzees that may activate chimpanzee TAS2Rs in each habitat, based on the ecological observations. *Vernonia* plants are candidate plants involved in the natural selection of *TAS2R*s. Chimpanzees in Mahale ingest *Vernonia* plants, but chimpanzees in Bossou and Kalinzu have not been reported to ingest them. These results indicated that genetic and functional diversity of *TAS2R*s and region-specific food choices of bitter plants could be organically linked with each other.

POB-070

Mitochondrial and exome diversity in Pan troglodytes schweinfurthii at Gombe National Park

Andrew T Ozga[1], Timothy H Webster[2], Maria A Nieves-Colon[3, 4], Kathleen Fowler[2], Rebecca Siford[4], Melissa Wilson Sayres[1, 2], Rebecca Nockerts[5], Michael L Wilson[5, 6], Ian Gilby C[4, 7], Anne E Pusey[8], YingYing Li[9], Beatrice H Hahn[9], Anne C Stone[1, 4, 7]

[1]Arizona State University (United States), [2]Arizona State University (United States), [3]Laboratorio Nacional de Genomica para la Biodiversidad (Mexico), [4]Arizona State University (United States), [5]University of Minnesota (United States), [6]University of Minnesota (United States), [7]Arizona State University (United States), [8]Duke University (United States), [9]University of Pennsylvania (United States)

One of the goals of primate genetics is to investigate levels of genetic diversity in wild populations over time. Strides have been made in understanding diversity in extant wild chimpanzee populations in African reserves,but long-term diversity (over a several decade span) had not yet fully been examined. Here, we present whole mitochondrial genome and exome data from living and dead eastern chimpanzees (Pan troglodytes schweinfurthii) from Gombe National Park in Tanzania. We use both modern and ancient DNA extraction and capture methods to chronicle the temporal changes in haplotype and genetic diversity in this small, protected habitat. We recovered 33 full mitochondrial genomes from chimpanzees that occupied the park from the 1960's to the 2000's, identified moderate diversity compared to other populations, and found a slight decrease in genetic diversity across decades (0.00257 to 0.00155). We identified 14 unique haplotypes at Gombe which group into six clusters when incorporated into a network analysis with 40 other previously published P. t. schweinfurthii genomes. Additionally, we use full exome data (9-90% coverage using IDT and Arbor Biosciences captures) sequenced from urine, dentine, dental calculus, and feces from living chimpanzees to estimate levels of heterozygosity compared to other wild Pan populations and assess the future trajectory of genetic diversity within Gombe. These results emphasize the importance of long term study of protected primate populations like those from Gombe National Park in order to ensure that inbreeding does not decrease genetic diversity to levels detrimental to these already endangered animals.

POB-071

Phenotypic variation in free-ranging rhesus macaques: heritability and selection

James P Higham[1], Clare Kimock[1], Alex DeCasien[1], Constance Dubuc[1]

[1]New York University (United States)

Evolutionary selection acts on heritable variation. Understanding the relative importance of genetic versus environmental determinants of traits, and the relationship between trait variation and reproductive success, is important for assessing their evolutionary landscape. However, the data required to measure this are extremely hard to come by for wild populations of large-bodied long-lived mammals. Here, we investigate the genetic and environmental sources of trait variation, and the links between trait variation and reproductive fitness, in free-ranging rhesus macaques on Cayo Santiago, Puerto Rico. We combine cranial, dental, and postcranial skeletal measurements of 223 individuals, and measurements of the facial coloration of 266 individuals, with a full genetic pedigree, and data on reproductive success, going back to 1985. Models of narrow-sense heritability estimates indicate that many, but not all, skeletal traits were significantly heritable, with greater additive genetic variance for measures of body size compared to bone proxies of muscle strength. Several skeletal traits, including canine length, showed lower heritability and greater variability in males than in females, suggesting a greater role for developmental determination in males. Coloration is heritable in both sexes, but with greater heritability estimates in analyses only including the same-sex parent (sex-linked inheritance). Selection gradients show that coloration is under linear selection in females, and correlational selection in males, with the most successful males combining dark face color with high dominance rank. We discuss these results in light of our numerous behavioral and experimental results, which together elucidate the evolutionary selection pressures acting on rhesus macaque biology.

POB-072

Interpreting functional genomic data from the Cayo Santiago rhesus macaques

Michael J Montague[1], Noah Snyder-Mackler[2], Lauren JN Brent[3], Julie E Horvath[4, 5, 6], Michael L Platt[1]

[1]University of Pennsylvania (United States), [2]University of Washington (United States), [3]University of Exeter (United Kingdom), [4]North Carolina Central University (United States), [5]North Carolina Museum of Natural Sciences (United States), [6]Duke University (United States)

Evidence suggests that individual variation in social behavior arises from a combination of genetic predispositions and individual experience, yet the underlying biological mechanisms remain poorly understood. To address this gap, we are investigating the genetic and epigenetic contributions to social behavior in a large, free-ranging population of rhesus macaques (*Macaca mulatta*) with a known pedigree and detailed behavioral phenotypes. We hypothesize that genetic variants underlying molecular differences in neural circuits are associated with behavioral variation in this socially complex species. Toward this end, over the past seven years, we have compiled genetic data and biological samples from approximately 1000 animals, while also amassing extensive observational behavioral data. After conducting whole genome sequencing for 250 individuals, we performed additional tissue sampling to biobank brain tissue and twenty peripheral tissues and organs sampled from a subset of animals, as well as microbiota from skin, lung and digestive tract. Whole blood and flash frozen brain tissues were recently processed for gene expression (RNAseq) and epigenetic (ATACseq) analyses. The brain bank is unprecedented in the number, diversity of ages, and degree of background behavioral information. This multi-faceted approach will generate valuable insights on the influences of social environment, including not only social stressors, but also social mechanisms that may be protective against stress. Here, we illustrate how this integrative dataset can be used to identify the functions and pathways underlying the relationship between social environment and the transcriptional and epigenomic signatures in different brain regions.

POB-073

Adaptation of owl monkeys to nocturnal lifestyle driven by rapid expansion of simple repeat sequence to form megasatellite DNA

Akihiko Koga[1], Hidenori Nishihara[2], Roscoe Stanyon[3], Hirohisa Hirai[1]

[1]Kyoto University (Japan), [2]Tokyo Institute of Technology (Japan), [3]University of Florence (Italy)

Rod cells of many nocturnal mammals possess a "non-standard" nuclear architecture. Heterochromatin localizes to the central region of the nucleus, providing elevated night vision by passing incoming light efficiently to the outer segments of photoreceptors. Owl monkeys (genus Aotus) are unique among New World monkeys (parvorder Platyrrhini) in that their ancestor underwent a shift from diurnal to nocturnal lifestyle. In their rod cell nucleus, a spherical heterochromatin block is found although it is not complete as compared with that of typical nocturnal mammals. Our experimental analysis demonstrated that the primary DNA component of this heterochromatin block is OwlRep, a megasatellite DNA consisting of 187-bp-long repeat units. Further, our bioinformatic analysis revealed its evolutionary origin and formation process. HSAT6 is a small-scale repetitive DNA that occurs widely in primates. It is likely to have been a single-copy sequence in early stages of primate radiation. HSAT6, with or without flanking sequences, was segmentally duplicated in New World monkeys or their ancestor. Then, in the owl monkey linage after its divergence from other New World monkeys, a copy of HSAT6 was tandemly amplified at chromosome ends, eventually forming OwlRep. It was also suggested that the formation of this megasatellite DNA started at or after 20 Mya and ended at or before 5 Mya. OwlRep thus provides a unique example of a satellite DNA that has a clear biological function directly related to organismal adaptation.

POB-074

Bitter taste receptor function in lemurs provides insight into the evolution of beta-glycoside sensing mechanism in primates

Akihiro Itoigawa[1], Takashi Hayakawa[1, 2], Nami Suzuki-Hashido[1], Hiroo Imai[1]

[1]Kyoto University (Japan), [2]Japan Monkey Centre (Japan)

Bitter perception is essential for survival because it is related to the detection of potential harmful substances. Since there are many kinds of toxic compounds in nature, animals have bitter perception adapting to their specific feeding habitat.

In this study, we focused on bitter taste receptors (TAS2Rs) detecting β-glycosides, which are kinds of secondary metabolites in plants. It is known that TAS2R16 in haplorrhines, one of the two monophyletic groups of primates, including humans, respond to β-glycosides, whereas TAS2R16 ortholog in mice does not respond. In contrast, human TAS2R41 does not respond to them, but mouse TAS2R41 ortholog responds. However, it is not clear whether the common ancestor of primates obtained β-glycosides sensitivity in TAS2R16 and/or TAS2R41 because β-glycoside sensitivity has been unknown in strepsirrhines, the other monophyletic group in primates.

Thus, we analyzed β-glycosides sensitivity of TAS2R16 and TAS2R41 in three lemur species (classified as strepsirrhines), using a calcium imaging assay. Salicin, a kind of β-glycosides, was used as a ligand for this analysis. As a result, we found that TAS2R16 responded to salicin and that TAS2R41 did not respond in all analyzed lemurs like haplorrhines. We also conducted food choice test using apple pieces for estimating the perception of salicin in a black lemur. Salicin-soaked pieces were significantly rejected, indicating that TAS2R16 function reflect its feeding behavior. In conclusion, β-glycosides sensitivity of TAS2R16 is conserved among all primate lineages and presumably obtained in the common ancestor of primates.

POB-075

Acquisition of Human-specific Characteristics of Skin through Gene Expression Changes

Nami Arakawa[1], Yohey Terai[1], Hiroo Imai[2], Yoko Satta[1]

[1]SOKENDAI (The Graduate University for Advanced Studies) (Japan), [2]Kyoto University (Japan)

Humans have unique skin characteristics compared to other primates. For example, the qualitative studies reported the thicker epidermis/dermis and the richer elastic fibers in humans, as well as the wavy shape of epidermal basement membrane (BM) to strengthen the connection between epidermis and dermis. It has been proposed that these human-specific characteristics protect the internal tissues from external physical stresses and compensate the loss of hair. The aim of this study is to understand how human-specific characteristics of skin had been genetically acquired.

Based on RNA-seq analyses of skin samples, the expression levels of 11 genes were significantly higher in humans (n=5) than in apes (chimpanzees: n=3, gorillas: n=3, orangutans: n=3). Those included the genes associated with the components of epidermal BM and the genes regulating the tension of skin by aggregating collagens. Their increased gene expression may contribute to the human-specific structure of epidermal BM and amount of elastic fibers in skin. Among conserved regions around these genes in apes, the putative transcriptional regulatory regions were inferred according to histone modifications showing active regulatory regions. Human-specific substitutions within the inferred regions were estimated to be responsible for the human-specific expression patterns of these genes, resulting in two to ten human-specific substitutions for each gene. All of them were located in binding sites of transcription factors, suggesting the possibility of expression changes by influencing affinity for transcription factors.

These human-specific substitutions may make human skin unique by changing the expression of the genes associated with skin traits.

POB-077

Gene Expression Plasticity At RNA And Protein Levels

Yifan Dai[2], Xinzhu Wei[1], Jianzhi Zhang[1]

[1]University of Michigan (United States), [2]Fudan University (China)

Gene expression plasticity across environments may affect adaptations to new environments. However, most studies on gene expression plasticity use only transcriptome data without analyzing proteome data despite that proteomic changes are likely functionally more important than transcriptomic changes. Here we study gene expression plasticity using the genome, liver transcriptome, and liver proteome data from 192 Diversity Outbred mice each fed with either a standard diet or a high fat diet. We found that about two thirds of all genes analyzed are more different in RNA concentration than protein concentration between the two diet groups and between the two sexes ($p < 10\text{-}8$). We are examining biological replicates to exclude the possibility that our observation is due to a potentially higher measurement error for RNA concentration than protein concentration. To understand the genetic basis of this phenomenon, we mapped cis- and trans-quantitative genetic loci for RNA concentration (eQTLs) and protein concentration (pQTLs). We plan to test (i) whether eQTLs tend to have smaller effects on protein concentrations than on RNA concentrations and (ii) whether eQTLs tend to be compensated by pQTLs that affect translation and post-translational processes. These analyses are expected to deepen our understanding of patterns as well as underlying genetic basis of gene expression plasticity.

POB-079

Locating adaptive events in the evolutionary history of the human coding genome

Ravi Patel[1, 2], Sudhir Kumar[1, 2, 3], Maxwell Sanderford[1], Tamera Lanham[1], Koichiro Tamura[4], Alexander Platt[1, 2], Benjamin Glicksberg[5], Ke Xu[5], Joel Dudley[5], Laura Scheinfeldt[1, 2, 6]

[1]Temple University (United States), [2]Temple University (United States), [3]King Abdulaziz University (Saudi Arabia), [4]Tokyo Metropolitan University (Japan), [5]Icahn School of Medicine at Mount Sinai (United States), [6]Coriell Institute (United States)

Of the hundreds of thousands of missense mutations detected in the human genome, only a handful have been determined to be adaptive. This suggests that adaptive events are exceedingly rare in human evolution. A vast majority of these adaptive variants have been identified using methods that are limited to very recent history, relying on signals that are slowly swept away with the passage of time. Here, we utilize an Evolutionary Probability Approach (EPA), which detects positive selection by contrasting neutral expectations generated from interspecific evolutionary history and intraspecific segregation patterns. The EPA approach discovers candidate adaptive polymorphisms (CAPs), and identifies time scales during which adaptive events may have occurred in the evolutionary history of a species. We report more than 18,000 missense CAPs currently segregating in human populations, and an additional 33,000 fixed alleles that likely experienced positive selection at various time points along the human lineage since the common ancestor with the Pan genus. Along with functional analyses that show >2,000 CAPs are associated with contemporary phenotypes, we use demographic simulations of human evolution to show that these CAPs likely contain a 15-fold increase in adaptive changes in human history than previously reported.

POB-085

Identifying Lineage-specific Targets of Darwin Selection by Bayesian Analysis of Genomic Polymorphism and Divergence from Multiple Species

Shilei Zhao[1], Tao Zhang[2, 3], Bing Su[2, 3], Peng Shi[2, 3], Hua Chen[1, 3]

[1]Beijing Institute of Genomics, Chinese Academy of Sciences (China), [2]Kunming Institute of Zoology, Chinese Academy of Sciences (China), [3]University of Chinese Academy of Sciences (China)

We present a method that jointly analyzes polymorphism and divergence sites from genomic sequences of multiple species to identify genes under positive or negative selection, and pinpoints its occurrence time to a specific lineage of the species phylogeny. The method provides posterior distributions of fitness effect of each gene, and parameters concerning the evolutionary history, including species divergence times and effective population sizes of external species. We demonstrate with simulation that our method provides accurate estimates of these population genetic parameters. We apply the method to genomic sequences of human, chimpanzee, gorilla and orangutan, and construct a spatial and temporal map of natural selection occurred during the evolution history of the four Hominidae species. In addition to FOXP2 and other known genes, we identify a new list of lineage-specific targets of Darwin selection. The positively selected genes in the human lineage are enriched in pathways of gene expression regulation, immune system, metabolism etc. Interestingly, some pathways, such as, gene expression, are significantly enriched with positively selected genes, while metabolism is enriched with both positive and negative selected genes. Our analysis provides insights into Darwin evolution in coding regions of human and great apes, serving as a basis for further molecular and functional study.

POB-086

Different types of cell migration during tumor growing process lead to spatial patterns of genetic variation

Yongsen Ruan Ruan[1], Ao Lan[1], Chung-I Wu[1, 2, 3]

[1] Sun Yat-sen University, Guangzhou (China), [2]Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing (China), [3]University of Chicago, Chicago (United States)

Tumorigenesis is an evolutionary process and populations of cancerous cells evolve, much like populations of organisms do. So cell migration during tumor growing process has a substantial effect on spatial genetic structure, but it remains unclear. Here, we analyze the genotyping data from a hepatocellular carcinoma with 286 multiregional samples and divide the tumor into 7 cell clones according to somatic mutations. We find that there is not distinct separation between any two adjacent clones and the high mixing scores show that the cells from two adjacent clones mix well with each other. Furthermore, we detect that 2 of the 7 clones have the clone variegation phenomena. All these imply that cell migration exists and affects the spatial genetic structure during tumor growth. In a spatially structured population, we divide cell migration into two types. One is called division dependent migration (D-migration). Migration only occurs in cell division process when offspring cells push neighboring cells to move. The other type is cell active migration (A-migration). A motile cell can independently migrate to a vacant position at any time. Computer simulations show that both types of migration can lead to clone variegation and high mixture between clones. However, clone patchiness is more pronounced in A-migration model and A-migration can lead to establishment of pioneer population in advance of the main expansion front and result in a quick fixation of a cell clone.

POB-087

Genomic Reconstruction of Transmission Networks in Malaria

Seth Redmond[1, 2], Bronwyn Maccinnis[1, 2], Selina Bopp[2, 1], Amy Bei[2], Daouda Ndiaye[3], Daniel Hartl[4, 1], Dyann Wirth[2, 1], Sarah Volkmann[2, 1, 5], Daniel Neafsey[2, 1]

[1]Broad Institute (United States), [2]Harvard TH Chan School of Public Health (United States), [3]Cheikh Anta Diop University, Dakar (Senegal), [4]Harvard University (United States), [5]Simmons College (United States)

Sequencing de novo mutations is a key technique in genomic epidemiology and has enabled researchers to reconstruct networks of disease transmission in many viral and bacterial pathogens. However it is not known whether eukaryotic pathogens acquire sufficient mutations within the host for this approach to be viable. The malaria parasite Plasmodium falciparum has a large genome, slow mutation rate, and an obligately sexual life cycle, making de novo mutations difficult to accurately identify.

Nevertheless the combination of large intra-host population sizes and severe transmission bottlenecks ensure that some de novo mutations are fixed between hosts. Using optimized library preparation and genotyping techniques we have gained access to a further 14% of the genome over previous methods, enabling us to target rapidly-mutating regions of the genome and giving us unprecedented access to de novo mutations. These techniques have been applied to a set of clinical samples - taken from a single location over a period of 7 years - and have enabled transmission networks to be reconstructed for malaria for the first time.

We will show how these methods are now being applied to Malaria outbreaks in Senegal and Haiti in order to aid in optimal targeting of disease interventions, and how they can generate unique insights into the transmission dynamics of recombining pathogens.

POB-090

Evolutionary analysis of epitopes and low complexity regions in Plasmodium

Sarah Medley[1], Alyssa Beaudet[1], Helen Piontkivska[2], Fabia Ursula Battistuzzi[1]

[1]Oakland University (United States), [2]Kent State University (United States)

The molecular sequence structure that characterizes epitopes contains similarities to low complexity regions (LCRs). It is therefore possible that LCRs have similar immunogenic functions to epitopes. We investigate this scenario in the agent of malaria (*Plasmodium*) in an effort to identify new or improved antimalarial drug targets. To examine their functional relationship, we have utilized an evolutionary approach comparing location and sequence conservation levels. Our goal is to identify epitopes and LCRs with high conservation and location overlap that can be candidates for future experimental studies. Based on pairwise species comparisons of 15 genes in 18 species, we found approximately 60% of LCRs and 50% epitopes with conserved amino acids. We also found nearly 40% of epitopes overlapping with LCRs and, of these regions, over 30% were conserved in both the epitope and LCR. Epitopes and LCRs especially were found to have a high rate of deletions. When pairwise deleted sites are ignored, LCRs exhibit lower conservation (<5%) compared to epitopes (18%). This is expected given the high evolutionary rate of LCRs. However, we also found different levels of conservation in species belonging to different phylogenetic subgroups (Laverania and Vinckei more conserved than Plasmodium and Haememoeba) and with different host preferences (Plasmodium in rodent and human hosts more conserved than those in primate and avian hosts). These results suggest that LCRs may be subject to different selective forces based on their evolutionary history and lifestyle. Genome specific analyses may lead to identification of new genomic regions with immunogenic properties.

POB-092

Recurrent loss of functional repressors in the evolution of *Escherichia coli in the mice's gut*

Mohamed GHALAYINI[1], *Sara DION*[1], *Melanie MAGNAN*[1], *Erick DENAMUR*[1], *Mathilde LESCAT*[1], *Olivier TENAILLON*[1]

[1]*INSERM (France)*

Escherichia coli is a highly studied pathogen, yet little is known about its long-term evolution within its host as a gut commensal. We therefore established two long-term *E. coli* evolution experiments in mice's gut over a full year. Our goal was to decipher how *E. coli* adapts to the gut and to the diet of its host. We first used streptomycin treated mice to follow the fate of the naturally streptomycin resistant *E. coli* strain 536. Whole genome sequencing revealed convergence at the gene level among the replicates. In particular, the functional repressor of a metabolic operon was inactivated. As other signs of convergence implied genes associated to streptomycin resistance, we launched an experiment mimicking the transmission of microbiota from mother to child in which the use of streptomycin was restricted to the inoculation of the strain in the pregnant mother. We observe a convergence in the inactivation of the *lacI* repressor but no clear sign of diet adaptation presumably due to low *E. coli* density. Interestingly, in both experiments, selection favored the inactivation of functional repressors. Hence, despite a functional induction, inactivation of the repressors conveyed a selective benefit in the gut or *in vitro*. These observations challenge the benefits and stability of operon regulation.

POB-093

Simulation of Intratumor Heterogeneity and its Medical Implication

Watal M. Iwasaki[1], Hideki Innan[1]

[1]SOKENDAI (Japan)

Cancers consist of heterogeneous subclones rather than a single type of homogeneous clonal cells.
This phenomenon, intratumor heterogeneity (ITH), has been a major obstacle to cancer screening and treatment.
Thus, understanding how tumors evolve and accumulate mutations is essential for early detection and treatment decisions.

We here developed a comprehensive and flexible framework, *tumopp*, to simulate spatio-temporal development of various solid tumors, and found that different assumptions can produce a huge variety of morphology and intratumor heterogeneity patterns even under neutrality.
As an application, *tumopp* was used to search for the optimal biopsy strategy to capture the entire genetic diversity within a tumor conditional on the status of the patient and tumor.
The results suggest that F_{ST} and physical distance between biopsy samples are informative to determine the minimum required number of samples.

POB-096

Rapid adaptations to the accidental human host in Legionella pneumophila

Daniel Dennis Leenheer[1, 2], Carmen Pelaz[3], Matilda Morin[4], Elisabeth Hallin[4], Daniella Klingenberg[4], Sophie Jarraud[5, 6], Christoph Ginevra[5, 6], Lionel Guy[1]

[1]Uppsala University (Sweden), [2]University of Tsukuba (Japan), [3]Centro Nacional De Microbiologia Instituto De Salud Carlos III (Spain), [4]Public Health Agency Sweden (Sweden), [5]Hospices Civils De Lyon (France), [6]Universite Lyon (France)

Legionella are host-adapted bacteria that infect, and primarily reproduce in, aquatic amoebae. Using similar infection mechanisms, they can infect human macrophages, and cause Legionnaire's disease and Pontiac fever. We hypothesise that, despite these similarities, the hosts are different enough so that there exists high selective value mutations that dramatically increase Legionella's fitness in the human host. As human-to-human transmission is very rare, fixation of these mutations into the population is unlikely, and mutations in the same genes would be observed in independent human infections, as examples of convergent evolution. Identifying these adaptive mutations would shed light on the mechanism of *Legionella* infection in its human host. By comparing a large enough number of independent infections, we expect these highly adaptive mutations to appear several times, despite the short duration of the infection.

Clinical isolates and isolates identified as their respective environmental source were sequenced and de-novo assembled using SPAdes. Variants (indels and SNPs) were called using RedDog. Variants between samples were compared to identify genes that are likely candidates for human-specific adaptations in multiple samples. Isolates from 58 independent infections or outbreaks were obtained and sequenced (161 isolates in total, from Sweden, France and Spain), and complemented by isolates from published studies. A preliminary analysis between clinical samples and their respective environmental source showed several genes were mutated independently multiple times.

Sequencing and comparing large numbers of *Legionella* clinical and environmental isolates enable us to identify adaptations that allow *Legionella* to infect the human host.

POB-097

Comparative genomics of *Mycobacterium tuberculosis*samples from patient serial isolates reveals antigenic variation during active TB disease

Roger Vargas[1, 2], Luca Freschi[2], Francis Drobniewski[3], Inaki Comas[4], Megan Murray[5], Maha Farhat[2, 6]

[1]Harvard Medical School (United States), [2]Harvard Medical School (United States), [3]Imperial College London (United Kingdom), [4]Instituto de Biomedicina De Valencia CSIC (Spain), [5]Harvard Medical School (United States), [6]Massachusetts General Hospital (United States)

Within-host evolution of *Mycobacterium tuberculosis* (Mtb) during active disease in humans is incompletely understood. While advancements have been made over the past decade in understanding the acquisition of antibiotic resistance *in vivo* with the aid of whole genome sequencing, debate surrounds the role antigenic variation during infection and establishment of Mtb disease. Several studies have concluded that CD4 T-cell epitopes are conserved in Mtb with respect to the variation observed in essential Mtb genes. Yet, other studies have found computational and experimental evidence that certain Mtb proteins evoke strong immune responses in humans and that the genes encoding them undergo diversifying selection. Particularly antigenic variation during infection within a host rather than at a population level has not been evaluated.

To study Mtb within-host evolution, we make use of paired serial isolates from the sputum of patients that were actively infected with Mtb. Our sample consists of 148 isolates from 74 patients representing the UK, Peru and China with the time between isolate collection for each patient ranging from two months to a few years.

We analyze the WGS for each pair of isolates and compare the intra-host diversity from each patient to the others. We find several coding regions of the Mtb genome that appear to be undergoing diversifying selection *in vivo*. Strikingly, many of these regions encode experimentally validated and computationally predicted CD4 T-cell epitopes. Some of our hits have previously been implicated as antigenic targets while others represent novel sources of variation.

POB-098

Search for the factors related to HCV replication in the HuH-7 cell line lineages

Masaki Kawamoto[1], Toshinori Endo[1], Masayoshi Fukasawa[2], Kentaro Hanada[2], Naoki Osada[1]

[1]Hokkaido University (Japan), [2]National Institute of Infectious Diseases (Japan)

Elucidating the evolutionary trajectory of artificially selected cell lines would provide interesting insight into the mechanisms of adaptation at the cellular/molecular level. The HuH-7 cell line, derived from a liver tumor, and its derivative Huh7.5.1-8 cell line have been revealed being capable of propagating HCVs more efficiently. In this research, in order to explore candidate genes that enhance or inhibit infection or replication of HCV in Huh7.5.1-8 cell lines, we mapped and aligned short reads of HuH-7 and Huh7.5.1-8 to the reference human genome and identified mutations, i.e. single nucleotide variants, copy number variants, insertions, deletions, duplications, inversions, and translocations, in each cell lines. Next, we compared the condition of these variants with the change of expression level data from Huh7.5.1 to Huh7.5.1-8 which obtained from the previous study. As a result of whole genome sequencing, 298 307 mutations were shared between HuH-7 Huh7.5.1-8. However, some mutations were specific to HuH-7 or Huh7.5.1-8 (75 880, 103 090, respectively). Among the genes harboring any mutations, with the change of expression level, 6939 expression level of genes changed. Among these genes, 2200 and 2227 genes were upregulated and downregulated in Huh7.5.1-8, respectively. In addition to these up- or downregulated genes, 1104 genes were expressed specifically in Huh7.5.1-8, and 1408 genes were expressed specifically in Huh7.5.1. We will detect causal genes by even more detail research about these fold change genes.

POB-099

Naturally occurring West Nile virus infections display no evidence for heightened selective constraint in avian hosts over mosquito vectors

Chase W. Nelson[1], Tony L. Goldberg[2, 3], Sergios-Orestis Kolokotronis[1, 4], Gabriel Hamer L.[5], Christina M. Newman[2], Samuel D. Sibley[2], Tavis K. Anderson[2], Edward D. Walker[6], Marilyn O. Ruiz[7], Jeffrey D. Brawn[8], Uriel D. Kitron[9]

[1]American Museum of Natural History (United States), [2]University of Wisconsin-Madison (United States), [3]University of Wisconsin-Madison (United States), [4]SUNY Downstate Medical Center (United States), [5]Texas A&M University (United States), [6]Michigan State University (United States), [7]University of Illinois (United States), [8]University of Illinois (United States), [9]Emory University (United States)

Although most RNA viruses are characterized by high genetic diversity, arboviruses such as West Nile virus (WNV) exhibit relatively low diversity but retain the tendency to "emerge". One explanation for this low diversity is the "trade-off" hypothesis, which posits differential selection in avian hosts and mosquito vectors. To test this hypothesis in the wild, we deep-sequenced spatio-temporally matched bird (n=6) and mosquito (n=14) samples retained from a long-term WNV screening study in Chicago, USA to characterize intrahost diversity. Single nucleotide variants were present at 7.5% of the 11kb genome, with bimodal F_{ST} values. However, no sites were fixed for unique variants in hosts or vectors, and phylogenetic analyses revealed bird/mosquito nonmonophyly. Intrahost nucleotide diversity showed no significant differences between samples from birds (π_N/π_S=0.23; π_S=0.0014) and mosquitoes (π_N/π_S=0.20; π_S=0.0005), and between-sample diversities for bird/mosquito comparisons (d_N/d_S=0.04; d_S=0.026) were indistinguishable from bird/bird and mosquito/mosquito comparisons. We preliminarily conclude that, in naturally occurring WNV infections, (1) interhost viral divergence exceeds intrahost viral diversity; (2) concurrently, interhost viral conservation exceeds intrahost viral constraint; (3) viral populations diverge substantially between host individuals; but (4) interhost divergence is not primarily due to bird vs. mosquito differences. Our findings suggest that the WNV genome is pre-adapted to both hosts and vectors, and that viral populations exhibit similar intrahost diversity in each. These results may help to explain how WNV and/or other arboviruses can maintain their emergent potential with limited genetic diversity.

POB-100

The effect of HIV co-infection in the evolution Mycobacterium tuberculosis

Daniela Brites[1, 2], Chloe Loiseau[1, 2], Eddie Wampande [3], Jan Hattedorf[1, 2], Henry Boom[4], Moses Joloba[4], Sebastien Gagneux[1, 2]

[1]Swiss Tropical Public Health Institute (Switzerland), [2]University of Basel (Switzerland), [3]Makerere Universtiy (Uganda), [4]Case Western Reserve University (United States)

The incidence of human tuberculosis (TB) has increased in many parts of the world, partially fueled by HIV co-infections. Adequate T-cell responses, in particular CD4+ T-Cells which are preferentially infected by the HIV virus, are essential for providing protective immunity against M. tuberculosis (Mtb) infection, but act as well as drivers of lung pathology and mediators of transmission. For most of its evolutionary history Mtb has evolved in the absence of HIV co-infection. We hypothesize thus that the immunocompromised human host environment affects Mtb fitness both by altering the immune landscape within host and by altering transmission dynamics. With the aim of investigating if and how HIV co-infection affects Mtb evolution, we apply population genomics to whole-genome polymorphism data obtained both from approximately 80 Mtb isolates evolving within patients with and without HIV co-infection as well as from approximately 500 Mtb strains isolated from HIV infected and uninfected TB patients from different geographical and epidemiological settings. Our results suggest that HIV co-infections do not structure Mtb populations significantly, despite causing dramatic changes in the clinical presentation of TB. However the level of immunosuppression of the host seems to affect the fate of mutations emerging in Mtb epitopes recognized by CD4+ T-cells. We explore different scenarios of Mtb evolution within and between immunocompromised patients which could possibly lead to the observed patterns in the Mtb population.

POB-101

Using genotype abundance to improve phylogenetic inference

William S. DeWitt[1, 2], Luka Mesin[4], Gabriel D. Victora[4], Vladimir N. Minin[3], Frederick A. Matsen[2]

[1]University of Washington (United States), [2]Fred Hutchinson Cancer Research Center (United States), [3]University of California Irvine (United States), [4]The Rockefeller University (United States)

Modern biological techniques enable very dense genetic sampling of unfolding evolutionary histories, and thus frequently sample some genotypes multiple times. This motivates strategies to incorporate genotype abundance information in phylogenetic inference. In this study, we synthesize a stochastic process model with standard sequence-based phylogenetic optimality, and show that tree estimation is substantially improved by doing so. Our method is validated with extensive simulations and an experimental single-cell lineage tracing study of germinal center B cell receptor (BCR) affinity maturation. We first validate the procedure against simulations of germinal center BCR diversification. We also empirically validate our method using an experimental lineage tracing approach combining multiphoton microscopy and single cell BCR sequencing, allowing us to study individual germinal center B cell lineages from brainbow mice. We have implemented our computational method, termed *GCtree*, as open source software.

POB-102

Rapid adaptation of bacteriophage to their host after host-switching

Xuhua Xia[1]

[1]University of Ottawa (Canada)

Host-switching in some bacteriophage is readily identifiable, such as Phage PRD1 parasitizing the gram-negative Escherichia coli but having all its phylogenetic relatives parasitizing gram-positive bacteria. Such host switching events provide an opportunity to study phage adaptation to a new host because one knows approximately the ancestral states. For example, the original and the new host have different 3' end of 16S rRNA (3'TAIL) and require different Shine-Dalgarno (SD) sequence for efficient localization of the start codon. The new host may have different tRNA pool from the original host and require new codon adaptation for efficient translation elongation. The new host may also have different abundance of release factors (RF1 and RF2) decoding stop codons which would drive phage genes to have different stop codon usage. We used RNA-Seq data to characterize the 3'TAIL of 16S rRNA (which tends out to be heterogeneous in most cases due to different RNA processing and degradation) and tRNA abundance, and compile RF1 and RF2 abundance from literature in the original and new hosts, to study how host-switched phage adapt to the new environment by altering their SD sequence, codon and stop codon usage.

POB-103

Characterising local epidemiology of *P. falciparum* through the structure of mixed infections

Sha Joe Zhu[1], Jason Hendry[1], Jacob Garcia[1], Gil McVean[1]

[1]University of Oxford (United Kingdom)

Individuals infected with *Plasmodium falciparum* malaria pathogens can carry multiple distinct, though sometimes related, strains. However, how the rate and relatedness of such mixed infections relates to local epidemiological parameters remains unclear.

Our program DEploid can infer the structure and identity of strains present in a sample by combining a reference panel of known haplotypes, and report the number of strains, their proportions, relatedness profiles and individual haplotypes.

We apply our method to 2,512 isolates from 14 countries. Our result show geographical differentiation and haplotype structure vary in the parasite. In particular, we find that 679 out of 1010 cases from Asia, and 622 out of 1502 cases from Africa present evidence of infection by a single parasite strain. Rates of mixed infection vary from 18%-63% across countries and correlate with independent estimates of prevalence (Pearson r = 0.76, p = 7.5e-06). 54% of mixed infections involve more than two strains and 71% include sibling strains likely to have to be have been co-transmitted from a single mosquito.

Relatedness among strains decreases with prevalence ($r = -0.75,~p = 1.4e-05$) and differs between continents, with Asian isolates typically showing elevated relatedness compared to African isolates. We show how characteristics of mixed infection relate to local epidemiological process and propose that monitoring spatial and temporal patterns of mixed infection will be highly informative in pathogen surveillance.

POB-104

Genome-wide variation in the Andes, Amazonia and Pacific coast of western South America

Chiara Barbieri[1, 2], Rodrigo Barquera[3], Leonardo Arias[4], Jose Raul Sandoval[5], Oscar Acosta[5], Ricardo Fujita[5], Camilo Zurita[6], Kentaro Shimizu[2], Paul Heggarty[1], Russell Gray[1], Mark Stoneking[4], Irina Pugach[4]

[1]Max Planck Institute for the Science of Human History (Germany), [2]University of Zurich (Switzerland), [3]Max Planck Institute for the Science of Human History (Germany), [4]Max Planck Institute for Evolutionary Anthropology (Germany), [5]Universidad de San Martin de Porres (Peru), [6]Laboratorios Zurita & Zurita (Ecuador)

The genetic diversity of the Americas has long been underestimated due to the paucity of high-resolution genetic data available. In South America, genetic studies have shown substantial differentiation between the Andes and Amazonia, with large communities connected by gene-flow in the Andes and small, isolated communities in Amazonia. Yet the diversity of the Pacific Coast, a key region for the migration history of the continent, has been missing from the picture, as has a fine-grained analysis of populations according to their different cultural, linguistic and historical backgrounds.

Here we explore the genetic structure of 24 populations from the Andes, Amazonia and the Pacific Coast, typed with the Affymetrix Human Origins chip. Populations selected to represent different geographical, cultural and linguistic domains show a predominant Native American component: this is particularly unexpected for the Pacific Coast, which historically has been exposed to gene-flow through maritime trade and migrations. This Native American component is a combination of one source of ancestry in the Coast and Andes and at least two distinct sources of ancestry in neighboring Amazonia. Finally, the Native American component admixed with European and African sources at different times, as estimated by a range of independent methods, reflecting the different histories of the encounters with waves of European colonization, in some cases followed by renewed isolation.

Our genetic results, contextualized with respect to the diffusion of major political entities and language families in western South America, shed light on various historical layers and present population diversity.

POB-107

Hybrid type of nuclear and plastid DNAs suggests the hypothesis for ancient propagation of American bottle gourd (Lagenaria siceraria)

dai watabe[1], Hiroshi Yuasa[2], Naoki Osada[1], Kazuhiro Satomura[1], Toshinori Endo[1]

[1]Hokkaido University (Japan), [2]The Research Institute of Evolutionary Biology (Japan)

Bottle gourds distributed widely through Asia, Africa, America and Oceania. They are thought to be one of the oldest cultivated plants, for archaeological records of 10,000 years ago were found in Mexico. Since the exclusive existence of its wild relatives, Africa is considered as the origin. For the first bottle gourd arrival in American Continents, two hypotheses had been proposed that bottle gourds were propagated by human carriage from Asia, and by current floating directly from Africa via Atlantic Ocean based on chloroplast molecular analysis. In this study, we aimed at clarifying the ancient propagation route of American bottle gourds. For this purpose, we analyzed 60 samples collected from local tribes at each site for nuclear and plastid DNAs. In nuclear analysis, samples were classified into three types, which were Asian, African, and both Asian and African hybrid types. All American samples were hybrid type except to one showing pure Asian type. This new findings suggested that Asian and African gourds crossed recently. In plastid analysis, we identified two characteristic insertions/deletions which can distinguish samples into Asian and African types. These results are congruent with Erickson *et al.* (2005) who suggested that ancient Asian domesticated gourds were brought to Americas by human. Given the results, we discussed the propagation route of ancient American bottle gourds. We also discuss the result of sequence mapping onto the genome sequence data recently published by Wu *et al.* (2017).

POB-108

The founding events in different Roma groups revealed through complete Y chromosome sequences

Carla Garcia Fernandez[1], Neus Sole Morata[1], Neus Font Porterias[1], Erica Bianco[1], David Comas[1], Francesc Calafell[1]

[1]University Pompeu Fabra - IBE (Spain)

Genetic, linguistic and cultural evidence point to the Roma having originated in North-Western India, and to have arrived at the Balkan Peninsula around 1kya. After their arrival to Europe, they diverged into socially distinct migrant groups, with different Romani dialects and specific cultural rules, that spread all across the continent.

Little is known about the demographic history and genetic diversity of these groups, because most genetic studies on the Roma have considered only the country of origin instead of the Roma group affiliation.

Given the patrilocal nature of the Roma, we generated 40 whole Y chromosome sequences from five different European Roma populations, by using whole-genome shotgun paired-end sequencing (Illumina HiSeq X Ten). The four major migrant Romani groups were represented, allowing us to study their diversity, admixture patterns and changes in population size.

Preliminary results indicate that three founder haplogroups (H1, I1, J2), comprise 57.5% of the Y Roma lineages, showing the low diversity in the Y chromosomes in the Roma, a consequence of their recent origin and spread.

Interestingly, North-Western Roma showed very similar haplogroup frequencies in Spain and Lithuania despite the large geographical distance between them. In contrast, Romungro Roma and Vlax Roma, who both live in close proximity in Hungary, show different haplogroup compositions. This supports that structure within Roma could be more related to migrant affiliation than to geographical origin.

POB-110

Linguistics as a Complementary Metric for Human Migration History and the Peopling of the Americas

Kara Boyer[1], Nicole Creanza[1], Maanasa Raghavan[2]

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

Despite ongoing efforts to sequence modern and ancient DNA from Native American populations, reconstructing a detailed demographic history of the Americas has remained an elusive goal. To help elucidate the population history of this region, we synthesized and analyzed in parallel genetic, geographic, and linguistic data from populations in the Americas and Siberia. In addition, we studied linguistic data, in the form of syntax and phoneme inventories, from over 500 populations in the Americas. Our previous analysis of worldwide linguistic and genetic data revealed signals of both ancestral relationships and more recent population contact, with geography showing a stronger relationship to genes than to languages. For the current analysis, we compiled 1,226 complete mtDNA genomes from 104 unique populations in northeast Asia and the Americas, and we matched each population to its native language. We then assessed the geographic patterns of genetic and linguistic variation in the Americas using distance-based metrics, principal component analyses, and Procrustes analysis. We found a surprising pattern that stands in contrast to the results of worldwide analysis: in the Americas, languages carry more geographic information than genes do. This pattern was consistent in both distance-based and principal-components-based analyses which showed that the geographic relationships between populations are more closely associated with linguistic variation than with genetic variation. These results hint that population-level language information might be particularly informative about human demographic history in the Americas, thus shedding new light on the patterns of human migrations across the Bering Strait and within the Americas.

POB-111

Population migration and dairy pastoralism on the Bronze Age Mongolian steppe

Christina Warinner[1, 2, 3], Choongwon Jeong[1], Shevan Wilkin[4], Tsend Amgalantugs[5], Abigail Bouwman[3], William Taylor[4], Sabri Bromage[6], Soninkhishig Tsolmon[12], Christian Trachsel[7], Judith Littleton[8], Cheryl Makarewicz[9], Erdene Myagmar[10], Bruno Frohlich[11], Jessica Hendy[4]

[1]Max Planck Institute for the Science of Human History (Germany), [2]University of Oklahoma (United States), [3]University of Zurich (Switzerland), [4]Max Planck Institute for the Science of Human History (Germany), [5]Mongolian Academy of Sciences (Mongolia), [6]Harvard School of Public Health (United States), [7]University of Zurich (Switzerland), [8]University of Auckland (New Zealand), [9]Christian Albrechts University, Kiel (Germany), [10]National University of Mongolia (Mongolia), [11]Smithsonian Institution (United States), [12]Mongolian University of Science and Technology (Mongolia)

The steppe belt that extends across Eurasia was the primary corridor of Eneolithic and Bronze Age migrations that reshaped the genetics of Europe and Asia and dispersed the Indo-European language family. Beginning in the Eneolithic, a new and highly mobile pastoralist society formed on the Western Steppe. These steppe herders expanded both westwards, contributing to the Corded Ware culture of Eastern and Central Europe, and eastwards, contributing to the mobile pastoralist Afanasevo, Sintashta, Andronovo, and Okunevo cultures in Central Asia. The eastern extent of this Western Steppe herder expansion is not well defined. Here we investigate genome-wide ancestry data obtained from 20 Late Bronze Age (16th-9th century BCE) khirigsuur burials from Khovsgol, Mongolia and further investigate evidence for dairy pastoralism by LC-MS/MS analysis of dental calculus. Overall, we observe limited Western Steppe gene flow into Late Bronze Age Mongolia, but adoption of Western ruminant dairying by ca. 1500 BCE.

POB-113

Human demographic history and adaptation in the abode of snow: the Himalayas and the Tibetan Plateau.

Elena Arciero[1], Thirsa Kraaijenbrink[2], Asan -[3], Marc Haber[1], Qasim Ayub[1, 6], Yuan Chen[1], Shane McCarthy[1, 7], Huanming Yang[3], Mark Jobling[4], George van Driem[5], Peter de Knijff[2], Yali Xue[1], Chris Tyler-Smith[1]

[1]Wellcome Sanger Institute (United Kingdom), [2]Leiden University Medical Centre (Netherlands), [3]BGI-Shenzhen (China), [4]University of Leicester (United Kingdom), [5]University of Bern (Switzerland), [6]Monash University Malaysia (Malaysia), [7]University of Cambridge (United Kingdom)

The Himalayas in South Asia provide a diversity of environments for humans, some of which have required substantial genetic adaptation. We have used a combination of SNP-chip data, genome sequences and functional studies to explore the demographic history, genetic structure and signatures of adaptation in the Himalayan populations. We genotyped ~600,000 genome-wide SNPs in 883 Himalayan individuals from 49 different autochthonous groups from Nepal, Bhutan, North India and the Tibetan Plateau in China, and generated high-coverage whole-genome sequences of 87 individuals from a subset of these populations plus three additional ones. We find that the Himalayan populations share a genetic component derived from a common ancestral population, followed by the development of local fine structure correlating with language and geographical distribution. High-altitude adaptation seems to have originated in a single ancestral population and spread widely across the Himalayas. We refined Himalayan population demographic history, using both autosomal and uniparental sequences. Himalayan populations display different proportions of gene flow with neighbouring populations and variable effective population sizes. The identified Y lineages are commonly found in South and East Asia and Tibet, but mostly form distinct clusters when compared to other populations in the region. We find signatures of adaptation to high altitude in EPAS1 and other genes involved in the hypoxic response. We explore in vitro functional validation of variants in the EPAS1 region that have been previously reported to result from introgression of DNA from the Denisovans using cell lines with and without the Denisovan introgressed haplotype.

POB-114

The genetic origins and admixed ancestry characterization of Japanese people

Wen-Ya Ko[1], Koichiro Higasa[2], Han-Yu Wang[1], Maiko Narahara[2], Kaori Iida[3], Fumihiko Matsuda[2], Ryo Yamada[2]

[1]National Yang-Ming University (Taiwan), [2]Kyoto University (Japan), [3]the Graduate University for Advanced Studies (Japan)

Modern human populations are often descended from multiple ethnic groups owning to the complex migration history of human expansion. Tracing the genomic signatures of admixture history can not only reveal previously unknown migration events but also provide critical information that can facilitate genetic profiling of disease susceptibility. In Japan, although it has been studied extensively over the past decades, the genetic origins of Japanese people remain controversial. Current studies support a dual model which suggests that Japanese people are constituted mainly by early human settlements during the Upper Paleolithic period (Jomon people), followed by an admixture event with people migrated from the Korean peninsula around 2300 year ago (Yayoi people). However, the genetic origin(s) of native Jomons remains unclear. Here, we analyzed a combined dataset of whole-genome genotyping data from 2,123 individuals across 97 populations. As a result, we showed that the genomes of most Japanese people contain three major ancestries with distinct geographic distributions. We performed principle component analysis and identified several candidate populations who may share ancestries with Japanese. Our results suggest that the genetic origins of Jomons may consist of multiple migration events from both Southeast and Northeast Asia. We further inferred local ancestry (LA) blocks on each chromosome of Japanese people. The length distributions of LA appear to differ significantly between Yayoi, Hmong, and Jomon ancestries (in a descending order), suggesting that the Hmong admixture could have occurred before the Korean-Yayoi migration.

POB-115

Detectng signatures of polygenic selection in East Asians

Kazuhiro Nakayama[1, 2]

[1]The University of Tokyo (Japan), [2]Jichi Medical Univresity (Japan)

Genetic adaptation to local environmental factors would have played important roles in the worldwide dispersals of prehistoric modern humans. East Asia is an environmentally diverse region and thus human ethnic groups in this region are expected to show heterogeneity in the genomic signatures of the past natural selection. Mongolians are a population of evolutionary interest since they successfully adapted to harsh inner East Asia by nomadism, in contrast to that majorities of other East Asians are agriculturalists and populate more temperate regions. We previously performed a genome scan for a selective sweep in East Asians and identified several genomic regions that would undergo Mongolian-specific adaptive evolution (Nakayama et al. Mol. Biol. Evol. 2017). However, this genome scan was based on outlier detection approaches and thus did not capture signature of polygenic selection effectively. Using dense single nucleotide polymorphism (SNP) array data and a highly consolidated set of human biological pathways, we performed gene set enrichment tests to detect signals of polygenic selection in East Asians. Fixation index values of SNPs were assigned to genes using whole genome expression quantitative trait loci data. We found that SNPs associated with expression levels of genes relevant to metabolism of various substrates, including amino acids, fatty acids, glycans, and sphingolipid, were significantly diversified between Mongolians and agricultural East Asians. Our results indicate that polygenic selection would partly explain adaptation to dietary changes in East Asians and ethnic differences in metabolic phenotypes of present East Asian populations.

POB-116

Population history of Native Siberians

Elena S. Gusareva[1], Hie Lim Kim[1, 2], Vladimir Kharkov N.[3], Vadim Stepanov A.[3], Stephan Schuster C.[1]

[1]Nanyang Technological University (NTU) (Singapore), [2]Nanyang Technological University (NTU) (Singapore), [3]Tomsk National Medical Research Center, Russian Academy of Science, Siberian Branch, Tomsk (Russian Federation)

Northern Eurasia, which spans from the Arctic Ocean down to Inner Asia, and from Eastern Europe to the Pacific Ocean, is a territory of a great historical interest. Over 40 culturally and linguistically diverse indigenous ethnic groups populate this area, while their genetic variation and histories of people migrations in this region and their further migrations to the New World are still poorly studied. We present deep whole-genome sequencing data (>30x) from 300 individuals belonging to 30 distinct indigenous populations living in the North Asia region. We analyzed these data sets with worldwide reference populations and found that West Siberian populations shared their ancestry with Northern Europeans, and also with the Finno-Ugric speakers (Veps and Karelians) and some portion of Southeast Asians. The Central and East Siberians shared much less with Europeans and the Finno-Ugric speakers, while the Northeast component becomes much more prevalent. Koryaks are most distinctive among Siberians. The closest groups to them are Chukchi and Eskimo that also share some ancestry with Native Americans. Our results provide new insights into population histories of Siberians and evidence of ancient gene flow between Siberia and Europe or the New World.

POB-117

Genetic adaptation and ancestry inference of the Taiwanese indigenous and Han people

Yun-Hua Lo[1]

[1]National Yang-Min University (Taiwan)

The Taiwanese indigenous people with an Austronesian ancestry and the Taiwanese Han people who were thought to originate from various provinces of China have co-inhabited in Taiwan for more than 400 years. The possible admixture between two populations complicated the genetic makeup of the recent populations, which might impede diagnostic progression of diseases. Here, we sought to detect genomic signatures of recent positive

selection in these two populations and study their genetic structure in order to better facilitate the genetic aspect of individualized medicine. To detect recent selection, we applied integrated haplotype score (iHS) in 95 samples with 990,732 SNPs of the Taiwanese indigenous people from Taitung and 14,401 Taiwanese Han samples with 562,983 SNPs collected by the Taiwanese Biobank. 12 and 26 loci that are potentially under positive selection were identified in the Taiwanese indigenous and Han people, respectively. In both populations, some SNPs found in the region of HLA genes showed strong selection signals whereas no other strong iHS signals were shared between populations. In addition, ADMIXTURE analysis showed that the Taiwanese indigenous people contain one major genetic component that contributes a great proportion to many of ISEA populations, which might reflect the events of Austronesian language dispersal. Three major components contribute to the genome of the Taiwanese Han people and they can be traced back to three geographic areas. Combining these two analyses, we provide a better understanding of genetic profile in Taiwanese people for further study in genetic variants underlying disease susceptibility.

POB-118

Lipidome Changes in Prefrontal Cortex of African American, Caucasian and Han Chinese Populations

Anna Tkachev[1, 2], Ekaterina Khrameeva[1, 2], Philipp Khaitovich[1, 3, 4]

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

Lipids constitute a large portion of the brain, and are an important manifestation of the individual's brain phenotype. Little is known about the degree to which genetic similarity characterizes similarity in brain phenotypes. In this study, we utilize large-scale mass-spectrometry to analyze the brain lipidome composition of three population groups, African American, Caucasian and Han Chinese(Asian), the last constituting the world's largest ethnic group. Surprisingly, we discover the brains of Asian individuals to have the highest amount of specific lipids between the three groups. Using machine learning techniques, we show that these differences in the brains of Asian individuals are particular to adult stages of development, and the same differences are not observable for the youngest ages studied (<1 year). Additionally, these differences are contrasted by expression data that was measured for the same three population groups, for which an excess of genes with expression patterns specific to Han Chinese population was not detected.

POB-120

Genetic diversity and admixture history of the Uyghurs in Xinjiang

Shuhua Xu[1]

[1]CAS-MPG Partner Institute for Computational Biology (China)

Asia harbors substantial cultural and linguistic diversity, but human genetic diversity in vast regions, such as western China which is geographically close to Central Asia, remains poorly investigated. Here we systematically assess genetic diversity and ancestry of the Uyghur people residing in Xinjiang (XJU), with a population size more than 10 million, by genotyping ~1,000 Uyghur individuals from 13 geographical regions with high-density single nucleotide polymorphism arrays. We unveiled a southwest-northeast differentiation within XJU, which is contrasted to the north-south divergence as previously expected assuming Tianshan Mountain to be a natural barrier. In the context of comparative analysis of 2,345 individuals representing 203 worldwide populations, we identified four major ancestries in XJU which were derived from East Asia (28.8%-46.5%), Siberia (15.2%-16.8%), West Eurasia (24.9%-36.6%) and South Asian (12.0%-19.9%). The time and sequence of each ancestry introduced into XJU gene pool are complicated, but can be modeled as an admixture of mixed Western ancestries and mixed Eastern ancestries. We develop a new method, MultiWaver, to infer the admixture history of XJU which can be largely explained by a process of two-wave admixture, with the ancient wave of admixture dated back to 3,750 years ago, and a more recent wave about 750 years ago. Our results provide further insights into human genetic diversity and history of human dispersion in Eurasia.

POB-121

Population genomics of ancient Beringia

Martin Sikora[1], Jose Victor Moreno-Mayar[1], Vitor C Sousa[2], Morten E Allentoft[1], Ben A Potter[3], Lasse Vinner[1], Matthias Steinrucken[8], Simon Rasmussen[4], Yun S Song[5], Rasmus Nielsen[5], Laurent Excoffier[7], David J Meltzer[6], Eske Willerslev[1]

[1]University of Copenhagen (Denmark), [2]Universidade de Lisboa (Portugal), [3]University of Alaska (United States), [4]Technical University of Denmark (Denmark), [5]University of California, Berkeley (United States), [6]Southern Methodist University (United States), [7]University of Bern (Switzerland), [8]University of Michigan (United States)

Beringia, the region comprising far northeast Asia and northwestern America, is an important region in human population history. There is broad agreement that modern humans initially populated the Americas via a land bridge exposed during the Pleistocene, which today forms the Bering Strait separating the two continents. Despite this agreement, many questions about its history and early peoples remain poorly understood, including for how long Native American ancestors were isolated from Asian groups in Beringia before entering the Americas, or whether one or more early migrations gave rise to the founding population of Native Americans.

To address these questions, we sequenced genomes of ancient human remains from Beringia, including two infants were recovered at Upward Sun River (USR) in Alaska, dated to around 11.5 thousand years ago (ka). We show that the "Ancient Beringians" form a previously unknown early American population, which forms an outgroup to all previously sequenced contemporary and ancient Native Americans, diverging from them between 22 - 18 thousand years ago. We also find evidence for additional gene flow events across the Bering Strait after 11.5 ka, including Siberian-related ancestry into some of the northern Native American populations. Finally, we will also present novel data on ancient Beringians from the old world. Our findings suggest that population history of Beringia was highly dynamic and complex from the Late Pleistocene well into the Late Holocene, shaped by multiple waves of migrations and large-scale population replacements.

POB-122

History of risk allele of diffuse-type gastric cancer in East Asian

Risa Iwasaki[1], Takehiro Sato[2], Ryosuke Kimura[3], Yoko Satta[1]

[1]SOKENDAI (The Graduate University for Advanced Studies) (Japan), [2]Kanazawa University (Japan), [3]University of the Ryukyus (Japan)

It is reported the incidence of diffuse-type gastric cancer is much higher in East Asian populations among the world populations. Especially, between Han Chinese and Japanese, this cancer is known to be more common in Japanese. In fact, risk allele of this cancer is higher in Japanese in Tokyo (JPT: 63%) than in Han Chinese in Bejing (CHB: 24%). In order to reveal how genetic differentiation in Japanese is associated with this higher incidence of gastric cancer, SNPs related to this disease in the 1000 genomes projects were examined. Through measuring F_{st} per SNP in the genome, the highest differentiated SNPs in JPT in comparison with CHB were located adjacent to and within *PSCA* gene on chromosome 8. *PSCA* gene is reported to be associated with various types of cancers. Interestingly, one of this highest differentiated SNPs is the risk SNP for diffuse-type gastric cancer. To elucidate what specific evolutionary forces elevated the SNP frequency in JPT, I performed several neutrality tests (EHH, nSL), but did not detect significant signals of selection. This suggests the involvement of demographic history specific to Japan, i.e., admixture of genetically distinct populations, the Yayoi and Jomon peoples (dual structure model by Hanihara). Frequencies of the disease-related SNPs/haplotypes in Ainu and Ryukyuan populations (possible descendants of the Jomon) as well as Asian continent populations (possible descendants of the Yayoi) are consistent with the model. In future, I will examine in details which model of the admixture explain the differentiation of the risk SNP.

POB-123

Genome-wide genetic variation of people in Japan and its implication of demographic history

Yosuke Kawai[1], Takahiro Mimori[2], Yuki Hitomi[1], Seik-Soon Khor[1], Olivier Gervais[2], Masao Nagasaki[2], Katsushi Tokunaga[1]

[1]The University of Tokyo (Japan), [2]Tohoku Univerisity (Japan)

The advent of genome-wide analysis technology allowed the collection of large amount of human genetic variation in population scale. This advantage allows us to investigate the recent population genetic events such as population expansion, migration and/or admixture. Japanese Archipelago is one of the end point in human migration. Population structure analyses of east Asian population inferred by principal component analysis have shown that populations in Japan are isolated from other east Asian populations. We investigated the genetic history of the Hondo Japanese population using the genome-wide data of people in near Tokyo, Japan. Genetic variations were first analyzed by whole genome sequencing and SNP array genotyping. We investigated haplotype sharing within the population to infer very recent population size change. We also analyzed cross coalescent rate between individuals to infer the time scale of genetic differentiation of Hondo Japanese. These results suggest that the genetic differentiation within Honshu island has begun after the last migration event in Yayoi period.

POB-127

Mitochondrial-encoded genes contribute to thermal divergence between Saccharomyces species

Xueying C. Li[1], David Peris[4], Chris Todd Hittinger[4], Elaine A. Sia[2], Justin C. Fay[3, 2]

[1]Washington University in St. Louis (United States), [2]University of Rochester (United States), [3]Washington University in St. Louis (United States), [4]University of Wisconsin-Madison (United States)

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

POB-129

A screen for mitochondrial genetic effects on locomotory performance of fruit flies across a thermal gradient, and implications for climatic adaptation

Venkatesh Nagarajan-Radha[1], Damian Kimon Dowling[1]

[1]Monash University (Australia)

The traditional view that the genetic variation that accumulates within the mitochondrial genome is selectively neutral has changed over the past decade. Numerous studies have now tied the sequence variation in the mtDNA sequences of metazoans to effects on the organismal phenotype. While the evolutionary forces that have facilitated the accumulation of this non-neutral mtDNA variation remain unclear, several studies have implicated a key role for adaptive selection, particularly selection to the thermal climate. The Australian distribution of fruit flies, Drosophila melanogaster, is characterised by the occurrence of two major mtDNA haplotypes. These haplotypes exhibit latitudinal clines, and previous experiments have shown that flies harbouring the northern-predominant (sub-tropical) haplotype exhibit greater resilience to heat stress, but decreased resilience to cold stress, than flies harbouring the southern-predominant (temperate-latitude) counterpart. Here, we test for differences in locomotory activity of flies of each haplotype, across a thermal gradient spanning six temperatures. We predicted the flies harbouring the northern-predominant haplotype would maintain higher activity levels at high temperatures than flies with the southern haplotpes, and vice versa, that the southern flies would be more active at cooler temperature. Our data suggests mitochondrial genetic effects on locomotory activity of flies in warmer but not in colder temperatures. Furthermore, we found significant temperature effects on the locomotory performance, but these effects were sex-specific, and also contingent on the time of day in which the flies were assayed. Our results thus suggest a contribution of the mitochondrial genome in mediating adaptive dynamics of thermal adaptation.

POB-130

Mitochondrial-Nuclear Interactions and the Thermal Sensitivity of Spermatogenesis

Abhilesh Dhawanjewar[1], Kristi Montooth[1], Colin Meiklejohn[1]

[1]University of Nebraska-Lincoln (United States)

Spermatogenesis appears to be sensitive to temperature in all animals and this thermal sensitivity may have important implications for the adaptation of ectothermic species to their local environments. Male sterility temperature thresholds show considerable variation within and between Drosophila species that correlates with geographical distributions and local climatic conditions. While mitochondrial genomic variation is often suggested to play a role in thermal adaptation, there is little empirical evidence that implicates mitochondrial variation in male thermal sterility. We investigated the role of the mitochondrial genome (mtDNA), the nuclear genome (nucDNA) and interactions between them in shaping these sharp thermal-sterility thresholds using characterized mitochondrial-nuclear (mito-nuclear) hybrid genotypes. We establish thermal-sterility thresholds for hybrid genotypes that combine mtDNAs from *D. melanogaster* and its sister species *D. simulans* with two *D. melanogaster* nuclear backgrounds. We confirm that a particular mito-nuclear incompatibility results in a lower thermal-sterility threshold, and present new data indicating that diet modifies this phenotype. Cytological dissections of male testes show the presence of sperm in all heat-induced sterile males indicating the sterility is due to the production of non-functional sperm. Preliminary results indicate a crucial checkpoint for spermatogenesis in early development, with the thermal-sterility phenotype being sensitive to the life stage at which flies experience thermal stress. Finally, we present results from experiments that shift diet and thermal regimes to determine the conditions under which this phenotype is reversible in order to gain insight on the developmental and energetic basis for this critical component of fitness.

POB-132

Intracellular and intraorganellar co-expression of divergent mitochondrial electron transport chain subunits in the germ line of a naturally heteroplasmic species

Fabrizio Ghiselli[1], Maria Gabriella Maurizii[1], Helena Arino[2], Carmine Cifaldi[1], Arkadiy Reunov[3], Yana Alexandrova[4], Andrea Pecci[1], Simone Bettini[1], Marco Passamonti[1], Valeria Franceschini[1], Liliana Milani[1]

[1]University of Bologna (Italy), [2]Universitat de Barcelona (Spain), [3]University of Ottawa (Canada), [4]Far Eastern Branch of Russian Academy of Sciences (Russian Federation)

Mitochondrial heteroplasmy is the presence of more than one type of mitochondrial genome (mtDNA) within an individual, and in most of the reported cases it seems to be an unfavourable condition. The only known natural and evolutionarily stable heteroplasmic system in Metazoa is the Doubly Uniparental Inheritance (DUI), reported in ~100 bivalve species, in which two mitochondrial lineages are present: one transmitted through females (F-type) and the other through males (M-type). The two lineages are highly divergent (up to 52% amino acid difference in protein coding sequences). Gametes are homoplasmic for the sex-specific type, female soma is homoplasmic for the F-type, and male soma is heteroplasmic to various degrees, depending on tissue and species. So far, no study has investigated such heteroplasmy at the protein level, and no analysis has been performed to clarify if it is present at the cell or organelle level. In the DUI species *Ruditapes philippinarum*, we immunolocalized F and M forms of three mitochondrially-encoded proteins (ND5, CYTB, COX3) in germ line and somatic tissues of females and males at different developmental stages. Somatic tissues resulted homoplasmic in females and heteroplasmic in males. While gametes turned out to be homoplasmic, both female and male stem cells and early germ cells unexpectedly showed the presence of both F- and M-type, suggesting a mechanism of sex-specific mtDNA selection during germ line differentiation. Most interestingly, the heteroplasmy in early germ cells was observed at the organelle level, namely, both F- and M-type variants were in the same mitochondrion.

POB-133

Molecular evolution of OXPHOS protein subunits in fishes with novel phenotypes

Ahmed A Elbassiouny[1, 2], Belinda S.W Chang[1, 3, 4], Nathan R. Lovejoy[1, 2, 3]

[1]University of Toronto (Canada), [2]University of Toronto Scarborough (Canada), [3]University of Toronto (Canada), [4]University of Toronto (Canada)

Mitochondria are crucial organelles for cellular function, because they carry out steps of cellular respiration that contribute ~80% of cellular ATP molecules. More notably, oxidative phosphorylation machinery (OXPHOS) in the mitochondrial inner membranes are key proteins in the generation of ATP molecules and eliminating reactive oxygen stress. OXPHOS proteins are known to be large, multi-subunit complexes that are encoded for by both the nuclear and mitochondrial genomes. Although non-synonymous mutations in OXPHOS subunits genes have been associated with detrimental human diseases, the exact effect of single mutations on the phenotype is hard to determine, mainly due to the large number of interacting subunits involved and the large number of mitochondria in each cell. Here, we investigate the molecular evolution of OXPHOS subunits, and utilize biophysical models of OXPHOS protein complexes to gain insights into functional consequences of non-synonymous mutations in fishes adapted to naturally-high metabolic burden. Using this approach, we are able to link molecular adaptation at the OXPHOS subunits level, to adaptations to high metabolic needs at the organismal level.

POB-137

Transition transversion ratio in mitochondrial genome is higher in long- versus short-lived mammalian species: effects of ROS and replication timing?

Alina G. Mikhaylova[1], Alina A. Mikhaylova[1], Kristina Ushakova[1], Evgenii Tretyakov[1], Andrey Yurchenko[2], Dmitry Knorre[3], Ilia Mazunin[1], Alexandre Reymond[5], Konstantin Gunbin[4, 1], Konstantin Popadin[5, 1]

[1]The School of Life Sciences (Russian Federation), [2]Institute of Biodiversity Animal Health and Comparative Medicine (United Kingdom), [3]The A.N. Belozersky Institute Of Physico-Chemical Biology (Russian Federation), [4]The Institute of Cytology and Genetics of the SB RAS (Russian Federation), [5]Center for Integrative Genomics (Switzerland)

Transition/transversion ratio (ts/tv) in mitochondrial genomes of animals significantly differs between species, families and orders, however no universal explanation has been suggested. Using four-fold degenerative synonymous polymorphisms of mammalian mitochondrial genomes here we reconstructed mutation spectra for 300 species. The average mutation signature is very similar to the recently described mutation signature derived from somatic mitochondrial mutations in human cancers: there are two most common types of substitutions (G->A and T->C transitions using notation of L chain) both of them demonstrating strong strand asymmetry (occurring mainly on a heavy chain: C->T and A->G transitions using notation of H chain). Comparing mutation spectra of long- versus short-lived mammals we observed a gradient: species-specific ts/tv increases with generation time and this correlation is robust to numerous potential confounders, such as nucleotide content, phylogenetic inertia and types of analyzed mitochondrial genes. Our findings might be explained by two, non-mutually exclusive hypotheses (i) short-lived species have increased basal metabolic rate and thus can suffer from the increased burden of ROS, manifesting itself by G:C->T:A transversions; (ii) long-lived species have prolonged replication of mtDNA and thus accumulate more C->T and A->G transitions occurring on single-stranded heavy chain during replication. Our findings are in line with previously observed correlations between mitochondrial nucleotide content and mammalian lifespan and emphasize that at least some of them are driven by purely mutagenic not selective forces.

POB-138

Variability in Gibbs energy of tRNA molecules in mitochondrial genomes of Chordates: neutral selection or evolution towards optimization of translation?

Kristina Ushakova[1], Alina A. Mikhailova[1], Alina G. Mikhailova[1], Dmitry Knorre[2], Ilia Mazunin[1], Alexandre Reymond[4], Konstantin Gunbin[3, 1], Konstantin Popadin[4, 1]

[1]Immanuel Kant Baltic Federal University (Russian Federation), [2]Moscow State University (Russian Federation), [3]Novosibirsk State University (Russian Federation), [4]University of Lausanne (Switzerland)

It is known that translation of frequent codons in prokaryotes and some eukaryotes is optimized by increasing the copy number of corresponding tRNA gene. However, highly streamlined Chordate mitochondrial genomes mostly hold only one tRNA gene for each amino acid. So how is mitochondrial translation optimized? We hypothesized that stability of tRNA molecules is an important variable, correlating with codon usage (CU). To test this hypothesis we reconstructed secondary structures and Gibbs Energy of each tRNA from ~4000 Chordata mito-genomes. We also reconstructed ancestral tRNA states, using CAT evolutionary model, at each internal node of phylogenetic tree to observe evolutionary stability trend. Observations:

In different classes of Chordates tRNA stabilities are highly variable: more stable in Aves than Mammalia, in Actinopterygii than Amphibia and Reptilia. GC% of the whole mito-genome demonstrates the same relationship, suggesting, tRNA stability, might be just a neutral consequence of the whole genome GC%. However, comparing tRNA GC% with whole genome - we observed that warm-blooded opposed to cold-blooded Chordates have increased tRNA GC% versus background - it is possibie tRNA stability might be under stronger selection in species with high basal metabolic rate.

Comparing different species within each class, we observed positive correlations between tRNA stability and whole genome GC%.

Comparing different tRNA molecules within the same genome of each species, we observed a positive correlation between tRNA stability and CU, especially in warm-blooded species. Conclusion: warm-blooded Chordate tRNA stabilities tend to be more selectionally constrained for translation efficiency than those of cold-blooded Chordates.

POB-141

Co-evolution within a synthetic microbial community

Philippe Piccardi[1], Sara Mitri[1]

[1]University of Lausanne (Switzerland)

Evolutionary dynamics are little understood within large ecosystems that contain hundreds of microbes species interacting with complex living organisms. Disentangling evolving human-microbe interactions is therefore challenging without a thorough understanding of simpler systems. In addition, since bacteria evolve faster than their hosts, it makes sense to first understand microbial co-evolution in the absence of a host. In this project, we focus on a synthetic bacterial community consisting of just four species involved in bioremediation. Using a well-defined medium, the four species grow in a closed system, in which we can closely follow bacterial abundances, the interactions between the four species and their genetic changes, during a long-term evolutionary experiment. Prior to the beginning of this evolutionary experiment, we assessed that the species have positive growth effects on one another. Our experiment should now address the following questions: what will happen to these positive interactions over time? What are the key genetic elements that drive these bacterial interactions? We will answer these by periodic deep sequencing to establish how mutations affect the behavior of the bacteria and their interactions. Although the results of this experiment are not yet available, we will be ready to present them during the conference. This project will help us to understand and better predict co-evolutionary dynamics in more complex models, such as in the human host.

POB-142

Reconstruction of Oral Microbiomes from Extinct and Extant Anthropoids through ancient DNA

James A Fellows Yates[1], Oral Microbiome Evolution Consortium -[1], Matthew C Curtis[2], J. Carlos Diez[3], Victoria E Gibbon[4], Mario Menedez[5], Marco Peresani[6], Mirjana Roksandic[7], Michael J Walker[8], Robert C Power[9], Domingo C Salazar-Garcia[1, 10], Johannes Krause[1], Alexander Herbig[1], Christina Warinner[1]

[1]Max Planck Institute for the Science of Human History (Germany), [2]University of California Los Angeles (United States), [3]Burgos University (Spain), [4]University of Cape Town (South Africa), [5]National University of Distance Education (Spain), [6]Universita di Ferrara (Italy), [7]University of Winnipeg (Canada), [8]Universidad de Murcia (Spain), [9]Max Planck Institute for Evolutionary Anthropology (Germany), [10]KERBASQUE-Basque Foundation for Science (Spain)

While modern microbiome research has shown the importance of our microbial communities in health and disease, research has tended to focus on the gut microbiome of either Western industrialised societies or captive animals. Recent discoveries in the field of archaeogenetics have revealed dental calculus from skeletal remains as a rich source of well-preserved bacterial DNA. In contrast to sampling from live individuals, dental calculus from archaeological remains presents an opportunity to 'non-invasively' study the oral microbiome from a wider diversity of species and populations.

We present results of a total over 3.5 billion shotgun DNA sequencing reads from ancient and modern dental calculus from over 90 hominids, including gorillas (29), chimps (20) and humans (45), as well as 14 Neanderthals from the Late Pleistocene and 5 New World monkeys. We will show initial (meta)genomic analysis of similarities and differences in the oral plaque microbiome at different stages of anthropoid evolution. While preservation is variable in archaeological samples, dental calculus from the Late Pleistocene can still yield authentic ancient DNA attributable to known microbial taxa found in the modern human oral cavity. Expanding our knowledge of the diversity of the human microbiome through time and space will be important in understanding the deep relationship between hosts and their microbial communities.

POB-145

The co-occurrence and co-exclusion of evolving objects in prokaryotes

Fiona Jane Whelan[1], Martin Rusilowicz[1], James McInerney[1, 2]

[1]The University of Manchester (United Kingdom), [2]University of Nottingham (United Kingdom)

Throughout evolution, evolving objects (domains, genes, operons etc.) have continuously combined, forming new proteins, gene clusters, and genomes. Horizontal gene transfer, particularly among prokaryotes, has facilitated this combinatorial process. Thus, evolving objects that interact positively or synergistically with each other are expected to co-occur more often than by chance; conversely, evolving objects may avoid co-occurrence, indicating an adverse interaction between objects. In this work, we use methods adapted from graph theory to understand patterns of co-occurrence and co-exclusion in prokaryotes. We have implemented multi-level graph models in which each node (vertex) is a domain, gene, operon, or (meta)genome connected by an edge (line) to another node to display co-occurrence and co-exclusion relationships. We demonstrate how these concepts can be used to show significant associations between units of selection. We apply these multi-level graph models to a variety of datasets including a representative set of prokaryotes, prokaryotic pangenomes, and microbial metagenomics. We find evidence for evolving objects that significantly co-occur with each other across each of these datasets; these genetic clusters include objects from characterized biological pathways but also include genes with unknown functions. Further, we identify co-occurring patterns that are only observed in particular datasets (for e.g. in human versus environmentally-associated metagenomes). We also identify genes that co-exclude each other, indicating evolving objects with non-complementary biological functions. This work represents a different approach to understanding the evolution of prokaryotes and allows us to draw novel hypotheses as to the potential role of these genetic clusters in prokaryotic biology.

POB-149

Paternal Lineage Bursts in the Western Mediterranean

Francesc Calafell[1], Neus Sole-Morata[1], Carla Garcia-Fernandez[1], Patricia Villaescusa[2], Marian M. de Pancorbo[2], David Comas[1]

[1]Universitat Pompeu Fabra (Spain), [2]Univerity of the Basque Country (Spain)

Resequencing the Y chromosome in human global samples (Poznik et al., Nat. Genet. 48:593-599) led to the discovery that many patrilineal lineages had experienced bursts, i.e., sudden increases in frequency, often linked to migrations and technological innovations. Here we show two examples of such bursts, which, even though they occurred quite recently over a limited geographic scale, thoroughly remodeled the Y-chromosome landscape in their respective populations.

Haplogroup R1b-DF27 is a branch of R1b-P312, the most common haplogroup in Western Europe. We genotyped it (as well as six SNPs downstream of it and 17 Y-STRs) and found that its frequency is about 40% in Iberians (up to 70% in Basques), drops to 5-20% in France and Britain, and is rare elsewhere. We estimated its age at about 4,500 ya, at the transition between the Neolithic and the Bronze Age, when the Y chromosome landscape of W Europe was thoroughly remodelled.

E-M183 (E-M81) is dominant in NW Africa (50-75%) and rare elsewhere. We resequenced 32 North Africans, and identified five new branches within E-M183. The validation of these variants in more than 200 North African samples, from which we also had information of 13 Y-STRs, revealed a strong resemblance among E-M183 Y-STR haplotypes that pointed to a rapid expansion of this haplogroup. Both SNP and STR data revealed an extremely recent origin (2,000 - 3,000 ya). Ancient DNA evidence together with our TMRCA estimates point to a local origin of E-M183 in NW Africa.

POB-150

Adaptive eQTLs in human populations reveal the evolutionary impacts of pleiotropy and tissue-specificity

Joseph Lachance[1], Melanie Quiver[1]

[1]Georgia Institute of Technology (United States)

Adaptive evolution is often due to regulatory sequence change, and allele frequencies can vary greatly across human populations at expression quantitative trait loci (eQTLs). We combined continental allele frequencies with data from 48 tissues and calculated population branch statistics (PBS) to identify adaptive eQTLs. Our results indicate that pleiotropic eQTLs are less likely to be adaptive than eQTLs that affect only a single tissue. We also find evidence that adaptive eQTLs tend to not have large effect sizes. The proportion of eQTLs that are adaptive varies by tissue, and we find that eQTLs that regulate expression in testes, blood, or sun-exposed skin are enriched for PBS outliers. By contrast, eQTLs that regulate expression in the cerebrum or female-specific tissues have a relative lack of PBS outliers. Scans of selection also reveal that the strongest adaptive signal in many regions of the human genome is an eQTL.

POB-151

Blunted nitric oxide regulation and high altitude adaptation in Tibetans

Yaoxi He[1], Xuebin Qi[1], Chaoying Cui[2], Bing Su[1]

[1]Kunming Institute of Zoology, Chinese Academy of Sciences (China), [2]Tibetan University (China)

Nitric oxide (NO) is an important molecule for vasomotor tone, and an elevated NO signaling was previously proposed as a unique and adaptive physiological change in highlander Tibetans. However, there has been a lack of NO data from Tibetans living at low altitude and lowlanders living at high altitude, which is critical to test the hypothesized adaptation in Tibetans. Here through cross-altitude (1,990m-5,018m) and between-population (Tibetans vs. Han Chinese) analysis of blood NO metabolites (NOx) of 2,086 individuals, we demonstrate that although Tibetans living at high altitude do have a higher blood NOx concentration than that of lowlanders living at low altitude, it is not unique to Tibetans. Han Chinese immigrants living at high altitude show an even higher blood NOx concentration than Tibetans. Consequently, our data rejects the previous proposal of an increased NO signaling in Tibetans as the adaptive strategy to hypobaric hypoxia. Instead, Tibetans have a blunted blood NOx regulation, similar with the known blunted regulation of hemoglobin. This observation is further confirmed by data from the hypoxic experiments using human umbilical vein endothelial cells and gene knockout mice. Furthermore, we identified a group of genes (GCH1 and others) carrying Tibetan-enriched mutations that contribute to the blunted NOx regulation in Tibetans. Collectively, our findings provide new insights into the physiological and genetic mechanisms of adaptation to high altitude hypoxia in Tibetans.

POB-152

High Resolution SNP Genotyping of Y-Chromosomes from Kazak Populations Living in Jetisuu Region, Kazakstan Affirms a Common Paternal Ancestry

Ayken Askapuli[1, 2], Miguel G Vilar[3], Maxat Zhabagin[2], Sabitov Zhaxylyk[4], Timothy A Jinam[6], Schurr G Theodore[5], Zhaxybay Zhumadilov[2], Naruya Saitou[6, 7]

[1]Astana Medical University, Astana (Kazakhstan), [2]Nazarbayev University, Astana (Kazakhstan), [3]National Geographic Society, Washington DC (United States), [4]L.N. Gumilyov Eurasian National University, Astana (Kazakhstan), [5]University of Pennsylvania, Philadelphia, PA (United States), [6]National Institute of Genetics, Mishima (Japan), [7]University of Tokyo, Tokyo (Japan)

In this study, we genotyped 90 Kazak individuals with 2140 Y-chromosome SNPs via the GenoChip 2.0 (Illumina iSelect HD custom genotyping bead array). Of the 90 individuals genotyped, 80 belonged to four Kazak populations (clans) living in the Jetisuu region in southern Kazakstan. The remaining 10 individuals, who belonged to other Kazak clans, were also analyzed for comparative purposes. In addition to the high-resolution genotyping of Y-chromosomes, the current study also paid special attention to the association between clan identity and Y-chromosome profile. This genotyping revealed that 62.5% of the 80 Jetisuu individuals harbored the same Y-chromosome haplotype from C3* Star Cluster (a.k.a. C3b1a3-F1918 or C2b1a3), which is hypothesized to represent the male lineage of Genghis Khan. However, as expected, other Y-chromosome lineages were also observed in these groups, including R1a (11.25%), N1a (6.25%), R2a (3.75%), J2a (3.75%), J1a (2.5%), Q1a (2.50%), I2a (2.50%), G2a (2.50%), T1b (1.25%), and E1b (1.25%). In Kazak genealogy, there are three major divisions called the Great Juz, Middle Juz, and Junior Juz. In this study, we explored genetic variation in four clans of the Great Juz, including the Jalayir, Alban, Shapirashti, and Suuan. We noted the genetic basis of the common ancestry of the Kazak clans, as indicated in their genealogy, albeit with certain degrees of admixture from other paternal lineages.

POB-154

Modelling human adaptations in mice

Michal Szpak[1], Yali Xue[1], Qasim Ayub[1, 5], Valerie Vancollie[1], Yvette Hooks[1], Neil J Ingham[2], Morag A Lewis[2], Mark Campbell[3], Sergio Rodriguez-Cuenca[3], Toomas Kivisild[4], Antonio Vidal-Puig[3, 1], Karen P Steel[2], Chris J Lelliott[1], Chris Tyler-Smith[1]

[1]Wellcome Sanger Institute (United Kingdom), [2]King's College London (United Kingdom), [3]University of Cambridge (United Kingdom), [4]University of Cambridge (United Kingdom), [5]Monash University Malaysia (Malaysia)

Following the out-of-Africa expansion, humans have adapted to a diverse range of new environments and selective pressures. Scanning genomes for population-genetic signatures of such adaptations yields vast lists of thousands of genetic candidates. Their functional validation and investigation of their biological consequences is a current roadblock in the field, limiting both our understanding of the selected phenotypes and more generally the importance of positive selection. Successful examples show that in-depth follow-up studies of putatively-selected variation using genome editing and model organisms constitute a suitable tool to uncover the underlying biology. However, modelling of non-pathological human variation has received limited attention to date. We therefore developed a tool for pinpointing and prioritisation of candidate selected variants for functional follow-up studies (*FineMAV*). After compiling lists of ~100 high-priority candidates across different continental populations, we have begun investigating several of them using CRISPR/Cas9 technology to generate 8 mouse knock-outs, and 21 knock-ins carrying the human selected allele. Examples particularly relevant to Asia include knock-outs of *HERC1* and *PRSS53* which reveal diverse phenotypes and a single hair-related phenotype, respectively. Mouse knock-in of a *CPT1A* allele selected in Siberians recapitulated some aspects of the human metabolic phenotype expected, while the *PCDH15* allele selected in East Asia confers a detectable hearing phenotype, although it is a progressive decrease in high-frequency hearing sensitivity, which is unexpected. Linking mouse phenotype to fitness in humans is thus complex; nevertheless, animal models provide one of the few ways to test hypotheses regarding recent human evolution and need large-scale evaluation.

POB-155

The European heritage of American populations

Francesco Montinaro[1, 2], Linda Ongaro[1, 3], Marilia Scliar[4], "Kristina Tambets," Jose Rodrigo Flores Espinosa[1], "Stefania Sarno," Guido Alberto Gnecchi Ruscone.[6], Donata Luiselli[6], Marta E. Alarcon-Riquelme[7], Andres Moreno Estrada[8], "Alessandro Achilli," "Ornella Semino," "Anna Olivieri,"[5], Antonio Torroni[5], Cristian Capelli[2], Eduardo Tarazona Santos[4], Luca Pagani[1, 9], Mait Metspalu[1]

[1]University of Tartu, Tartu (Estonia), [2], University of Oxford, Oxford (United Kingdom), [3]University of Tartu (Estonia), [4]Departamento de Biologia Geral, Instituto de Ciencias Biologicas, Universidade Federal de Minas Gerais (Brazil), [5] University of Pavia, Pavia (Italy), [6]University of Bologna, Bologna (Italy), [7]Centro Pfizer - Universidad de Granada - Junta de Andalucia de Genomica e Investigacion Oncologic (Spain), [8]Cinvestav (Mexico), [9] University of Padua, Padua (United Kingdom)

The genetic diversity of the Americas has been shaped by several events of admixture and gene flow from both Europe and Africa as a result of the Colonial Era and the Atlantic Slave Trade. Additional waves of admixture began in the XIX century, contributing further layers of complexity to the whole picture. These multiple incoming population entities added further variation to the genomic landscape of the double continent, thus a precise characterization of the admixture dynamics is now crucial for a wide range of disciplines.

The recent explosion of genomic data production, coupled with the development of methods harnessing the variability created by recombination, make now possible a fine reconstruction of the recent ancestry profile of human populations.

Here, to finely dissect the genomic landscape of the Americas, we first compiled a genome-wide dataset of ~13,000 individuals from twelve American countries, which was combined with genome-wide data from ~7,000 individuals of European descent. A number of haplotype-based methods were then employed to evaluate admixture dynamics through time.

Focusing on the European influence in the continent, we were able to decrypt and describe the high complexity of the genetic contribution of European populations in America, both under geographic and temporal perspectives.

POB-156

Characterizing archaic hominid's contribution to genetic variation in modern human populations following migration out of Africa

Kelsey E. Witt[1], Jonathan Rice[1], Emilia Huerta-Sanchez[1]

[1]University of California - Merced (United States)

The out-of-Africa bottleneck was instrumental in shaping patterns of modern human genetic variation. Traditionally, studies have characterized genetic variation in non-Africans using a serial bottleneck model, and assumed that new genetic variation in non-Africans was generated via *de novo* mutations. However, ancient DNA studies have revealed that interbreeding between humans and archaic humans (Neanderthals and Denisovans) outside of Africa introduced additional genetic variation to non-African populations. Here we quantify how much "archaic" variants contributed to new genetic variation in non-African populations by examining uniquely shared mutations between archaic humans and non-Africans that are rare in Africa ($f < 0.01$), and therefore likely to have arisen after humans left Africa. We examined both ancestral and derived variants, relative to the chimpanzee genome, and compared the variants shared with archaic hominids to those found only in modern human populations. We found that 5.5-7% of derived variants not found in Africans, and 65% of non-African ancestral alleles, are archaic alleles, and so were likely introduced through introgression. The majority of these variants are shared with Neanderthals, although the ratio of Denisovan to Neanderthal variants differs between populations. The patterns of "archaic" genetic variation in non-African populations largely match what is expected given the successive bottlenecks that occurred as humans migrated across Eurasia and into the Americas. However, some populations had more common variants ($f > 0.2$) or rare variants ($f < 0.2$) than expected, suggesting that local selection events have impacted the archaic variation present in modern populations.

POB-157

Recent human demographic history confounds population genetic inference of recessive selection

Daniel M. Jordan[1], Daniel J. Balick[1, 2], Shamil R. Sunyaev[2], Ron Do[1]

[1]Icahn School of Medicine at Mount Sinai (United States), [2]Brigham and Women's Hospital and Harvard Medical School (United States)

Over the last decade, many large-scale sequencing studies of human populations have been performed. Each such study produces its own slightly different estimate of parameters of recent human demographic history, including the rate of recent explosive growth since the out-of-Africa bottleneck. Methods that use demographic models to infer the parameters of natural selection typically ignore these quantitative differences; however, this may not be appropriate when inferring dominance as well as selection, as recessive selection is highly sensitive to historical population size changes. We perform forward simulations of three realistic models of European demographic history, derived from three different large human population sequencing studies, and apply a likelihood framework to infer dominance and selection jointly. We demonstrate that strong additive purifying selection can be distinguished from all other regimes regardless of demography, but incorrectly estimated demography may dramatically reduce power to distinguish strong recessive purifying selection from weaker additive purifying selection or neutrality. This is evident across the range of demographies consistent with currently available population data. To address this problem, we propose an alternative to the widely used Poisson random field likelihood function. Our likelihood function relies on summary statistics that capture the features of the site frequency spectrum that are most sensitive to recessive selection, while remaining robust to specific parameters of recent demography since the out-of-Africa bottleneck. Using both our simulations and European samples from the Exome Aggregation Consortium (ExAC), we confirm that the inferences produced by applying this likelihood function are much more robust to demography.

POB-159

Inference of time-varying mutation rates in human populations

Jeffrey P. Spence[1], Yun S. Song[2, 3]

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

The human mutation rate has been the subject of much debate. Direct methods, such as counting *de novo* mutations in trios, have estimated the rate to be at least 25% slower than estimates based on population genetics theory. Because direct methods infer the mutation rate in present day humans, and population genetics-based methods infer a historical average of the mutation rate, one explanation for the discrepancy between the estimates is a slowdown in the per-generation mutation rate. To explore this hypothesis, we developed a coalescent-based method that allows accurate inference of time-varying mutation rates from sequence data from individuals sampled at the present. Using our method, we found that the per-generation mutation rate has been broadly constant for at least a million years, and we inferred a present day mutation rate in line with direct methods, casting doubt on other population genetics-based inferences. We inferred a 10% slower mutation rate for times older than about one hundred thousand years ago. We also explored temporal variation in the trinucleotide mutational spectrum, finding that some trinucleotide mutation rates have been variable throughout human history. In particular, we found evidence that a number of trinucleotide mutation have changed since the out-of-Africa event including additional evidence for a previously described increase in the rate of TCC->TTC mutations in Europeans.

POB-160

Uniparental Markers Show Sex-biased Admixture in Ancient Pacific Populations

Kathrin Naegele[1], Cosimo Posth[1], Frederique Valentin[2], Stuart Bedford[3, 4], Monica Tromp[1, 5], Jana Zech[1], Patrick Roberts[1], Johannes Moser[6], Julia Gresky[7], Russell Gray[1, 8], Adam Powell[1], Johannes Krause[1, 9, 10]

[1]Max Planck Institute (Germany), [2]CNRS (France), [3]Australian National University (Australia), [4]Vanuatu Cultural Centre (Vanuatu), [5]University of Otago (New Zealand), [6]German Archaeological Institute (Germany), [7]Department of Natural Sciences (Germany), [8]University of Auckland (New Zealand), [9]University of Tuebingen (Germany), [10]University of Tuebingen (Germany)

Analyses of uniparental markers have shown that present-day populations across Remote Oceania received mitochondrial DNA (mtDNA) largely of East Asian origin while the majority of Y-chromosome lineages derive from Near Oceania. Ancient DNA time transects across different regions of the Southwest Pacific offer the opportunity to elucidate the timing and formation of this genetic landscape shaped by sex-biased admixture. Here, we report uniparentally inherited mtDNA and Y-chromosomes of ancient individuals from Vanuatu, Tonga and French Polynesia in Remote Oceania, as well as from the Solomon Islands and Papua New Guinea in Near Oceania. We confirm that the first settlers of Remote Oceania associated with the Lapita culture and the spread of Austronesian languages approximately 3,000 years before present (yBP) carried mtDNA haplogroup B4a1a1, the so-called 'Polynesian motif'. We also present the first Y-chromosome haplogroup for a Lapita individual, O1a1a1a, further supporting the East Asian origins of the Austronesian expansion. We identify the arrival of Near Oceanic ancestry and signatures of sex-biased admixture as early as 2,500 yBP in Vanuatu. While autosomal DNA analyses of ancient individuals provide greater resolution in reconstructing population history, uniparental markers still offer valuable insights into admixture dynamics and especially where DNA preservation is not sufficient for genome-wide analyses.

POB-162

Phenotype and phylogeny of Neolithic Japanese hunter-gatherers, Jomon people, based on whole nuclear genome sequences

Hideaki Kanzawa-Kiriyama[1], Timothy Jinam[2], Yoshuke Kawai[3], Takehiro Sato[4], Kazuyoshi Hosomichi[4], Atsushi Tajima[4], Kryukov Kirill[5], Noboru Adachi[6], Naruya Saitou[2, 7], Ken-ichi Shinoda[1]

[1]National Museum of Nature and Science (Japan), [2]National Institute of Genetics (Japan), [3]University of Tokyo (Japan), [4]Kanzawa University (Japan), [5]Tokai University (Japan), [6]University of Yamanashi (Japan), [7]Graduate University fo Advanced Studies (SOKENDAI) (Japan)

Funadomari Jomon are 3,500-3,800 year-old northern Japanese hunter-gatherer. Here, we determined high depth and low depth nuclear genome sequences from Funadomari Jomon female (F23) and male (F5). F5 belongs to Y chromosome haplogroup D1b2b, which is rare haplogroup in modern Japanese. We genotyped the genome of F23, and HLA class-I type (homozygous for A*24:02:01, B*15:01:01, and C*03:03:01) and many phenotypic traits (e.g. non-shovel shape incisor and wet-type earwax) were determined. We observed high HBD similar to modern southern Native Americans, but long HBD tracts longer than 10cM is small, suggesting no close consanguineous mating but small population size in northern Jomon populations. MSMC analysis also suggested their small population size for more than 10,000 years. Phylogenetic relationship between F23 and modern/ancient Eurasians and Native Americans showed a deep divergence of F23 in East Eurasia, probably before the split of the ancestor of Native Americans from East Eurasians, but after the split of 40,000-year-old Tianyuan. This indicates that Northern Jomon people are genetically isolated from continental East Eurasians for long time. Intriguingly, we found that not only modern three Japanese (Ainu, mainland Japanese, and Ryukyuan) but also Ulchi, Koreans, the aboriginal Taiwanese and the Philippines are genetically closer to F23 compared to Han Chinese at the significant level ($Z>3$). These informations are one of a key to reconstruct ancient population structure and history in East Eurasia.

POB-163

Strong selection favoring Neanderthal-free segments on the X chromosome in out-of-Africa populations

Kasper Munch[1], Mikkel Heide Schierup[1], Laurits Skov[1], Moises Coll Marcia[1], Elise Lucotte[1]

[1]Aarhus University (Denmark)

In an analysis of the Simons genome diversity data set, we discover megabase long segments of the X-chromosome that have repeatedly lost most of the variation in out-of-Africa human populations. The loss of diversity is consistent with very strong selective sweeps occurring after the out-of-Africa expansion. The genomic positions of these sweeps overlap hotspots of selection in the great apes that may be subject to genomic conflict between the X and the Y chromosome (meiotic drive). Using a new approach to inferring individual Neanderthal-derived haplotypes, which do not rely on the Neanderthal reference genome, we discover that the swept regions are completely devoid of Neanderthal ancestry. The swept haplotypes are more divergent from present-day sub-Saharan African populations than the rest of the X chromosome. These findings prompt us to speculate that the swept regions are remnants of an initial out-of-Africa expansion that was subsequently replaced by a later expansion. However, the initial expansion may have contained strong meiotic drivers that resisted this genetic replacement.

POB-164

Origins of the Livestock in Madagascar

Takahiro Yonezawa[1], Yoshio Yamamoto[2], Felix Rakotondraparany[3], Masahide Nishibori[2]

[1]Tokyo University of Agriculture (Japan), [2]Hiroshima University (Japan), [3]University of Antananarivo (Madagascar)

Madagascar, known as an experimental field of evolution, is the fourth largest island located 400 km off the coast of East Africa, in the Indian Ocean. Malagasy people are thought to be dual origins (Austronesian people from Borneo and Bantu people from East coast of Africa) based on the genomic evidences. Accordingly, how, when, and from where, their material and non-material cultures came is very interesting issue.

Livestock are one of the most important innovation in human history, and have played important roles for the developments of the human societies in many aspects. Therefore, the domestication origins and the dispersal processes of livestock are one of the hot topics in evolution. However, although it is known that the most of the Malagasy cultivated plants were originated from Asian rather than from Africa, the origins of the Malagasy livestock are little known. Since Madagascar played important role in the Indian Ocean Trading Network, it also makes their origins as enigmatic issues. Here we carried out genetic analyses to elucidate the origins of the Malagasy livestock and we report the preliminary results of this study.

POB-165

Extremely rare variants reveal complex patterns of germline mutation in humans

Jedidiah Carlson[2], Jun Li[2, 3], Sebastian Zollner[1, 4]

[1]University of Michigan (United States), [2]University of Michigan (United States), [3]University of Michigan (United States), [4]University of Michigan (United States)

Precise understanding of the mutation process is essential to the study of human genome evolution. Rate heterogeneity and events that create multiple polymorphic sites substantially influence all population genetic inference. However, rate estimates using common variants are biased while analyzing de novo variants provides insufficient observations to consider sequence context. Here we use ~36 million singleton variants from 3,560 whole-genome sequences to characterize the heterogeneity of germline mutation rates across the genome. These singletons arose very recently in the population, and are thus largely unaffected by confounding evolutionary factors. We show that nucleotide context is the strongest predictor of mutability. Histone modifications, replication timing, recombination rate, and other local genomic features further modify mutability; overall, the mutation rate can vary by over 4 orders of magnitude. Moreover, we demonstrate that ~3% of all polymorphisms are the result of multi-nucleotide mutations. Several types of such events can be differentiated by the distance between polymorphisms and the type of mutation. Moreover, the type of the multi-nucleotide mutations determines its distribution across the genome, indicating that each types of multi-nucleotide mutations is the result of a distinct mutation process.

We evaluate the estimated models in an independent dataset of ~46,000 de novo mutations and show that singleton-based estimates provide a more accurate prediction of the mutation patterns than estimates based on common variants used in previous approaches. The effects of sequence contexts interacting with genomic features we present capture a refined portrait of germline mutation patterns in humans.

POB-166

Regulatory variation and selection in traditional populations of Southeast Asia

Katalina Sara Bobowik[1], Irene Gallego Romero[1], Murray Cox[2]

[1]University of Melbourne (Australia), [2]Massey University (New Zealand)

Regulatory variation and selection in traditional populations of Southeast Asia

Katalina Bobowik, Murray P. Cox, Irene Gallego Romero

The vast majority of human functional genomics studies to date have focused on large, outbred populations of primarily European descent. Given the diverse demographic history of our species, this fails to capture the full landscape of human regulatory variation, and has potential implications for genomic medicine and our understanding of genomic function. We have characterised regulatory variation in small, structured populations within islands of eastern Indonesia. We use whole genome, transcriptome, and DNA methylation data from nearly 100 individuals to identify loci that contribute to regulatory variation within and between islands, and identify differentially expressed genes that correlate with changes in environmental variables and genetic ancestry. This expands the repertoire of regulatory variation, particularly in small islands experiencing unique selective pressures, and adds to our database of human functional variation.

POB-169

Study of the North African human genetic landscape through the analysis of complete genome sequences

Gerard Serra-Vidal[1], David Comas[1]

[1]Universitat Pompeu Fabra (Spain)

The African continent is known be the place of origin for modern humans, which makes the understanding of its internal genetic variation especially relevant for reconstructing human complex demographic history. Nevertheless, the North of the continent has a somewhat independent history from the rest and has a highly complex demographic history due to its singular geographical location - between the Sahara desert, the Mediterranean Sea and the Middle East - which has historically favored a complex demographic history (including back migrations, gene flow, bottlenecks or genetic isolation) involving both autochthonous and neighboring populations.

Here, we study the recent history and relationships among 10 different North African populations, by analyzing individual whole-genome sequence data at deep coverage covering all major ethnogeographic groups within North Africa (from Western Sahara to Egypt, including Arab and Berber groups). We analysed this data and compared them to current and ancient Sub-Saharan, Middle Eastern and European groups to have them in a broad genomic context.

More than 10 million single nucleotide substitutions were found, providing an unprecedentedly rich picture of the genome diversity and population history in North Africa, which allows a deeper insight into complex demographic features, such as the evolution of effective population sizes, split times, migration rates within North African groups and between surrounding populations. North African samples were found to be genetically heterogeneous and show different patterns of admixture and demographic histories, though a clear correlation between samples differentiation and geographic origins or ethnicity could not be established.

POB-170

Predicting pathways to multiple drug resistance

Pamela Yeh[1]

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

Whether evolution occurs predictably and is repeatable, or occurs with significant stochastic components, has long been debated in evolutionary biology. In pathogen evolution, this debate also has important medical implications: If evolution is predictable when a bacteria is subjected to a specific drug environment, we would have information regarding how to combat the bacteria. If evolution is unpredictable, the types and amounts of drugs needed to treat an infection may also be unpredictable. Here, we ask whether the genetics underlying the mutant prevention concentration (MPC) for each drug-pathogen combination is predictable based on environmental selection pressures.
Using Staphylococcus epidermidis and eight antibiotics with different mechanisms of action, we examined the frequency of resistance for a wide range of concentrations per antibiotic, measured the repeatability of the MPC, and examined cross-resistance and cross-sensitivity. We conducted whole genome sequencing to determine the genetic changes underlying the differences in MPCs. We found that the MPC varied widely over replicates and patterns of cross-resistance and cross-sensitivity were complex, with different mutations responsible for the different MPC values. The results suggest that, there is a significant stochastic element in the evolution of MPC traits.

POB-172

H3N2 influenza virus fitness prediction

Ksenia Safina[1, 2], Pavel Dvurechensky[2, 3], Alexey Neverov[4], Alexander Gasnikov[2, 5], Alexander Favorov[6], Georgii Bazykin[1, 2]

[1] Skolkovo Institute of Science and Technology (Russian Federation), [2] Institute for Information Transmission Problems RAS (Russian Federation), [3]Weierstrass Institute for Applied Analysis and Stochastics (Germany), [4] Central Research Institute for Epidemiology (Russian Federation), [5]Moscow Institute of Physics and Technology (Russian Federation), [6]Johns Hopkins University School of Medicine (United States)

H3N2 influenza virus is a major cause of seasonal flu which kills up to half a million people annually. Influenza vaccines need to be updated regularly because of the fast evolution of viral surface antigens (neuraminidase and hemagglutinin) driven by collective immunity selection pressure. Each year, the World Health Organization recommends three influenza strains (one for each of subtypes A/H1N1, A/H3N2, and B/ influenza) as vaccine candidates that should be antigenically close to the strains that are expected to dominate in the upcoming season. The effectiveness of vaccines is poor if WHO picks up wrong strains.

The strain selection procedure may be improved by modeling influenza evolution. We develop a new model that predicts seasonal population dynamics of influenza clades on H3N2 phylogenetic tree accounting for pairwise epistasis among sites in hemagglutinin.

POB-173

A "molecular tango" of sperm-egg recognition proteins in Pacific abalone

Damien B Wilburn[1], Joshua G Schraiber[2], Willie J Swanson[1]

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

A hallmark of interacting reproductive proteins in males and females is rapid evolution. Such accelerated evolution is thought to arise from sexual selection driving continuous coevolution between the interacting protein partners. Classically, sexual selection has been used to describe the evolution of macroscopic phenotypes such as coloration or ornamentation, yet its role in shaping molecular traits has been minimally examined. Here we describe a "molecular tango" model of molecular coevolution between interacting egg-sperm proteins in the marine gastropod abalone. As one of the few biological systems where both the male and female proteins have been biochemically characterized, the abalone system provides a unique opportunity to dissect the role of sexual selection in their molecular coevolution. The sperm protein lysin evolves extraordinary rapidly, which is contrast to its egg receptor VERL that acquires non-synonymous mutations approximately five times more slowly. To characterize lysin-VERL coevolution, we leveraged knowledge of abalone natural history and the biochemistry of lysin-VERL interactions to develop a novel model of coadaptation. Based on the Gillespie-Orr framework of modeling adaptive walks, we argue that lysin-VERL coevolve via a "molecular tango": VERL evolves by genetic drift, followed by lysin acquiring a sequence of adaptive mutations to restore binding affinity. We test this model using *in silico* protein docking simulations and stochastic mapping with data from 7 Pacific abalone species.

POB-174

Origin and evolution of ORFan genes in Lactobacillus kunkeei

Karl David Dyrhage[1], Andrea Garcia Montaner[1], Christian Seeger[1], Siv Andersson[1]

[1]Uppsala University (Sweden)

With the ever-increasing amount of genomic data available, it has become evident that the genomes of most newly sequenced organisms contain relatively large numbers of open reading frames (ORFs) that bear no similarity to genes or protein domains in other taxa. Apart from artifacts from faulty gene prediction, these hypothetical genes, or ORFans, could arise through several mechanisms: 1) *de novo* emergence from non-coding DNA, 2) rapid evolution following gene duplication and diversification or gene decay, or 3) horizontal gene transfer from previously unobserved organisms. Currently, the rates at which these appear, and subsequently how they evolve and spread through the population, is poorly understood.

Here, we use bioinformatic and experimental methods to detect and investigate the evolutionary origins of ORFans in the honey bee symbiont *Lactobacillus kunkeei*. A comparative genomic study of 43 strains revealed over 200 ORFans restricted to *L. kunkeei*, about 100 of which are strain specific. Our analysis has shown that several ORFans are the result of short duplications followed by gene fractionation. To estimate how many of these ORFans are functional, we study the expression patterns, fold properties and structures of the ORFan proteins. Many ORFans cluster near the origin of replication and we discuss putative functions of these ORFans along with possible mechanisms that could account for such a biased pattern of gene birth.

POB-177

Statistical tests for genomic time series data

Andreas Futschik[1]

[1]JKU University Linz (Austria)

Genomic time series data arise for instance in experimental evolution, where several populations of fast reproducing organisms, such as yeast or Drosophila, are kept under different environmental conditions. The goal is to find genomic signatures of adaptive selection. Recent experiments revealed however, that different experimental populations do not necessarily show the same genomic signatures of adaptation. This may be for instance due to the random loss of beneficial variants due to genetic drift, due to initial linkage with deleterious alleles, or to different possibilities for adaptation. We investigate different possibilities for identifying selection under such circumstances, and propose new more powerful tests of a global null hypothesis. In part this work relies on Futschik, Taus & Zehetmayer (2017)

POB-179

Effect of binding interference on the divergence of paralogous genes that encode homodimers

Angel Fernando Cisneros[1, 2], Christian Landry[1, 2, 3]

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

Protein complex formation is a key process for cell biology as many proteins need to physically associate to carry out their functions. Interaction among proteins is therefore an important component of their evolution. We are interested in the evolution of these interactions following gene duplication. Of particular interest are paralogs that derive from the duplication of self-interacting proteins because those give rise to heterodimers of paralogs in addition to two self-interacting paralogs. My objective is to evaluate the effect of the competition between heterodimers and homodimers on their sequence divergence.

Such competition introduces a molecular conflict that can lead to several different scenarios, depending on which type of complexes is favored by natural selection over time. I am studying these scenarios and their consequences on the divergence of paralogs by simulating the evolution of protein complexes using crystal structures, and the estimation of binding energies, protein stability, and specific fitness functions. Because a large fraction of genes encode self-interacting proteins, these molecular conflicts may be an important factor in the divergence of paralogous proteins and the evolution of protein interaction networks.

POB-180

Genetic framework for brain and cranium development at the transcriptional level

Derek Caetano-Anolles[1]

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

Exploring and identifying the genetic components that direct the structure of organisms has remained a major focus of developmental and evolutionary biology. Morphology can evolve very rapidly in populations, which can lead to novel adaptations in response to environmental changes. This research focuses on identifying genes responsible for developing the morphological features of the skull and mandible of the mouse (Mus musculus), while comparing these morphological effects to those found in rat (Rattus norvegicus). Previous Quantitative Trait Locus (QTL) studies have advanced the identification of the genetic basis for craniofacial shape in mice have shown that the phenotype is controlled by genes in several genomic regions. Recently, candidate genes involved in naturally occurring craniofacial shape variation were identified in natural hybrid mice using genome-wide association studies (GWAS). As these were natural hybrids obtained from wild caught mice, the genes identified could therefore be the genes being acted upon by natural selection and affecting craniofacial shape evolution, leading to our use of knockout mice for these candidate genes are being used to confirm their genetic effects on mouse facial development.

POB-181

RNA-seq sex-specific mapping protocol increases expression value on the sex chromosomes in both genetic male and female samples

Sarah Brotman[1], Kimberly Olney[1], Valeria Valverde-Vesling[1], Jocelyn Andrews[1], Melissa Wilson Sayres[1, 2]

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

Different read aligners, reference genomes, and differential expression programs can impact the assessment of differential expression in RNA-seq studies. For this reason we tested several mapping scenarios in human RNA-seq tissue samples from the brain cortex, lung, and whole blood of 5 genetic female (46, XX) and 5 genetic male (46, XY) samples to determine how differential gene expression patterns change depending on the reference genome used for alignment, the choice of read aligner, and the differential expression software. We implemented a mapping protocol where we mapped all individuals using two different read aligners, to the GRCh38.p10 human reference genome, which includes the full sequence of both the X and Y chromosomes, and computed differential expression between the set of genetic male (46, XY) and genetic female (46, XX) samples between each tissue type using several differential expression programs. We then mapped the genetic female samples (46, XX) to the human reference genome GRCh38.p10 with the Y chromosome hard-masked (Y-masked) and the genetic male samples (46, XY) to the human reference genome (including the Y chromosome), but with the pseudoautosomal regions of the Y chromosome hard-masked (YPARs-masked) creating two sex-specific mappings, Y-masked and YPARs-masked. We investigated expression values in all mapping and differential expression scenarios and show that regardless of choice of read aligner or differential expression program, using a sex-specific mapping protocol showed higher expression values on the sex-chromosomes in both genetic male and genetic female across all tissue types.

POB-182

Deciphering genetic basis and copulatory effect underlying the rapid diversification of male genitalia between sibling species of Drosophila melanogaster

Kentaro M. Tanaka[1], Yoshitaka Kamimura[2], Aya Takahashi[1, 3]

[1]Tokyo Metropolitan University (Japan), [2]Keio University (Japan), [3]Tokyo Metropolitan University (Japan)

Among animals with internal fertilization, genital morphology, especially in male, exhibits rapid diversification even in recently diverged species, and plays an important role in reproductive isolation. To expand our knowledge on sexual isolation by genital divergence, it is important to understand the genetic basis and copulatory effects of the morphological changes. To address these questions, we used a young species pair, *Drosophila simulans* and *D. mauritiana*, which have distinctive male genital structures but still able to produce fertile hybrid females in a laboratory condition. Taking advantage of the fertilility, we generated a series of introgression lines that carry different 6-15 Mb *D. mauritiana* segments in a *D. simulans* genomic background. Observation of genital structures indicated that males of one introgression line possessed significantly larger anal plates than pure *D. simulans*. This enlargement explained about 45% of the interspecific difference. Also, a mass of ejaculate leakage on abdominal body surface was frequently detected after copulation in pairs with the introgression male. Further fine-scale mapping by recombinant males generated from this introgression line allowed us to narrow down the causative region to 17 kb, from where there was no detectable transcript obtained by RNA-seq during genital morphogenesis. These results suggest that regulatory modification of a gene outside the introgressed region has contributed to the male genital divergence and has caused mechanical incompatibility in male and female genital coupling.

POB-183

Differentiation in the facial morphology between Ryukyuans and mainland Japanese

Ryosuke Kimura[1], Chiaki Watanabe[1], Eri Miyazato[1], Kyoko Yamaguchi[2], Takehiro Sato[3], Tsuyoshi Ito[4], Akira Kawaguchi[1], Ken Yamamoto[5], Hajime Ishida[1]

[1]University of the Ryukyus (Japan), [2]Liverpool John Moore's University (United Kingdom), [3]Kanazawa University (Japan), [4]Kyoto University (Japan), [5]Kurume University (Japan)

The cause of the phenotypic differentiation in facial features among modern human populations -genetic drift or natural/sexual selection- is still under debate. To answer this question, we performed a neutrality test using the F_{ST}-Q_{ST} comparison method on the facial difference between Ryukyuans and mainland Japanese. Using a handheld 3D scanner, we obtained facial morphology data from 734 adults living in Okinawa. After procedures of landmarking (23 points) and homologous modeling using 2,596 corresponding points, we performed principal component analysis (PCA), in which top 17 PCs contributed to 90% of the morphological variation. We considered these PCs as quantitative traits and performed subsequent analyses. To calculate Q_{ST} values for PCs, we only used individuals whose four grandparents had originated all from Okinawa or all from mainland Japan. We also obtained genome-wide SNP genotype data from the same individuals and estimated the F_{ST} value. We found that the mean Q_{ST} value in the 17 PCs were significantly higher than the F_{ST} value, which indicates that the facial differentiation between Ryukyuans and mainland Japanese cannot be explained only by genetic drift and that the presence of some selective pressure is expected.

POB-185

Genome-wide allele-specific expression of the transcriptional landscape associated to Capsicum domestication

Erik Diaz-Valenzuela[1], Angelica Cibrian-Jaramillo[1], Ruairidh Sawers[2]

[1]Centro De Investigacion Y De Estudios Avanzados Del Instituto Politecnico Nacional (Mexico), [2]Centro De Investigacion Y De Estudios Avanzados Del Instituto Politecnico Nacional (Mexico)

Plants have evolved an ample diversity of morphologies that correlate with changes in gene expression driven by the rewiring of gene regulatory networks (GRN), rather than changes in the coding sequences of the genes. GRNs control morphology by means of interactions between *cis* (promoters) and *trans* (transcription factors) regulatory sequences. Our main hypothesis is that genetic variation associated to morphological evolution in plant domestication should also be contained in regulatory sequences. We utilized *Capsicum* fruits (chilli pepper, Solanaceae) to investigate the transcriptional changes derived from its domestication, and its underlying genetics. Genome-wide transcriptional divergence between wild (W) and cultivated (C) Capsicum revealed that 51% of the orthologs between C and W show differential expression with a bias towards the up-regulation of C, and that differentially expressed genes (DEG) include reproduction and fruit development processes. An Allele Specific Expression analysis (ASE) in F1 hybrids showed that 66% of the DEG in the fruit was driven by trans-acting variation and 44% by cis-acting variation. In contrast to most studies in domesticated plants, we also discovered that recessive variation underlays most of the trans-regulatory changes, while cis-regulatory changes are mostly due to additive variation. It is possible that these patterns relate to how Capsicum has been cultivated for thousands of years, at least in its center of domestication. Overall, we provide insights to the overall understanding of genome evolution and the nature of genetic interactions that take place during plant domestication.

POB-186

The genetic architecture of the divergent ecotypes of Littorina saxatilis

Pragya Chaube[1], Anja Westram[1], Roger Butlin[1]

[1]University of Sheffield (United Kingdom)

Hybrid zones present an excellent opportunity to study diverging populations and examine the traits which are under selection and contribute towards reproductive isolation. In the current study, we exploited this feature of hybrid zones to study the rough periwinkle, Littorina saxatilis. Littorina saxatilis is a well-characterized species to study adaptation and speciation. It is suggested that the distinct ecotypes of Littorina saxatilis have evolved at different geographical locations from a local population under similar divergent selection pressures. In order to gain an insight into the genetic architecture of the adaptive phenotypes, we performed NGs and association mapping in the hybrid zones.

POB-187

Transcriptome analysis of the reef-building octocoral, Heliopora coerulea

Christine Guzman[1, 2], Chuya Shinzato[3], Tsai-Ming Lu[4], Cecilia Conaco[1]

[1]University of the Philippines (Philippines), [2]Okinawa Institute of Science and Technology Graduate University (Japan), [3]The Universtity of Tokyo (Japan), [4]Okinawa Institute of Science and Technology Graduate University (Japan)

The blue coral, *Heliopora coerulea*, is a reef-building octocoral that prefers shallow water and exhibits optimal growth at a temperature close to that which causes bleaching in scleractinian corals. To better understand molecular mechanisms that may contribute to its temperature tolerance and its reef-building capacity, we generated a reference transcriptome for *H. coerulea* using next-generation sequencing. We successfully distinguished genes originating from the host coral and its symbionts by aligning sequences to cnidarian and *Symbiodinium* genomes, respectively. Analysis of the blue coral transcriptome revealed enrichment of genes involved in stress response, including heat-shock proteins and antioxidants, as well as genes participating in signal transduction and stimulus response. Its tissues host *Symbiodinium* related to the thermotolerant C3-Gulf ITS2 type, known from the Persian Gulf. These features suggest a basis for the reported tolerance of H. coerulea to elevated temperature. Furthermore, the blue coral possesses homologs of biomineralization genes found in other corals and may use a biomineralization strategy similar to that of scleractinians to build its massive aragonite skeleton. These findings thus offer insights into the ecology of *H. coerulea* and suggest gene networks that may govern its interactions with its environment.

POB-189

Comparative avian genomics: Identifying the genetic basis of beak shape variation.

Leeban Yusuf[1], Toni Gossmann[1], Chris Cooney[1]

[1]University of Sheffield (United Kingdom)

Comprising over 10,000 species, birds are among the most speciose and diverse taxonomic groups. A recent morphometric analysis by Cooney et al. (2017) from over 2000 bird species found that large variation exists in beak shape evolution across avian taxa. As an ecologically-important trait, there have been many recent attempts to understand the genetic basis of beak shape evolution in a microevolutionary context, but how the large morphological variation that exists across distant taxa can be explained remains unanswered. Here, we establish the link between molecular evolution and the disjunction in beak shape evolution in different lineages observed in these morphometric analyses. For this, we combine genomic data with morphometric data by Cooney et al. (2017) to identify putatively adaptive genes linked to beak shape evolution. We focus on coding regions and examine orthologous sequences from over 60 fully-sequenced bird genomes. We calculated rates of evolutionary change at coding sites using the nonsynonyomous to synonymous rate ratio. Our study combines comprehensive morphometric data with genetic data from a range of avian taxa to determine genes that explain morphological shifts in beak shape.

POB-190

Is gene transfer a source of ecological adaptation in eukaryotes? The case of nitrate metabolism

Eduard Ocana-Pallares[1], Sebastian Najle[1, 2], Claudio Scazzocchio[3, 4], Inaki Ruiz-Trillo[1, 5, 6]

[1]Institut de Biologia Evolutiva (CSIC-UPF) (Spain), [2]Instituto de Biologia Molecular y Celular de Rosario (IBR) CONICET and Facultad de Ciencias Bioquimicas y Farmaceuticas, Universidad Nacional de Rosario (Argentina), [3]Imperial College London (United Kingdom), [4]Institute for Integrative Biology of the Cell (I2BC) (France), [5]Universitat de Barcelona (Spain), [6]Institucio Catalana de Recerca i Estudis Avancats (ICREA) (Spain)

Nitrogen is an essential element for life. However, while it constitutes about 80% of the atmosphere, only a handful of organisms can fix it directly. Therefore, most taxa require alternative nitrogen sources that are mostly produced through biological reactions. Because nitrate is the most abundant source in the oceans and in many terrestrial environments, characterizing its metabolism is important to address the role of a given organism in an ecosystem. We here assessed the distribution of the genes involved in nitrate assimilation (NAPs) in eukaryotes. We found that this metabolic pathway is completely absent in phagotrophic lineages while is strongly associated with autotrophy and osmotrophy. The phylogenetic signal indicate multiple transfers of NAPs between osmotrophic and autotrophic lineages, the last possibly related with the endosymbiotic events that led to the origin of complex plastids. These results suggest that gene transfer could be an important source of adaptive ecological traits also for eukaryotes. It is also relevant the finding of NAPs in two recently sequenced ichthyosporean species which are able to grow in a nitrate minimal medium. Interestingly, they present a novel nitrate reductase that replaced the canonical C-ter region for the N-ter region of the nitrite reductase. While most of ichthyosporeans are described as parasites, the metabolic complexity of these two species agrees with environmental data suggesting that non-parasitic members may exist. As ichthyosporeans and animals are closely related in the eukaryotic tree, these findings have implications for the potential living styles of the animal ancestors.

POB-191

Photosensitive alternative splicing of the circadian clock gene timeless is population specific in a cold-adapted fly, Drosophila montana

Riikka Tapanainen[1], Darren Parker J[2], Maaria Kankare[1]

[1]University of Jyvaskyla (Finland), [2]University of Lausanne (Switzerland)

To function properly, organisms must adjust their physiology, behavior and metabolism in response to a suite of varying environmental conditions. One of the central regulators of these changes is organisms internal circadian clock, and recent evidence has suggested that the clock genes are also important in the regulation of seasonal adjustments. In particular, thermosensitive splicing of the core clock gene timeless in a cosmopolitan fly, *Drosophila melanogaster*, has implicated this gene to be involved in thermal adaptation. To further investigate this link we examined the splicing of timeless in a northern malt fly species, *Drosophila montana*, which can withstand much colder climatic conditions than its southern relatives. We studied northern and southern populations from two different continents (North America and Europe) to find out whether and how the splicing of this gene varies in response to different temperatures and day lengths. Interestingly, we found that the expression of timeless splice variants was sensitive to differences in light conditions, and while flies of all study populations showed a change in the usage of splice variants in constant light compared to LD 22:2, the direction of the shift varied between populations. Overall, our findings suggest that the splicing of timeless in northern *D. montana* flies is photosensitive, rather than thermosensitive and highlights the value of studying multiple species and populations in order to gain perspective on the generality of gene function changes in different kinds of environmental conditions.

POB-194

The genomic basis for convergent evolution of carnivory in shrew rats

Emily J Roycroft[1, 2], Anang S Achmadi[3], Jacob A Esselstyn[4, 5], Jeffrey M Good[6], Adnan Moussalli[1], Kevin C Rowe[1]

[1]Museums Victoria (Australia), [2]The University of Melbourne (Australia), [3]Research Center for Biology (Indonesia), [4]Louisiana State University (United States), [5]Louisiana State University (United States), [6]University of Montana (United States)

Comparative genomic studies across distantly related taxa have increasingly suggested that convergent molecular changes may underlie phenotypic convergence. Understanding to what extent phenotypic convergence shares a common genomic signature, and to what extent this is meaningful, allows us to gain insight into possible constraints on selection at the molecular level.

The ecological state of exclusive carnivory has evolved on at least four separate occasions in the murine rodents of the Indo-Australian Archipelago. Of these, the "shrew rats" of Sulawesi and the Philippines are a remarkable case of phenotypic convergence. Despite being approximately 12 million years divergent, these independent carnivorous clades have evolved a number of striking ecomorphological similarities (e.g. skull shape, dentition and digestive tract). In addition to shared features associated with carnivory, within these clades, the genera *Paucidentomys*(Sulawesi) and *Rhynchomys* (Philippines) have converged on a diet of exclusively earthworms, and both exhibit highly modified craniofacial features (elongated rostrum and reduced/absent molars) associated with "sucking" worms.

Here, we investigate the genomic basis of convergent evolution in these strikingly convergent carnivorous ecomorphs. Via sequencing whole exomes (~20,000 protein coding genes) from representative species across the murine tree, we screen for convergent molecular selection in the carnivorous shrew rats of Sulawesi and the Philippines, and discuss the implications of our results in terms of how ecological opportunity can drive evolution at the molecular level.

POB-195

Evolution of sperm proteins and non-fertilizing sperm in two Lepidoptera with different mating systems

Andrew Joseph Mongue[1], James R. Walters[1]

[1]University of Kansas (United States)

Sperm dimorphism is a relatively common phenomenon in the animal kingdom, but perhaps nowhere is it more striking than in Lepidoptera (butterflies and moths). Males of this order make both fertilizing (nucleated) sperm and a second sperm type (anucleated) that lacks any nuclear DNA. Production of both types is under hormonal control and ultimately both types are transferred to the female during mating. Indeed anucleated sperm outnumber nucleated sperm by 10:1 or more in the ejaculate, suggesting some functional role for this non-fertilizing sperm; yet its evolutionary significance remains unclear. One popular hypothesis posits that auncleated sperm are agents of sperm competition, in some way hindering the reproductive success of rival males. Here we investigate the role of auncleated sperm and the selective strength of sperm competition on proteins in both sperm types in a monandrous moth (Manduca sexta) and a highly polyandrous butterfly (Danaus plexippus). Using genes identified by targeted proteomics of sperm cells and whole genome resequencing of wild individuals, we estimate the rates of adaptation of sperm proteins in both the presence and absence of sperm competition in natural populations. These estimates narrowly give us a better understanding of this unique reproductive system and broadly provide insight into the role of adaptation vs. drift in divergence of reproductive proteins.

POB-198

Ancient and modern evolution of opsins in *Daphnia*.

Jeffry L. Dudycha[1], Matthew J. Greenwold[1], Christopher S. Brandon[2]

[1]University of South Carolina (United States), [2]Florida State College (United States)

Vision allows animals to acquire information about their environment, and is critically dependent on opsins to capture light. The genome of *Daphnia pulex* has a larger opsin gene family than any other animal. We compared the *D. pulex* and *D. magna* genomes to infer the complement of opsins present in the ancestral *Daphnia*, which dates back to the Mesozoic Era, and then trace the history of gene duplication and loss in the two lineages. We inferred a total of 29 opsins in the *Daphnia* progenitor, of which 13 were likely visual opsins. Nearly all of the ancient opsins have one or more descendent genes in modern *Daphnia*. This is considerably more than necessary to explain the tetrachromatic vision of *Daphnia*, but their long persistence suggests some kind of selective advantage for maintaining a diverse family of opsins. We then used a population genomic sample of >80 clones of *D. pulex* from a single population to construct a dataset of sequence variation of opsins within a population. We applied selection analyses to evaulate the degree of selection on each opsin and determine A) whether each descendant of an ancient opsin was under similar selection and B) whether a single opsin for each wavelength class (UV, blue, green, red) was under particularly strong selection relative to other members of its class. We further evaluated whether sites known to be associated with spectral sensitivity were under stronger selection than other parts of the genes.

POB-199

Gene expression variation and diapause regulation in wood white (*Leptidea*) butterflies.

Niclas Backstrom[1], Luis Leal[1], Venkat Talla[1], Thomas Kallman[2], Magne Friberg[3], Christer Wiklund[4], Vlad Dinca[5], Roger Vila[6]

[1]Uppsala University (Sweden), [2]Uppsala University (Sweden), [3]Lund University (Sweden), [4]Stockholm University (Sweden), [5]University of Oulu (Finland), [6]Institut de Biologia Evolutiva (CSIC-UPF) (Spain)

In temperate latitudes, many insects enter diapause (dormancy) during the cold season, a period during which developmental processes come to a standstill. The wood white (*Leptidea sinapis*) is a butterfly species distributed across western Eurasia that shows photoperiod-induced diapause with variation in critical day-length across populations at different latitudes. We assembled transcriptomes and estimated gene expression levels at different developmental stages in experimentally induced directly developing and diapausing cohorts of a single Swedish population of *L. sinapis* to investigate the regulatory mechanisms underpinning diapause initiation. Different day lengths resulted in expression changes of developmental genes and affected the rate of accumulation of signal molecules, suggesting that diapause induction might be controlled by increased activity of monoamine neurotransmitters in larvae reared under short-day light conditions. Expression differences between light treatment groups of two monoamine regulator genes (*DDC* and *ST*) were observed already in instar III larvae. Once developmental pathways were irreversibly set at instar V, a handful of genes related to dopamine production were differentially expressed leading to a significant decrease in expression of global metabolic genes and increase in expression of genes related to fatty acid synthesis and sequestration. This is in line with a time-dependent (hour-glass) model of diapause regulation where a gradual shift in the concentration of monoamine neurotransmitters and their metabolites during development of larvae under short-day conditions leads to increased storage of fat, decreased energy expenditures, and ultimately developmental stasis at the pupal stage.

POB-200

Tissue- and temperature-specific RNAseq differential gene expression and SNP variation by sex, geography, and linkage reveal adaptive potential and population genomic health of marine turtles

Andrew M Shedlock[1, 2, 3]

[1]College of Charleston (United States), [2]Medical University of South Carolina (United States), [3]Hollings Marine Laboratory (United States)

Many turtle species have now become threatened by local habitat loss, pollution, and excessive mortality due to human activities. Despite this we know little about the genomic basis for many of their unique adaptations, especially those species which have developed pelagic marine life histories and astonishing natal beach homing capabilities, exemplified by the loggerhead turtle, *Caretta caretta*, a species of ecological and economic importance in both Japan and the US. Presently our lab has identified 402,258 transcripts in more than 790 million Illumina cDNA sequence reads that have revealed 13,867 annotated genes, about 7% of which are differentially expressed above and below the pivotal temperature of 29C for *C. caretta* environmental sex determination. These candidate genes have been analyzed for tissue-specific co-expression and functional pathway clustering from 12 *de novo* transcriptome assemblies of sea turtle hatchlings collected legally in the wild and raised experimentally in the laboratory. Additionally, over 125 million Illumina reads from ApeKI-indexed genomic libraries were compiled for 48 population genomic samples collected from SE US Coastal Atlantic waters. Using reference-based SNP calling and varying parameters in independent runs of *Stacks*, 2.79 million SNPs and up to 15 potentially sex-specific loci were identified, with 0.127% of loci in linkage disequilibrium, indicating regions of potential natural selection and adaptive evolution. RNAseq results are being integrated with reference-based SNP variance due to sex, geography, and linkage, thereby moving the field toward a more genomics-enabled predictive approach to environmental management of threatened turtle populations worldwide.

POB-201

Transcriptome sequencing reveals genomic adaptations to life at high elevations in Ethiopian frogs

Jacobo Reyes-Velasco[1], Yann Bourgeois[1], Stephane Boissinot[1]

[1]New York University Abu Dhabi (United Arab Emirates)

Understanding the genomic changes that occur when organisms adapt to new environments is a fundamental question in evolutionary biology. High elevations can be particularly challenging for organisms, as temperature, UV radiation and low oxygen levels can be difficult to overcome, especially for slow moving poikilotherms such as amphibians. Despite these adversities, several species of amphibians have adapted to extreme environments at elevations above 3,000m in the mountains of Central Asia, South America, and East Africa.

In Ethiopia, some populations of the frog Leptopelis gramineus occur at elevations of more than 3,500m and span more than 1,500m in their elevation range. This tolerance to multiple environmental conditions at different elevations makes it a good model to study which genomic changes that underlie local adaptation in populations of this species. In this study, we sequenced the transcriptome of several individuals of L. gramineus from three different elevations. Our goal was to compare genes that might be up or down regulated between populations from different elevations.

We looked for loci that might be under selection, particularly those related to functions such as oxygen intake, UV resistance, and cell to cell binding, which are known to be important in other vertebrates that live at high elevations.

Our study sheds light into the genomic changes that have occurred in amphibians in order to survive at high elevations. This research is of particular importance as climate change is threatening the survival of amphibians living in montane areas across the world.

POB-203

Local adaptation under gene flow: Recombination, conditional neutrality and genetic trade-offs shape genomic patterns in Arabidopsis lyrata

Tuomas Hamala[1], Outi Savolainen[1]

[1]University of Oulu (Finland)

Short-scale local adaptation is a complex process involving selection, migration and drift. The expected effects on the genome are well grounded in theory, but to examine these on an empirical level has proven difficult, as it requires information about local selection, demographic history and recombination rate variation. Here, we use locally adapted and phenotypically differentiated *Arabidopsis lyrata* populations from two altitudinal gradients in Norway to test these expectations on a whole-genome level. Demography modelling indicated that populations within the gradients diverged less than 2 kya and that the sites are connected by gene flow. The gene flow estimates were, however, highly asymmetric with migration from high to low altitudes being several times more frequent than vice versa. To detect signatures of selection for local adaptation, we estimated patterns of lineage specific differentiation among these populations. Concordant with theory, adaptive loci in the low-altitude populations were concentrated in areas of lower recombination rates than in the high-altitude populations. Although most selected loci displayed patterns of conditional neutrality, we found indications of antagonistic pleiotropy, with one locus particularly showing high divergence and signs of selection in both populations. Our results further suggest that resistance to solar radiation is an important adaptation to alpine environments, while vegetative growth and bacterial defense were indicated as selected traits in the lowland habitats. These results provide insights into genomic architectures and evolutionary processes driving local adaptation under gene flow. We also contribute to understanding of traits and biological processes underlying alpine adaptation in northern latitudes.

POB-204

Uncovering the demographic history of endangered Formosan fruit bats (Pteropus dasymallus formosus) in Taiwan

Kung-Ping Lin[1], Yun-Hua Lo[1], Shiang-Fan Chen[2], Masako Izawa[3], Wen-Ya Ko[1]

[1]National Yang-Ming University (Taiwan), [2]National Taipei University (Taiwan), [3]University of the Ryukyus (Japan)

Formosan fruit bat (*Pteropus dasymallus formosus*) is one of the five subspecies of Ryukyu fruit bats which distributes on the West Pacific islands stretching from Ryukyu to Taiwan and Philippines. Formosan fruit bats were once abundant in Taiwan. However, the population had suffered from habitat loss and extreme hunting in the past, likely causing a dramatic population decline. Now the population is on the blink of extinction. However, little is known about the genetic diversity and levels of population inbreeding of this endangered subspecies as well as the degree of gene flows with other subspecies dwelling on neighboring islands. Here, we studied genetic diversity and demographic history of *P. d. formosus* from four individuals. We applied restriction-site-associated DNA sequencing technology (RADSeq) to uncover novel genetic variants across nuclear genomes. As a result, we have successfully obtained 10,226 assembled scaffolds/contigs that can be aligned unambiguously between these individuals and identified 8,836 SNPs (θ_W=0.0025). We plotted the folded allele frequency spectrum (f-AFS) for minor alleles and identified an excess of intermediate alleles in comparison with the simulated f-AFS under the model of neutral evolution of constant population size. The f-AFS becomes even more biased toward intermediate levels after excluding one specimen smuggled into Taiwan illegally. Our results have revealed that the *P. d. formosus* population might have been declining severely. Future work will focus on studying the levels of population differentiation and gene flows between *P. d. formosus* and its sister subspecies inhabiting on the Ryukyu islands.

POB-205

The genomics of local adaptation in the alpine Brassicaceae Arabis alpina

Aude Rogivue[1], Kevin Leempoel[2], Rimjhim Roy Choudhury[3], Stefan Zoller[4], Stephane Cretegny[5], Francois Felber[6], Michel Kasser[5], Christian Parisod[3], Stephane Joost[7], Felix Gugerli[1]

[1]WSL Swiss Federal Research Institute (Switzerland), [2]Stanford University (United States), [3]University of Bern (Switzerland), [4]ETH Zuerich (Switzerland), [5]HEIG-VD (Switzerland), [6]University of Neuchatel (Switzerland), [7]EPF Lausanne (Switzerland)

Fitness differences among individuals, for example due to changing environment, are mostly genetically determined. It is therefore important to study the genomic processes driving local adaptation of individuals to heterogeneous and temporally changing environmental conditions. Here we investigate the extent of adaptive variation that can be found in genic and non-genic regions of the genome. Three hundred and four individuals of the Alpine rock cress Arabis alpina, an emerging model plant for ecological genomics, were sampled in contrasting environments in four regions of the western Swiss Alps. We performed association analyses of environmental factors derived from high-resolution LIDAR data and measured in-situ with 135,000 SNPs and 3,800 transposable elements present near genic regions, identified through whole-genome re-sequencing. Latent factor mixed models indicate that different types of genomic polymorphisms show signatures of local adaptation, after controlling for population structure. Few SNPs and transposable element are significantly associated with environmental factors that are proxies for water flow, snow movement, erosion and solar radiation. We will present genes that are involved in local adaptation and how they interact to control traits under selection, by considering analyses targeted at signals of polygenic adaptation. In addition, multiscale sampling of individuals suggests that the signature of selection is population-specific in highly heterogeneous and topographically structured alpine environment with limited gene flow, supporting the notion that local adaptation remains local.

POB-206

THE IMPACT OF INCREASED SANDSTORM ACTIVITIES ON THE RED SEA MICROBIOTA

Hayedeh Behzad[1], Hajime Ohyanagi[1], Badr Alharbi[2], katsuhiko Mineta[1], Takashi Gojobori[1]

[1]King Abdullah University of Science and Technology (KAUST) (Saudi Arabia), [2]King Abdulaziz City for Science and Technology (Saudi Arabia)

Large areas of lands around the world are faced with increased desertification and dust activities, in part due to global climate changes. Dusts are important sources of essential nutrients and trace metals required for growth and propagation of microorganisms in marine environments. From the perspective of ecological and evolutionary assessments, it is important to examine whether increased dust deposit due to enhanced sandstorm activities could significantly impact microbial community dynamics in marine ecosystems. We, thus, examined the effect of sandstorms on the Red Sea microbiota, since the Red Sea is a relatively isolated oligotrophic seawater region with no river inflow, minimal precipitation, and in close proximity to large deserts. Metagenomics was used to investigate potential changes in microbial diversity in the Red Sea in response to three major sandstorms originating from the neighboring Syrian (Sep 2015), Nubian (July 2016), and Libyan (March 2017) Deserts. Phylogenetic analysis of the 16S rRNA using QIIME demonstrated that despite differences in dust origin and time, all three sandstorms consistently changed the relative abundance of reads corresponding to similar sets of autotrophic and heterotrophic microorganisms, with a significant increase in the phylum Bacteroidetes and a sharp decline in cyanobacteria. Our data supports the notion that the predicted increase in sandstorm activities due to global climate changes could significantly impact marine ecosystems through modulations in their microbial community dynamics.

POB-207

Capybara genome sequencing scrutinizes factors affecting evolutionary rates

Isaac Adeyemi Babarinde[1], Naruya Saitou[2, 3]

[1]Southern University of Science and Technology (China), [2]Graduate University for Advanced Studies (Japan), [3]National Institute of Genetics (Japan)

Capybara is the extant rodent species with the largest body weight. To investigate the evolutionary rate dynamics and impacts of body weight, the whole genome of capybara was sequenced on Illumina platform to the depth of about 14X.The estimated genome size was 2.7Gbp. Guinea pig is the closest species to capybara with determined genome sequences, and their divergence time was estimated to be 14-35 million years ago. Although the body size of capybara is more than 50 times that of guinea pig, the neutral evolutionary rates of the two species were not significantly different. However, analyses of multiple genomes revealed high heterogeneity and significantly higher evolutionary rates among rodents than other mammalian orders. The highest evolutionary rate, found in the Cricetidae-Muridae common ancestor after the divergence of Spalacidae, was more than 8 times higher than the human rate. Although all the factors investigated were significantly correlated with evolutionary rates, regression analyses showed that model containing only order, generation interval and age at maturity was the optimal (adjusted R2 = 0. 73). Including longevity, gestation length, litter frequency, litter size and body weight affected the model. Interestingly, after adjusting for order, the apparent relationship between evolutionary rate and body size disappeared. This study challenges some previous assertions and scrutinizes factors affecting evolutionary rate dynamics. The acceleration of evolutionary rate in Cricetidae-Muridae common ancestor might be related to species explosions in the descendant families as predicted by mutation-driven hypothesis of speciation.

POB-208

Seasonal dynamics of the microbial communities by time-series metagenomics in the Red Sea

Katsuhiko Mineta[1], Yoshimoto Saito[1], Intikhab Alam[1], Hayedeh Behzad[1], Mohamed Alarawi[1], Hiroshi Shimizu[1], Siham Fatani[1], Amani Al-Ma'abadi[1], Kosuke Goto[1], Hajime Ohyanagi[1], Magbubah Essack[1], John Archer[1], Vladimir Bajic[1], Takashi Gojobori[1]

[1]King Abdullah University of Science and Technology (Saudi Arabia)

Metagenomics is an emerging approach to provide a significant amount of sequence data from the organisms in the environment. For the marine environment, metagenomics is a promising and powerful approach to study the microbial community for understanding their ecosystem, environmental/evolutionary adaptation and diversity, comprehensively. Here, we focused on the Rea Sea as a target since the Red Sea has several unique environmental features such as high salinity, high temperature, and low nutrition. These features are characteristic of harsh environments that could drive the evolution of the extraordinary microbial community and potentially provide genetic resources for the pharmaceutical and bio-industrial applications. However, little is known about the dynamics of the community structure, the functional gene content, and evolutionary and environmental background of the organisms in the Red Sea environment. To better characterize the dynamics and variety of the gene contents and the microbial communities, we are conducting time-series comparative metagenomics in the Red Sea. So far, more than 400 metagenome samples were collected from seawater and sediment on a monthly basis. We took the whole-genome shotgun metagenomics approach so that we can infer the functional diversity as well as the diversity in the microbial community along with the time series. In this presentation, we will provide an overview of our Red Sea metagenome project and details of the outcome from our metagenomics analyses with the evolutionary insights.

POB-209

Uncovering the genomic determinants of adaptive radiations: Comparative transcriptomics of convergent prey specialization in Dysdera spiders from the Canary Islands

Joel Vizueta[1], Nuria Macias-Hernandez[2], Miquel A. Arnedo[2], Alejandro Sanchez-Gracia[1], Julio Rozas[1]

[1]Universitat de Barcelona (Spain), [2]Universitat de Barcelona (Spain)

During their evolutionary history, spiders diversified into a large number of different habitats, sometimes involving dietary specializations to feed on specific preys. The adaptive radiation of the genus *Dysdera* in the Canary Islands is an outstanding example of diversification associated with dietary shifts. Some endemic species of this genus have adapted to capture and feed on terrestrial isopods, a prey with prominent morphological, chemical and behavioral defenses. These spiders show specific morphological modifications of chelicerae along with behavioral and metabolic adaptations that allow them to capture and digest these preys. Preliminary phylogenetic data suggest that diet specialization has evolved multiple times independently within this archipelago. To date, however, little is known about the genetic basis of such a remarkable phenotypic convergence. To address this question, we sequenced and compared the complete transcriptome of two pairs of closely related generalist-specialist (with respect to diet) *Dysdera* species. To identify putative parallel molecular changes associated to the observed convergence, we searched for protein-coding genes with significant changes in their functional constraints, including the footprint of positive selection, and/or showing differential expression patterns in both specialist species. Based on gene ontology annotations, we propose a set of candidate genes that could have promoted prey specialization, including genes involved in metabolism and heavy metal assimilation. Since terrestrial isopods accumulate toxic substances and heavy metals, such as copper or cadmium, parallel specific changes in these genes involved in their metabolism could have promoted the evolutionary convergence of prey specialization observed in *Dysdera*.

POB-210

Using a combined RNAi/RNA-seq approach to assess the role of heat shock transcription factor (HSF) in the evolution of thermal adaptation

Alice Harada[1], Sumaetee Tangwancharoen[1], Ronald Burton[1]

[1]Scripps Institution of Oceanography, University of California San Diego (United States)

Genetically isolated populations of conspecifics that are distributed across broad geographic ranges allow for the study of adaptation to varying local environments. The intertidal copepod *Tigriopus californicus* experiences a latitudinal gradient in thermal regimes across its range along the west coast of North America, making it an attractive system for examining the evolution of thermal tolerance. Previous studies have shown that the southernmost populations of *T. californicus* have the highest survivorship following thermal stress, and RNA-seq experiments suggest that more effective *cis*-regulation of heat shock protein genes may be a key factor in the ability of southern populations to survive acute heat stress. To better understand the potential role of *trans*-regulation of the heat shock response and thermal tolerance, we used RNAi to knock down expression of the heat shock transcription factor (HSF) in populations of *T. californicus*. Following confirmed knockdown, we saw a significant drop in survivorship of adults following heat stress. We then used RNA-seq to compare gene expression in control and HSF knockdown groups. The resulting transcriptome-wide patterns of expression provide an assessment of the role of HSF in the heat shock response; ongoing experiments will allow us to compare the response in populations with varied thermal tolerance. Using the combined RNAi/RNA-seq approach, we can begin to assess the relative importance of *cis*- and *trans*-regulation in the evolution of thermal tolerance among *T. californicus* populations.

POB-211

Transcription Termination is a Mutational Target During Extreme Feast/Famine Cycles in Escherichia coli.

Megan G Behringer[1], Samuel F Miller[1], Gwyneth Boyer[1], Michael Lynch[1]

[1]Arizona State University (United States)

Microbes are robust organisms capable of persisting in the face of extreme environments. In nature, access to nutrients is not always consistent and populations commonly experience cycles of feast and famine. To further understand how microbes persist and eventually thrive in environments with sporadic nutrient availability, we performed an evolution experiment varying intervals between feeding times in laboratory *Escherichia coli* populations. Over 1000-days, 48 populations of *E. coli* K-12 were challenged with three different starvation-interval treatments, with either 1-, 10-, or 100-days between feedings. Using whole-genome population-level sequencing, we identified 19 genes enriched for nonsynonymous mutations in starved populations, eight of which are associated with transcription (*crp*, *hns*, *rho*, and *rpoC*) or translation (*fusA*, *infB*, *rplF*, and *rpsL*).

Further investigation reveals that loss-of-function of many of these genes results in lethal phenotypes, as such mutation of these genes is likely adaptive. Two mutations in particular: 1) a R109H substitution in the RNA binding 1 domain of Rho, one of three transcription termination factors in *E. coli*, has fixed in five-independent starved populations; and 2) a R93H substitution in the DNA binding domain of H-NS, a histone-like transcriptional regulator, has also fixed in these populations. Previous studies suggest interactions exist between Rho and H-NS in order to repress antisense RNA transcription. As many *E. coli* antisense RNAs are involved in environmental stress-responses, reduced binding- activity of Rho and H-NS, resulting in de-repression of antisense RNA, may be beneficial in populations enduring long-term starvation.

POB-212

Adaptive divergence and molecular evolution of venom gland transcriptomes of long-jawed orb weavers Tetragnatha versicolor (Araneae: Tetragnathidae)

Michael Brewer[1]

[1]East Carolina University (United States)

Spiders are the most successful group of venomous animals, comprising more than 47,000 described species with total estimates up to 120,000. Spiders inhabit most terrestrial ecosystems and have persisted for more than 380 million years. Much of their success can be attributed to their approximately 2 million unique venom peptides. Despite this, many spider venomes have not been characterized, including many of the most speciose and/or ecologically diverse taxa. Members of the genus *Tetragnatha* are found in diverse habitats worldwide, including a well-described Hawaiian adaptive radiation. To investigate the intraspecific evolutionary ecology of spider venomes, Atlantic Coastal Plain and Appalachian Highland populations of the widespread North American species *Tetragnatha versicolor* were studied. Divergences in trophic niches were determined via stable isotope analyses. Differential expression of venom gland transcripts corresponding to trophic niche differences indicated expression-level variation of venom components, likely due to local adaptation. Using three novel bioinformatics tools, the sequence-level molecular evolution of venom peptides was examined. Venom gland expressed gene families were reconstructed and analyzed for evidence of pervasive selection. Additionally, orthologs were identified and tested under explicit phylogenetic frameworks for branch-specific rates of protein evolution between populations. Lastly, putative venom peptides were identified, many never before characterized in any spider, using a newly developed machine learning technique. These results indicate a complex yet logical framework of population-level evolution in ecologically divergent populations of *T. versicolor*.

POB-213

Microbiome analysis of raw fish ceviche through a metagenomic approach

Humberto Martinez Montoya[1], Guadalupe Concepcion Rodriguez Castillenjos[1], Regulo Ruiz Salazar[1], Maria Cristina Hernandez Jimenez[2]

[1]Universidad Autonoma de Tamaulipas (Mexico), [2]Universidad Autonoma de Tamaulipas (Mexico)

Ceviche is a traditional dish made using raw fish meat marinated few hours in lime juice and do not have any heat cooking step throughout its preparation process. Although the use of organic acids as antibacterial agents is well known, recent research indicate that lime juice actually can reduce the risk of V. parahemolyticus, infections but it is not effective against other potential pathogens such as S. enterica. In Mexico, diarrheic diseases are mainly caused by contaminated food, it has been estimated that nearly 67% of infections due contaminated food occasioned by presence of bacterial agents revealing that fresh and frozen fish are the most affected group of food throughout the country. Ceviche usually is made using meat obtained from either marine or freshwater species. Despite the fact that fresh fish meat is safe, exposed organs including skin, gills, and guts represent a potential source of bacterial contamination. Moreover, handling, transportation and preparation could also alter significantly the structure of microbial communities present in Ceviche and increase the risk of infections due consumption of raw seafood products. In this project, we performed the characterization of the microbial communities present in ready-to-eat Ceviche using a metagenomic approach sequencing the ribosomal gene 16S. Bacterial taxonomic diversity was determined using the V3 and V4 regions of the 16S gene. In this study of microbial diversity in food we present a general panorama of the taxonomic groups present. Our results potentially will allow detection of pathogenic bacteria not detected by conventional methods.

POB-214

Comparisons of behavioral and TRPA1 heat sensitivities in three Cuban Anolis lizards from distinct thermal microhabitats

Hiroshi Akashi[1, 2], Shigeru Saito[3], Antonio Cadiz Diaz[4], Takashi Makino[1], Makoto Tominaga[3], Masakado Kawata[1]

[1]Tohoku University (Japan), [2]The Graduate University for Advanced Studies (Japan), [3]National Institute for Physiological Sciences (Japan), [4]University of Havana (Cuba)

Temperature sensation is essential for behavioral and physiological thermoregulation. While several members of transient receptor potential ion channels have been reported as molecular temperature sensors (thermo-TRPs), the divergence of thermo-TRPs among species has rarely been investigated particularly with respect to their thermal habitats. In Cuba, three *Anolis* lizards, *Anolis allogus*, *Anolis homolechis*, and *Anolis sagrei*, coexist in the same community by inhabiting distinct thermal microhabitats. The set of these species, thus, provides excellent opportunity to investigate how molecular temperature sensors diverge among species with distinct thermal physiology. Here, we focused on these three species and their TRPA1, a member of thermo-TRPs, which detects temperatures noxious for the animals. We performed behavioral experiments on these three species and electrophysiological analysis on their TRPA1 to investigate temperatures that elicit avoidance behaviors and also that activate their TRPA1 under heat stimuli. We found that temperatures triggering behavioral and TRPA1 responses were significantly lower for shade-dwelling species (*A. allogus*) than for sun-dwelling species (*A. homolechis* and *A. sagrei*). The ambient temperature of shade habitats where *A. allogus* occurs stays relatively cool compared to that of open habitats where *A. homolechis* and *A. sagrei* occur and bask. The high temperature thresholds of *A. homolechis* and *A. sagrei* may reflect their heat tolerances that would benefit these species to inhabit the open habitats.

POB-215

Detection of genes involved in adaptive evolution to hot-open habitat in Cuban Anolis lizards

Shunsuke Kanamori[1]

[1]Tohoku University (Japan)

Understanding genetic basis of adaptive evolution to different thermal habitat of ectotherms is important not only to understand the history of their evolution and diversity but also to preserve biodiversity from environmental fluctuations. In this study, evolutionary history to different thermal habitat was estimated for Cuban *Anolis*lizards, and genes positively selected in the lineages, in which species had evolved from cool-shaded to hot-open habitats, was detected. Parsimony estimation of ancestral traits indicates that the evolution from cool-shaded habitat to hot-open habitat occurred 5 times independently. Comparative coding sequence analysis found that *tgfb1* parallel evolved in species inhabiting hot-open habitat. Furthermore, positive selection was detected on several genes involved in the TGF-β/Smad signaling pathway and collagen synthesis and genes encoding type I collagen itself only on one lineage inhabiting hot-open environment. The denaturation and water retention of collagen could be important for adaptation to hot and strong UV radiation environment. Thus, the results suggest that the evolution of genes involved in collagen synthesis is important for adaptation to hot-open habitat in Cuban *Anolis* lizards. In addition, we detected positive selection in genes associated with circadian rhythm and neurotransmission, which have been suggested as genes related to thermal adaptation by previous studies.

POB-216

Genetic architecture of mate preference behaviours and the corresponding warning pattern cue facilitates ecological speciation in sympatric tropical butterflies

Richard Merrill[1, 2], Matteo Rossi[1], Pasi Rastas[3], John Davey[3], W Owen McMillan[2], Chris Jiggins[3]

[1]Ludwig-Maximilians-Universitat, Munich (Germany), [2]Smithsonian Tropical Research Institute (Panama), [3]University of Cambridge (United Kingdom)

The evolution of new species is facilitated when traits under divergent ecological selection are also mating cues. The extent to which these traits contribute to assortative mating depends on the evolution of corresponding preference behaviours, but we still know little about the genetics of behaviour in natural populations. The warning patterns of Heliconius melpomene and H. cydno are under disruptive selection for mimicry, and are also used during mate recognition. We report a genome-wide QTL analysis which reveals that divergent male preference between these species has a surprisingly simple genetic basis. Three QTLs explain 60% of the difference in preference behavior observed between the parental species. Two of these QTLs are on chromosomes with major colour pattern loci, and one is very tightly linked to the key wing patterning gene optix. Different loci influence different aspects of attraction, suggesting that behavioural isolation in Heliconius involves the evolution of independently segregating modules, similar to those for the corresponding wing pattern cues. We are now integrating these results with gene expression analyses of developing brains, measures of positive selection and admixture between the two species, and linkage analyses of further taxon pairs to better understand the evolution of visual mate recognition. Recent research has shown that mimicry in Heliconius can be achieved through hybridization and subsequent sharing of key loci between species. The existence of large effect preference loci, influencing different aspects of behaviour, could similarly facilitate the origin of novel phenotypes through recombination and introgression, and should facilitate rapid speciation.

POB-217

The rises and falls of opsin genes in 59 ray-finned fish genomes and their implications for environmental adaptation

Jinn-Jy Lin[1], Feng-Yu Wang[2], Wen-Hsiung Li[1, 3], Tzi-Yuan Wang[1]

[1]Academia Sinica (Taiwan), [2]National Applied Research Laboratories (Taiwan), [3]University of Chicago (United States)

We studied the evolution of opsin genes in 59 ray-finned fish genomes. We identified the opsin genes and adjacent genes (syntenies) in each genome. Then we inferred the changes in gene copy number (*N*), syntenies, and tuning sites along each phylogenetic branch during evolution. The Exorh (rod opsin) gene has been retained in 56 genomes. Rh1, the intronless rod opsin gene, first emerged in ancestral Actinopterygii, and N increased to 2 by the teleost-specific whole genome duplication, but then decreased to 1 in the ancestor of Neoteleostei fishes. For cone opsin genes, the rhodopsin-like (Rh2) and long-wave-sensitive (LWS) genes showed great variation in N among species, ranging from 0 to 5 and from 0 to 4, respectively. The two short-wave-sensitive genes, SWS1 and SWS2, were lost in 23 and 6 species, respectively. The syntenies involving LWS, SWS2 and Rh2 underwent complex changes, while the evolution of the other opsin gene syntenies was much simpler. Evolutionary adaptation in tuning sites under different living environments was discussed. Our study provides a detailed view of opsin gene gains and losses, synteny changes and tuning site changes during ray-finned fish evolution.

POB-218

Inferring global population structure and genetic diversity in wild house mice (Mus musculus) using a genome-wide SNP array

Jonathan Hughes[1], Andrew Parker Morgan[2], John Didion[2], Jeremy Searle[1], Fernando Pardo-Manuel de Villena[1]

[1]Cornell University (United States), [2]University Of North Carolina At Chapel Hill (United States)

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

POB-219

Diet and roosting behavior shape the color vision of leaf-nosed bats (Phyllostomidae)

Amanda Dawn Melin[1, 2], Kelly Kries[3], Marilia A. S. Barros[4], Gwen Duytschaever[1], Joseph Orkin[1], Daniel Pessoa[4]

[1]University of Calgary (Canada), [2]University of Calgary (Canada), [3]Washington University in St. Louis (United States), [4]Federal University of Rio Grande do Norte (Brazil)

Bats are a diverse radiation of mammals of enduring interest for understanding the evolution of sensory specialization. Color vision variation among species has previously been linked to roosting preferences and echolocation form in the suborder Yinpterochiroptera, yet questions remain about the roles of diet and habitat in shaping bat visual ecology. We sequenced OPN1SW and OPN1LW opsin genes for 20 species of leaf-nosed bats (family Phyllostomidae; suborder Yangochiroptera) with diverse roosting and dietary ecologies, along with one vespertilionid species (*Myotis lavali*). OPN1LW genes appear intact for all species examined. OPN1SW genes appear intact and under purifying selection for *Myotis lavali* and most phyllostomid bats, with two exceptions: 1) we found evidence of ancient OPN1SW pseudogenization in vampire bats and loss-of-function mutations in all three extant species; 2) we additionally found a recent, independently-derived OPN1SW pseudogene in *Lonchophylla mordax*, a cave roosting species. These mutations in leaf-nosed bats are independent of the OPN1SW pseudogenization events previously reported in Yinpterochiropterans. Accordingly, the evolution of monochromacy (complete colorblindness) has occurred in both suborders of bats, and under various evolutionary drivers; we find independent support for the hypothesis that obligate cave roosting drives color vision loss. We additionally suggest that haematophagous dietary specialization, and corresponding selection on non-visual senses led to loss of color vision through evolutionary sensory trade-off. Evidence of functional opsins among other phyllostomid bats, despite variation in ecology, suggests that dichromatic color vision remains adaptive for a wide range of visual tasks.

POB-221

Influence of sexual conflict on reproductive mode and fecundity in a facultative parthenogen

Mercedes Burns[1], Nobuo Tsurusaki[2]

[1]University of Maryland, Baltimore County (United States), [2]Tottori University (Japan)

Sexual conflict may be uniquely manifested in facultatively parthenogenetic systems. Theory suggests that reproductive mode switching can produce conflict between resistant parthenogenetic females and coercive males, affecting population structure. The extent to which this theory helps explain evolutionary dynamics in populations of facultative parthenogens, and the relationship between mating success, fecundity, and the factors that maintain males remains unclear. The facultative parthenogens *Leiobunum manubriatum* and *L. globosum* are Japanese harvestman species with morphological and population variation well-suited for studies of mating system maintenance of reproductive polymorphisms. Males have sexual traits that may affect coercive ability, and populations vary in the relative frequency of males. If males are present at high frequency in a population of facultative parthenogens, will those parthenogens actually mate, and how will their offspring production compare to conspecifics in low frequency male populations? We addressed these questions by collecting egg clutches of both harvestmen species from populations that varied in male frequency. In *L. manubriatum*, females from populations with few males were not more fecund than females from populations with higher male frequency, despite the potential release of the former from sexual conflict. A preliminary genomic panel revealed offspring from these high male populations were primarily produced via asexual reproduction. This is surprising because sex ratios in southern populations approach equality, increasing the probability for females to encounter mates and produce offspring sexually. We are now applying a higher-throughput genotyping-by-sequencing (GBS) approach to identify asexual versus sexual offspring across a gradient of male frequency/density.

POB-222

The reconstruction of ecological interactions through eDNA metabarcoding analyses on ice cores extracted from the largest and deepest southern European Alps glacier: Adamello, Italy

Cristiano Vernesi[1], Alexis Marchesini[1], Matteo Montagna[2], Matteo Girardi[1], Antonella Cristofori[1], Valter Maggi[3], Daniela Festi[4], Camilla Wellstein[4], Stefan Zerbe[4], Klaus Oeggl[5]

[1]Fondazione Edmund Mach/Research and Innovation Centre (Italy), [2]Universita degli Studi di Milano (Italy), [3]University of Milano Bicocca (Italy), [4]Freie Universitat Bozen-Bolzano (Italy), [5]University of Innsbruck (Italy)

For inferring environmental changes through time, ice cores offer a twofold advantage: good preservation of eDNA and detailed timescale, possibly at year and season level. In the CALICE project, we aim at estimating biodiversity changes in the Adamello glacier catchment area, experiencing in recent times dramatic land-use and climatic modifications. The retrieved 45 m ice core would in fact encompass the last 70-80 years, according to tritium radioactivity results.
Identifying plant and insect species would reveal how their ecological interaction (e.g. pollination) responded to global changes, thus shedding light on functional biodiversity modifications from macro- to micro-scale landscape level.

Co-amplification of untargeted organisms, taxonomic resolution, preferential amplification of some taxa and contamination represent the major limitations of PCR-based eDNA metabarcoding.
We developed a target enrichment approach on eDNA extracted from the first basal 10 m of the ice core. We targeted, through a synthesised bait, about 10 kb representing the most variable regions of cpDNA across a wide range of plant taxa. Approximately 8 kb of mtDNA were targeted in insects. After baits hybridization with eDNA, libraries assembly and massive sequencing, the resulted reads were compared to specific reference databases: PhyloAlps (cpDNAs of all the plant taxa of the Alps) and a custom-made database for plants and insects, respectively.
This sequence capture metabarcoding proved to be more efficient to infer plant and insects biodiversity, therefore showing the potential of ice cores as very informative climatic and biological archive for addressing the impact of global change on biodiversity.

POB-223

Why is the dN/dS ratio substantially lower in nature than in long-term laboratory evolution?

Piaopiao Chen[1], Jianzhi George Zhang[1]

[1]University of Michigan (United States)

Lenski's long-term experimental evolution of *Escherichia coli* in a constant low-glucose environment is characterized with a gradual decline of the nonsynonymous to synonymous substitution rate ratio (d_N/d_S) over time. However, the overall d_N/d_S ratio for the first 50,000 generations is still much greater than 1. Similar observations have been made in other experimental evolution studies of microbes in constant environments. Because the environment fluctuates more frequently in nature than in these laboratory settings, positive selection should be stronger and more frequent in natural evolution. Surprisingly, however, comparisons of natural strains of microbes typically yield much lower d_N/d_S ratios (<0.1). We hypothesize that the low d_N/d_S ratio in nature is due to the high prevalence of antagonistic pleiotropy such that nonsynonymous mutations beneficial in one environment are harmful in the next environment, preventing their fixations. To test this hypothesis, we conducted 1120 generations of experimental evolution of yeast in constant as well as changing environments. We predict that d_N/d_S is lower in changing environments than in constant environments and that d_N/d_S should further decrease as the changing environments become more antagonistic to one another. Comparing the genome sequences of the evolved strains shows clear evidence for the above predictions. Hence, observations of low d_N/d_S ratios in nature should not be interpreted as a scarcity of positive selection. This finding has profound implications for understanding the relative roles of genetic drift and positive selection in evolution.

POB-225

Early environmental changes and metabolism evolution

Anne Oudart[1], Celine Brochier-Armanet[1]

[1]Biometry and Evolutionary Biology laboratory, University of Lyon (France)

Geosphere and Biosphere have co-evolved for billions of years. Thus, understanding the evolutionary history of metabolisms is essential to decipher the history of Earth, and reciprocally. The evolutionary period spanning from the Last Universal Common Ancestor (LUCA) to the diversification of the three Domains of Life (*Archaea*, *Bacteria*, and *Eucarya*) has witnessed the emergence of the present-day biodiversity as well as major environmental changes such as the massive oxygenation of the primitive atmosphere and oceans (i.e. Great Oxidation Event) linked to the emergence of *Cyanobacteria* and oxygenic photosynthesis.

Determining the relative emergence order of enzymes provides important information about the origin and evolution of the corresponding metabolisms and the availability of chemical compounds in ancient ecosystems. The recent outburst of genome sequencing projects, in particular for *Archaea*, offers a unique opportunity to tackle this issue. To understand the impact of the rise of dioxygen on past biospheres, we performed a large scale phylogenomic analysis of the 709 enzymes described to use and/or produce dioxygen, nitrite oxide, and superoxide. Our data revealed that more than a hundred of them are of recent origin, while 7 seemed to be very ancient, having originated before the emergence of *Cyanobacteria*, and thus of the oxygenic photosynthesis. This suggested that the use of dioxygen in metabolic processes have likely preceded the Great Oxidation Event and thus that some sources of dioxygen were available prior to the emergence of oxygenic photosynthesis.

POB-226

Detecting signatures of convergent adaptation in population genomic data

Kristin M Lee[1], Jessica Selby[2], John Willis[2], Graham Coop[1]

[1]University of California, Davis (United States), [2]Duke University (United States)

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

POB-227

Turning sex into apomixis: Genetic crossing reveals the mode of apomixis transfer into sexual populations of the genus Boechera

Martin Mau[1, 2], Tiina Liiving[1, 2], Liza Fomenko[1], Richard Goertzen[1], Laura Boettner[2, 3], Timothy Francis Sharbel[1, 2]

[1]University of Saskatchewan (Canada), [2]Leibniz Institute of Plant Genetics and Crop Plant Research (Germany), [3]Max-Planck Institute for Chemical Ecology (Germany)

Natural selection has led to the evolution of asexual seed production (apomixis) from sexual ancestors in a number of flowering plants (Angiospermae). Engineering apomixis in major crop plants, which predominantly reproduce sexually, represents a significant enabling technology that would be able to fix and maintain valuable genotypes and associated traits and could avoid inbreeding depression effects. Hence, understanding the transfer and dispersion of apomixis in natural populations is of great interest to breeders and evolutionary biologists, although measuring the introgression of apomixis into sexual populations is clouded by ploidy barriers and mentor effects. Here we describe our use of a large intra- and interspecific genetic crossing panel to elucidate the transfer and maintenance of apomixis in sexual populations of the genus *Boechera* (Crucifereae). The frequencies of apomixis transfer and segregation patterns were evaluated by flow cytometric seed screens and by the presence/absence of known apomeiosis factors in parents and progenies. We identified one sexual and two apomictic pollen classes based on ploidy, germination rates and fitness, and demonstrate that only meiotically-reduced apomictic pollen enables the spread of apomixis into sexual populations. Homoploid hybridization can thus lead to the establishment of highly facultative or even obligate apomicts within a single generation. In contrast, fertilization with unreduced apomictic pollen largely produces highly sterile polyploids. Ultimately, we aim to identify all contributing genetic factors to apomictic reproduction in *Boechera* by narrowing down donor male genomic regions using a comprehensive backcrossing strategy in combination with large scale resequencing of all parents and backcrosses.

POB-228

From Genotyping-by-Sequencing to aDNA, a peek into the past and future of the New Zealand Kaka

Denise Martini[1], Neil Gemmell[1], Bruce Robertson[2], Michael Knapp[1]

[1]University of Otago (New Zealand), [2]University of Otago (New Zealand)

The Kaka (*Nestor meridionalis*) is a large parrot endemic to New Zealand and at risk of extinction. Populations are fragmented and subject to conservation management aiming to reduce inbreeding, avoid outbreeding depression and reinstate the original phylogeographic structure of the species. However, key information to support such management is missing. Kaka display phenotypic and behavioural variation across their range and two subspecies, the North Island and the South Island Kaka, are recognised. Nevertheless, recent genetic studies using neutral markers failed to find any population structure.

We are using a combination of ancient DNA data and population genomics to reconstruct the past population structure of the Kaka and identify functional genomic structure across present day populations.

Whole-genome sequencing of the two subspecies provided a reference genome for this study. Preliminary Genotyping-by-Sequencing data indicate that Kaka retain a good amount of genetic diversity overall. Across all markers, the data confirms the lack of population structure but mapping variation against our Kaka genomes allows us to focus on functional variation and to identify potential local adaptation to the diverse biota that Kaka encounter across their range. Together with our data from subfossil remains, our results provide insights into how long populations have been isolated and how this isolation has influenced their genetic structure and variability. Our study gives a perspective on the differential impact that human colonisation and climate change had in shaping the Kaka's present-day variability and will help to develop conservation strategies for the recovery of this species.

POB-229

Genomic analyses of the population structure in *Daphnia pulex*.

Takahiro Maruki[1], Zhiqiang Ye[1], Michael Lynch[1]

[1]Arizona State University (United States)

Studying population structure is important for understanding the evolutionary features of an organism. Previous studies using a small number of molecular markers found high degree of genetic differentiation among the populations in the freshwater microcrustacean *Daphnia pulex*. However, the dispersal ability of *D. pulex* is potentially high, as they can produce resting eggs that can be transported to other areas by the wind, flowing water, and other animals.

In this study, we investigate the population structure in *D. pulex* using high-throughput sequencing data of 96 individuals from each of ten temporary ponds mainly in the Midwestern part of the United States. We show that most of the analyzed populations are in Hardy-Weinberg equilibrium. To examine the population structure, we estimate Wright's fixation indices at single-nucleotide polymorphism sites. Our F_{ST} estimate is much lower than those in previous studies, raising the possibility that confident identification of targets of local adaptation may be possible. Pairwise F_{ST} estimates show that the degree of genetic differentiation is positively correlated with geographic distance between populations. To identify signatures of natural selection, we carried out sliding-window analyses of F_{IS} and F_{ST} estimates, identifying top and bottom outliers using the bootstrap. We compare the estimates of the heterozygosity and genetic divergence from the sister species *D. pulicaria* between silent and replacement sites in protein-coding sequences to find functional changes involved in the putative targets of selection. We infer roles of different evolutionary forces in shaping the population structure in *D. pulex*.

POB-230

Speciation and secondary contact between two Sulawesi macaque species, Macaca tonkeana and M. hecki

Yohey Terai[1], Sohei Takuno[1], Hiroo Imai[2], Laurentia Henrieta Purba[3], Kanthi Arum Widayati[3], Bambang Suryobroto[3]

[1]SOKENDAI (The Graduate University for Advanced Studies) (Japan), [2]Kyoto University (Japan), [3]Bogor Agricultural University (Indonesia)

The Sulawesi macaque species endemic to island of Sulawesi (Central Indonesia) have differentiated into seven morphologically distinct species in seven allopatric areas. The evolution of these species has been studied, however the evolutionary process and its genetic basis are still unclear. The remarkable point of these species is that species hybridized at the boundary of distribution (hybrid zones), while the hybrid zones have not extended. Therefore, the genetic differences between species are expected at loci responsible for local adaption, as a consequence, for preventing expansion of hybrid zones. Here, we report the pattern of genomic differentiation between two Sulawesi macaque species. We determined exome sequences of Macaca tonkeana (n=11) and M. hecki (n=11) that distribute side-by-side and had been reported the hybrid populations in their hybrid zone. Single nucleotide polymorphisms (SNPs) were extracted from the exome sequences. Most of the SNPs were observed in one species (~90%) or shared between two species, whereas a handful of SNPs were fixed differences. These fixed differences were located in ~150 genes including genes responsible for olfaction, detoxification, hair formation, and reproduction in female. The analyses by STRUCTURE and site frequency spectrum showed that these two species have started to differentiate 50 thousand generation ago, and admixed recently by hybridization. These results demonstrate that most of genomic regions have not completely differentiated and the small number of genes with fixed differences may be responsible for local adaption and for preventing expansion of hybrid zones.

POB-231

Genomes of Hawaiian Canacidae (Beach Flies)

Nina Pak[1]

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

Transitioning from marine to freshwater habitats has allowed multiple radiation and speciation events for many taxa. On macro-evolutionary time scales, this transition has occurred more frequently than the transition from freshwater to land, and thus provides opportunities for understanding rapid responses to environmental change and freshwater invasion. Salinity is a source of stress for many animal phyla and insect lineages, shaping distributions and influencing community structures of saltwater ecosystems. The inability to respond to high salt conditions may compound negative consequences of climate change, such as drought in arid and hyper-arid climate conditions. Evolving salt tolerance may allow these groups to escape predation, to reduce competition, and to avoid water loss. Canacidae (beach flies) are mostly found in intertidal ecosystems where their larvae feed on algae. Within the last 5 million years, one lineage of Hawaiian species has lost the ability to tolerate saline habitats and now occupy high elevation freshwater streams. By investigating whether genes that are known for salt tolerance are present in Canacidae genomes, testing whether these salt tolerance genes are under selection, and examining other areas of selection in the genomes, we can further our understanding of freshwater invasions and pathways behind osmoregulatory systems. We expect that Canacidae may employ some, but not all, of the same genes to cope with saline environments.

POB-233

Population genomics of woodland strawberry reveal strong selection of extensive adaptive introgressions in different climates

Tuomas Toivainen[1], Samia Samad[1], Takeshi Kurokura[4], Ari Loytynoja[2], Patrick P Edger[5], David Pose[6], Lars Paulin[2], Petri Auvinen[2], Timo Hytonen[1, 3]

[1]University of Helsinki (Finland), [2]University of Helsinki (Finland), [3]University of Helsinki (Finland), [4]University of Utsunomiya (Japan), [5]Michigan State University (United States), [6]University of Malaga (Spain)

Plants have colonized large parts of Europe and adapted to new environments after the postglacial warming of the climate. Woodland strawberry (Fragaria vesca ssp. vesca) is a perennial model species with an especially wide geographical distribution. It grows from the Mediterranean region with hot and dry summers (southern Spain, latitude 37N) to the arctic climate of the most northern parts of Norway (latitude 70N). European populations show phenotypic differentiation along a latitudinal cline especially in reproductive traits, which suggests that most populations have adapted to local climates. To explore drivers of climatic adaptation in woodland strawberry we sequenced 217 accessions using Illumina technology. Population structure clearly shows that postglacial woodland strawberry populations originate from southern and eastern refugia. Effective population sizes have strongly declined during and after the last glacial maximum. Interestingly, both northern Norwegian and Icelandic populations have gone through strong bottlenecks about 1000 years ago coinciding with Viking raids and settlements in the regions. Gene flow has been an important factor for moving adaptive variation between populations. We show adaptive introgressions and selective sweeps of up to 10Mb in both southern and northern European populations and discuss their possible roles in the adaptation of woodland strawberry to climatic extremes within its geographical range.

POB-234

Evolution of RNA polymerase subunits in endosymbionts proteobacteria with extreme genome reduction

CYNTHIA RANGEL[1], Edgardo Galan[2], Agustino Martinez[1]

[1]Cinvestav (Mexico), [2]Universidad autonoma de san luis potosi (Mexico)

The RNA polymerase (RNAp) is essential in all organisms. In bacteria it is composed of six subunits (α_2,β, β', ω and σ). However, some endosymbiotic bacteria with extreme genome reduction retain a minor number of these subunits. Additionally, in these organisms the protein subunits lack of apparently essential domains if compared to E. coli. Nevertheless, the effect of these reductions in the function of these proteins is unknown.

In this work, we studied the evolution of RNAp of four groups of endosymbiont proteobacteria with the smallest genomes sequenced to date, through comparative genomics, phylogenetic analysis and prediction of tridimensional models of each subunit.

We identified loss of fragments on important domains for the RNAp assembly which can be compensated by some amino acid residues under positive selection. In addition, coevolution analyses indicate the existence of covariation between amino acids that participate in the interaction between different subunits. Beside, we found exclusive modifications with experimental evidence in sites and regions involved on the recognition and binding to the promoter. All these suggest changes in the specificity for the canonic -10box(TATAAT). Furthermore, RNAp of some strains apparently lost the capacity to bind the -35 box and instead have structural similarities with the RNA pol II of Archaea. Finally, we propose structural models for RNAp of endosymbiont bacteria with minimal genomes. We conclude RNAp of each these endosymbiotic bacteria, follows a unique evolutionary path which differs from the rest of bacteria, probably mediated by interactions with their hosts.

POB-237

Searching for deep prokaryotic lineages in the ocean.

Romain Lannes[1], Philippe Lopez[1], Eric Bapteste[1]

[1]Pierre et Marie Curie (France)

Metagenomic enables the recovery of genetic information from an environment, even from uncultivated microorganisms that are thought to represent from 60 up to 99% of the microbial diversity. Thus, searching for distant homologous of known genes in environmental metagenome datasets could lead to the discovery of new lineages or new enzymes of biological/industrial interest. Mining huge metagenome datasets is challenging: for example, the TARA ocean expedition produced to $44x10^6$representative genes. Sequences similarity networks are useful tools for identification of homologous sequences, but require an all against all pairwise alignment, which becomes computationally challenging when applied to huge datasets. We developed a tool to identify distantly related homologs in huge environmental datasets by extracting a sequence similarity network focusing on the gene family of interest. Our method is based on an iterative process that uses each sequence previously gathered as a new seed for the next step. Because such iterative approaches can lead to false positives we enforce specificity by adding two rules to the iterative process. Firstly, we conserve the alignment position from the initial search step in subsequent iterations. Secondly, we only consider sequences that have a similar size compared to the initial query. Currently, our pipeline can use two aligners NCBI BLAST or DIAMOND. We will present the pipeline and one application on TARA OCEAN metagenome dataset, identifying extremely distant homologs of ribosomal proteins that are conserved in archaea and bacteria, hinting at the presence of deep novel prokaryotic lineages.

POB-238

Comparative genomics of bioluminescent fungi

Isheng Jason Tsai[1], Huei-Mien Ke[1], Chan-Yi Lin[1], Pei Hsuan Wu[1], Chiung-Chih Chang[2], Hsiao-Wei Kao[2]

[1]Academia Sinica (Taiwan), [2]National Chung Hsing University (Taiwan)

More than 75 bioluminescent fungi have been described with more than half of them described in the *Mycena* genus. Despite its fascinating phenotypic feature, little is known about the genomic basis of bioluminescence and evolution of these species. Here, we report five whole-genomes of bioluminescent fungi including *Mycena kentingensis* which is endemic in Taiwan. These 42- to 165-megabases genomes were sequenced using long read sequencing and assembled into N50 of 1-3Mb. Comparative genomics analysis revelaed many genes reflecting their ability as saprobic fungi in forest ecosystems. Transcriptome data from comparing bioluminescent- against normal mycelium revealed upregulation of members of Carbohydrate active enzymes, P450 and heme peroxidase. These results shed light on the basic biology and genomic basis underlying the bioluminescence of fungi.

POB-240

Population divergence and thermal selection: Combining reverse ecology with population transcriptomics

Maike Herrmann[1, 2]

[1]Paul-Ehrlich-Institut (Germany), [2]University Koblenz-Landau (Germany)

Revealing the genetic basis underlying adaptive responses is essential for a comprehensive understanding of evolutionary processes. Attaining this objective is, complicated by the interplay of forces influencing genetic divergence among populations. To assess the contribution of local thermal selection to population divergence, we combined reverse ecology and population transcriptomic approaches; thereby considering transcriptome-wide variation in both gene expression profiles and DNA sequences. Based on an extensive literature search for candidate genes under thermal selection in arthropods and comparisons of transcriptomes among four *Daphnia galeata* populations, we identified candidate transcripts potentially responding to local thermal selection.

Temperature-relevant candidate genes were overrepresented among transcripts strongly contributing to sequence divergence among populations; indicating that local thermal selection acted on the coding sequence level. Based on outlier tests and distinctive expression profiles, we identified a large number of transcripts which may contribute to local thermal adaptation. Temperature-relevant candidate genes were, however, not overrepresented among these, compared to the global gene set. The majority of genes contributing strongly to sequence divergence did not contribute strongly to divergence at the expression level and vice versa, but interestingly, gene functions were largely consistent between the two data sets. Thus, we conclude that thermal selection played a minor role in divergence among *Daphnia* populations and that genetic and regulatory variation constitute alternative routes for responses to natural selection. Our combination of literature-based identification of ecologically informative candidate genes and a population transcriptomics approach represents a powerful methodology with a wide range of applications in evolutionary biology.

POB-241

Genomic architecture of adaptation to the freshwater in ten independently adapted freshwater populations of G. aculeatus from the White Sea basin.

Nadezhda V. Terekhanova[1, 2], Alexey S. Kondrashov[3], Georgii A. Bazykin[1, 2], Nikolai S. Mugue[4]

[1]Skolkovo Institute of Science and Technology (Russian Federation), [2]Institute for Information Transmission Problems of the RAS (Russian Federation), [3]University of Michigan (United States), [4]Russian Federal Research Institute of Fisheries and Oceanography (Russian Federation)

Threespine sticklebacks adapted to freshwater environments all over the Northern Hemisphere. This adaptation involved parallel recruitment of freshwater alleles in clusters of closely linked sites, or divergence islands (DIs). We examine 10 freshwater populations of similar ages from the White Sea basin, and study the repeatability of patterns of adaptation in them. Overall, the 65 detected DIs tend to reside in regions of low recombination, underlining the role of reduced recombination in their establishment. Moreover, the DIs are clustered in the genome to the extent that is not explainable by the recombination rate alone. 21 out of the 65 DIs are universal; i.e., the frequency of freshwater alleles in them is increased in all analyzed populations. Universal DIs tend to be longer, and the divergence between the marine and the freshwater haplotypes in them is higher, implying that they are older. Across the genome, the frequencies of freshwater haplotypes in 8 pairs of DIs are significantly correlated, implying similar response to ecological differences between populations and/or epistatic interactions between DIs. Within most DIs, the same set of sites distinguished the marine and the freshwater haplotypes in all populations; however, in some of the DIs, the genetic architecture of the freshwater haplotype differed between populations, suggesting that they could have been established by soft selective sweeps.

POB-242

Genome re-sequencing of natural isolates reveals patterns of local and continental population structure in the green alga *Chlamydomonas reinhardtii*.

Rory J Craig[1], Katharina B Boendel[1], Rob W Ness[2], Nick Colegrave[1], Peter D Keightley[1]

[1]University of Edinburgh (United Kingdom), [2]University of Toronto Mississauga (Canada)

Recent studies exploring the biogeography of terrestrial and freshwater unicellular eukaryotes have found evidence of substantial population structure, contrary to the widely-held "Everything is everywhere" view of microbial evolution. However, free-living and non-domesticated unicellular eukaryotes remain greatly under-studied, with analyses of genome-wide genetic variation and structure currently limited to the yeast *Saccharomyces paradoxus* and the ciliate genus *Paramecium*. We here present whole-genome resequencing data for 20 natural isolates of Chlamydomonas reinhardtii, sampled from two sites in Quebec, separated by ~80 Km, with one site sampled at two time points. Analysed together with existing data for other North American isolates, we find evidence of substantial genetic variation both within and between locations. These data support two distinct lineages of *C. reinhardtii*, forming North-Eastern and South-Midwest-Middle Atlantic clades with ~5% divergence (~26 million generations, using a mutation rate of 9.63×10^{-10} from mutation accumulation experiments). Various species (including *S. paradoxus*) exhibit similar distribution patterns reputedly due to allopatric separation during the last glacial period, and although the generation time of *C. reinhardtii* in the wild is unknown, the observed divergence in not inconsistent with such a scenario explaining the two clades. Within Quebec, we find limited spatial or temporal differentiation, with low K_{st} (<0.1) between the two sample sites and time points. Genetic diversity (π) for both locations at four-fold degenerate sites is ~2.8%, resulting in an effective population size estimate of $\sim 1.4 \times 10^{7}$. Overall, our results support the moderate endemicity model of microbial evolution for *C. reinhardtii*.

POB-243

Lineage specific deferentially expressed genes in eight endangered amphibians in Ryukyu Archipelago

Takeshi Igawa[1], Quintin Lau[2], Masatoshi Matsunami[3], Masayuki Sumida[1]

[1]Hiroshima Univesity (Japan), [2]Sokendai (The Grad. Univ. Adv. Std.) (Japan), [3]University of the Ryukyus (Japan)

How functions and diversities of the genes contributed to characterize the extant biodiversity is the focus of interest in current biology. For endangered species, genetic diversity of functional genes is also important for their survival through the changing environment. In Japan, higher number of endangered amphibians distribute in Ryukyu archipelago. For sustainable conservation of these species, we sequenced skin and spleen transcriptomes of eight endangered amphibians (7 anurans: *Odorrana splendida*, *O. ishikawae*, *O. amamiensis*, *O. narina*, *O. supranarina*, *Babina subaspera* and *B. holsti* and a urodele: *E. andersoni*) in Ryukyu archipelago and compared expression patterns and repertoire of the genes. As the result, more than one hundred thousand genes were obtained and approximately 20 % were annotated by comparing proteins of X. tropicalis. Comparing expression patterns of spleen and skins, we found numerous specifically expressed genes in each tissue, such as innate and acquired immunity related genes. We also found lineage specific expression patterns, which should be related to their behavior and ecology. Using tree-based screening approach, more than five thousand putatively orthologous genes were isolated and some genes showed deviation from negative selection.

POB-246

Genomic investigation of hybridisation and incipient speciation in Eucalyptus

Kevin D Murray[1], Tim Collins[2], Justin Borevitz[1], Rose Andrew[2]

[1]Australian National University (Australia), [2]University of New England (Australia)

Eucalyptus species are diploid, highly out-crossing and known to readily form inter-specific hybrids. This study investigates the putative genomic mosaic of a putative incipient species. Initial evidence suggests hybridization as a plausible mechanism of speciation. This study uses new whole-genome sequencing data to resolve the contribution of hybridization and ancestral population structure to the speciation of this lineage.

POB-248

Molecular Basis of Variable Glyphosate Resistance in Bacterial Gut Symbionts of Honey Bees

Erick Vicente Da Silva Motta[1], Kasie Raymann[2], Nancy Moran[1]

[1]University of Texas at Austin (United States), [2]University of North Carolina at Greensboro (United States)

Glyphosate, the primary herbicide used for weed control in agriculture, targets the EPSPS enzyme in the shikimate pathway found in plants and some microorganisms. Thus, glyphosate may affect the microbiota of animals living near agricultural sites, including pollinators such as bees. The bee gut microbiota is dominated by eight bacterial species which have been found to affect bee growth and pathogen susceptibility. These bacterial species exhibit a high level of strain diversity and possess diverse metabolic capabilities. We identified the gene coding for the EPSPS enzyme in all sequenced genomes of bacterial members of the bee gut, but *Lactobacillus* Firm-5, suggesting that these strains are susceptible to glyphosate. Then, we demonstrate that bees exposed to sub-lethal concentrations of glyphosate have their gut microbial communities disturbed. Bacterial species containing a Type I EPSPS enzyme (sensitive to glyphosate) were decreased in abundance in glyphosate exposed bees, whereas species containing a Type II EPSPS enzyme (insensitive to glyphosate) were either not affected or increased in abundance following exposure. In vitro experiments largely confirm the molecular mechanisms of glyphosate tolerance in these bee gut bacteria. However, a subset of *Snodgrassella alvi* strains is tolerant to glyphosate despite presenting a susceptible EPSPS enzyme. Also, glyphosate exposure on early gut colonization disturbs the microbiota and increases mortality of bees exposed to the opportunistic pathogen *Serratia*. We conclude that exposure of bees to glyphosate alters their beneficial gut microbiota, potentially affecting bee health and their effectiveness as agricultural pollinators.

POB-250

Presence-absence polymorphisms of highly expressed FP sequences contribute to fluorescent polymorphisms in the stony coral, Acropora digitifera

Shiho Takahashi-Kariyazono[1], Yohey Terai[1]

[1]SOKENDAI (The Graduate University for Advanced Studies) (Japan)

Despite the biological roles of fluorescent proteins (FPs) have been proposed based on natural variation in fluorescence in corals, their roles and the genetic basis of FP-mediated color polymorphisms in *Acropora* remain unclear. In this study, we determined the genetic mechanism underlying fluorescent polymorphisms in *A. digitifera*. Using a high-throughput sequencing approach, we found that *FP* gene sequences in the multi-gene family exhibit presence-absence polymorphism among individuals. A few particular sequences in short-to-middle wavelength emission (S/MWE) and middle-to-long wavelength emission (M/LWE) clades were highly expressed in adults, and different sequences were highly expressed in larvae. These highly expressed sequences were absent in the genomes of individuals with low total FP gene expression. In adults, presence-absence differences of the highly expressed *FP*sequences were consistent with measurements of emission spectra of corals, suggesting that presence-absence polymorphisms of these *FP* sequences contributed to the fluorescent polymorphisms. In a functional analysis of recombinant FPs that were highly expressed in larval and adult stages, we found that adults utilize UV to short wavelength light and larvae utilize middle-to-long wavelength light via FPs. The highly expressed *FP* sequences exhibited presence-absence polymorphisms in four populations with geographical separation, suggesting that the presence-absence status was maintained during the evolution of *A. digitifera* populations. The difference in FPs between adults and larvae and the polymorphisms of highly expressed *FP* genes may provide key insight into the biological roles of FPs in corals.

POB-252

Why fly when you can walk? The convergent evolution of flightlessness in rails.

Gillian Gibb[1], Steve Trewick[1]

[1]Massey University (New Zealand)

It is a curious, even paradoxical, phenomenon that many birds are flightless, when for most of us birds are quintessentially flyers. For a flightless species to evolve from a flying ancestor there must be good ecological reasons and pathways for selection: options for being flightless must exist. The rails (Rallidae) are a cosmopolitan group of birds with a broad distribution through the world. Flightlessness has evolved independently multiple times within the group, often in association with the colonisation of islands and there are numerous examples in the New Zealand/Pacific region. Nearly all islands of the Pacific have been colonized by one or more rail lineages, and before human contact it is likely that most islands had endemic flightless rails. Evolution of flightlessness in rails has been shown to occur rapidly, and may be measured in 'generations rather than millennia'. Therefore rails are a highly appropriate case study of convergent evolution, and an excellent group for the comparative analyses of flightlessness. In New Zealand, the South Island Takahe (*Porphyrio porphyrio*) and Weka (*Gallirallus australis*) are living examples of flightless rails. A number of recently extinct species are also known, including *Gallirallus dieffenbachi, Gallirallus modestus* and *Porphyrio mantelli*. While ecological reasons for loss of flight may be obvious, what sort of genetic adaptations might be underlying the evolution of flightlessness? We use comparative genomics of developmental and metabolic genes, coupled with ancient DNA and a temporal-spatial framework to investigate the evolution of flightlessness in New Zealand rails.

POB-256

Genetic analysis of diversifying genitalia in ripening-fruit-consuming Drosophila species

Leona Muto[1], Yoshitaka Kamimura[2], Kanoko Takahashi[1], Airi Sato[1], Kentaro M. Tanaka[1], Aya Takahashi[1, 3]

[1]Tokyo Metropolitan University (Japan), [2]Keio University (Japan), [3]Tokyo Metropolitan University (Japan)

Drosophila suzukii, which have recently become a serious fruit pest, have uniquely modified ovipositors, which have enabled them to lay eggs inside the hard skin of ripening fruits. Their ovipositors have a larger number of modified hard bristles lined up along the edges and a more elongated and less curved shape compared to those of a closely related species, *D. subpulchrella*. The latter species is also found on ripening fruits but cannot puncture as hard surface as *D. suzukii*. In *Drosophila* spp., the ovipositor is also known to play an important role in copulation. Thus, we conducted a detailed observation of the spatial configuration of genitalia of copulating pairs and found interspecific differences in copulating postures and coupling positions of genital structures associated with the elongation of the ovipositor in *D. suzukii*. These observations suggest an intriguing possibility that coevolution in genitalia morphology between male and female *D. suzukii* has been driven by changes in female ovipositor shape due to a shift in host fruit preference to less ripening fruits with harder skins. Furthermore, the changes in genital coupling positions can cause mechanical reproductive incompatibility between divergent forms. Therefore, we have conducted a genetic analysis of genital morphology in these two species to investigate the genetic basis underlying this evolutionary process potentially leading to ecological speciation.

POB-257

Transcriptomic response of the Drosophila melanogaster Minutes

Jai Andrew Denton[1], Floyd A. Reed [2]

[1]OIST (Japan), [2]University Of Hawaii At Manoa (United States)

Minutes, first named for their shorter bristles, are a class of *Drosophila melanogaster*mutants described almost a century ago. Extensive investigation revealed this class of mutants to be almost exclusively disruption of ribosomal proteins. Minutes are dominant phenotypes due to haploinsufficiency, a phenotype that is broadly shared across eukaryotes, and with rare exception are homozygous lethal.

Here we describe the influence of mutations in ribosomal protein encoding genes on the *Drosophila melanogaster* transcriptome. Each of the mutants displays a complex and unique transcriptomic response but shared between these mutants are overlapping groups of differentially expressed genes. Moreover, there is a strong indication that these ribosomal protein genes act as genomic gatekeepers.

POB-263

The Methylome of the Lakes : environmental epigenomics in East African cichlids

Gregoire Vernaz[1, 2, 3], Eric A. Miska[1, 2, 3], Richard Durbin[3], Milan Malinsky[4], Hannes Svardal[3], George Turner[5], Emilia Santos[1, 2, 3]

[1]University of Cambridge (United Kingdom), [2]University of Cambridge (United Kingdom), [3]University of Cambridge (United Kingdom), [4]University of Basel (Switzerland), [5]Bangor University (United Kingdom)

The hundreds of species of the Lake Malawi cichlid adaptive radiation show a remarkable diversity of phenotypic and ecological adaptations. Strikingly, recent studies highlighted that genetic diversity within the radiation is among the lowest ever observed in vertebrates. Such a high phenotype/genotype diversity ratio makes Lake Malawi cichlids a promising system to investigate the role of genetics and, for the first time at a species level, epigenetics in the context of adaptation. Yet, the molecular mechanisms and, in particular, any epigenetic aspects underlying such phenotypic diversity and speciation success remain unknown. Here, we focus on whole-genome DNA methylation landscape (methylome), a heritable and dynamic epigenetic mark that has been reported to be responsible for rapid and transmissible changes in phenotype in plants and mammals. We hypothesise that the liver methylome would affect liver function and thus be related to diet. We thus performed sequencing of the liver of five species with different feeding strategies. Preliminary results reveal striking differences in methylome at conserved genomic regions, with the most divergent patterns seen between a pelagic fish-eating and an algae-eater rock-dwelling species. Interestingly, GO analyses of differentially methylated regions (DMRs) and differentially expressed genes suggest enrichment for terms related to liver metabolic processes. Furthermore, we observe a significant enrichment of DMRs localised within transposable elements (TE), suggesting a role of TE methylation levels in the adaptation of liver function. We conclude there might be a crosstalk between local environment and methylome in cichlids. Future experiments would investigate methylome heritability in inter-species-hybrids.

POB-265

Convergent increase of gene body methylation in mangroves and its role in gene expression homeostasis

Yushuai Wang[1], Aimei Dai[1], Tian Tang[1]

[1]Sun Yat-sen University (China)

Gene body methylation (gbM) is a common feature in plants but its function and evolutionary consequence remain largely unknown. It is suggested gbM may play a role in stabilizing gene expression. One may expect an increase of gbM in plants that inhabit fluctuating environments. Here we test the homeostatic functions of gbM in mangroves that have adapted to the tropical and subtropical intertidal zones. Using bisulfite sequencing, we found gbM of orthologs is convergently 2-3 times more prevalent in three lineages of mangroves in comparison to their terrestrial relatives. More than three hundred gbM are highly conserved and specific to mangroves, whereas only one of the gbM conserved between orthologs of more than two species are specific to nonmangroves. Genes with conserved gbM specific to mangroves are featured with long coding-region, low CpG density (measured by CpG[O/E] value) and slow evolutionary rate. Despite the high expression level of genes with gbM, genes with conserved mangrove-specific gbM are expressed in a narrow range, less disturbed in response to salt stress and show smaller expression divergence between mangrove species in compared with the unconserved mangrove-specific gbM genes. Our findings support the role of gbM in gene expression homeostasis and highlight the evolutionary significance of gbM in plant stress adaptation.

POB-266

Origination of Human Exons through Differential Nucleosome Occupancy

Yumei Li[1], Chen Li[1], Chuan-Yun Li[1], Aibin He[1]

[1]Peking University (China)

Modifications on nucleosome have been implicated in fundamental epigenetic regulations, while the function of the nucleosome binding itself remains to be addressed. Here we present high-resolution nucleosome occupancy profiles for multiple tissues derived from human, monkey, tree shrew, mouse and pig. Genome-wide comparison reveals conserved nucleosome profiles across different species and tissue types, indicating its potential role in shaping long-term, cross-species changes. Notably, we found significantly higher nucleosome occupancy in exons than introns, a pattern correlated with the different exon-intron GC content. We then determine whether the biased occupancy may have some regulatory roles in exon formation, or merely represent a downstream effect after the formation of exons, through a comparative approach. By identifying human-specific exons using matched RNA-seq, we found that higher exonic nucleosome occupancy also existed in orthologous regions in rhesus macaque without these exons. Such biased nucleosome occupancy presumptively facilitates the origin of the new exons by maintaining a pre-existing ancestral splicing motif. We thus propose a "nucleosome-first" model for the origination of exons *via* differential nucleosome occupancy.

POB-267

On the track of the evolutionary forces shaping gene-body methylation conservation and variation in *Brassicaceae*

Robert Horvath[1], Shohei Takuno[2], Benjamin Laenen[1], Tajan Slotte[1]

[1]Science for Life Laboratory, Stockholm University (Sweden), [2]SOKENDAI (The Graduate University for Advanced Studies), Hayama, Kanagawa (Japan)

Gene-body methylation (gbM) refers to an increased level of methylated cytosines specifically in a CG sequence context within genes. Gene-body methylation is found in plant genes with intermediate expression level, that evolve slowly, and is often broadly conserved across millions of years of evolution. Intriguingly however, some plants lack gbM, and thus it remains unclear whether gbM has a function. In animals, there is support for a role of gbM in reducing erroneous transcription, but so far most studies in plants have tested for an effect of gbM on expression level, not noise. Here, we therefore test whether gbM is associated with reduced expression noise or erroneous transcription using single-cell expression data. To assess whether changes in gbM status are associated with shifts in selective pressures, we analyse genome-wide polymorphism data from four crucifer species to quantify selection pressures on genes with evolutionarily conserved vs. variable gbM status. The results contribute to our understanding of the evolutionary and functional importance of gbM.

POB-268

Characterisation of extinct bison methylomes using bisulphite sequencing

Bastien Llamas[1], Holly Heiniger[1], Graham Gower[1], Paul Gooding[2], Catherine M Suter[3, 4], Stefan Hiendleder[5, 6], Jeremy F Taylor[7], John R Stephen[8], Alan Cooper[1]

[1]University of Adelaide (Australia), [2]Australian Genome Research Facility (Australia), [3]Victor Chang Cardiac Research Institute (Australia), [4]University of New South Wales (Australia), [5]University of Adelaide (Australia), [6]University of Adelaide (Australia), [7]University of Missouri (United States), [8]University of Adelaide (Australia)

Epigenetics encompasses a suite of mechanisms that potentially enable the adaptation of species to rapidly changing environments. We propose that mammal populations from the Quaternary-the current geological period characterised by dramatic climate oscillations-represent a unique model to study epigenetic responses to environmental cues, and their role in adaptation and extinction. Statistical methods have recently been developed to infer the methylation status of cytosines from ancient mammalian genome datasets, albeit at a relatively low resolution. On the other hand, experimental studies of ancient DNA methylation have been restricted to a limited number of target loci and a small sample size, due to pronounced DNA degradation and low levels of endogenous DNA in sub-fossil remains. Here, we present a method to perform whole-genome bisulphite sequencing of ancient DNA extracts. To demonstrate the power of this method, we characterised methylomes at a single nucleotide resolution using 10 extinct and 14 modern bison samples from North America, which span a time range of more than 50,000 years that includes key climate cooling and warming events. Amongst all identified differentially methylated regions, an intron of the *MET* oncogene shows consistent differential methylation between extinct and modern bison samples. Our method provides a unique opportunity to study the methylomes of extinct mammals at an unprecedented level of resolution.

POB-269

Differences in open chromatin states reveal heterogeneous dosage effects along the Z-chromosome

Ana Catalan[1, 2]

[1]Ludwig Maximilians University (Germany), [2]Uppsala University (Sweden)

In most birds where dosage compensation has been assessed, it is usually found an absence of complete Z-chromosome dosage compensation in the heterogametic sex, where the female's Z-chromosome expression is usually significantly lower than the autosomes and the male's Z-chromosome. These conclusions have often been drawn from chromosome wide mean expression values between the Z and the autosomes. Here we use the European crow to assess dosage compensation in a locus specific manner. We generated ATAC-seq and RNA-seq data from spleen, liver, and gonads and characterized chromatin state differences between compensated and not compensated loci. When interrogating dosage compensation along each chromosome we were able to identify those genes that are subjected to compensation, but also the degree of which other loci escape compensation. We also found that the level of dosage compensation across the three organs tested variates as well as the chromatin states between compensated and uncompensated loci. For those genes identified as being dosage compensated, we observe that flanking genomic regions show higher chromatin accessibility when compared to uncompensated genes. We hypothesize that more accessible open chromatin regions facilitate the access of dosage compensation complexes only at specific loci. Whether such complexes are present in the European crow system is yet to be elucidated. With this study, we show that although birds do not show a pattern of global dosage compensation, there are some loci that do show dosage in the Z, a phenomenon that might be link to gene functionality or influenced by gene-gene interactions.

POB-270

Drought stress-specific DNA methylation differences found in tolerant and susceptible maize varieties.

Ryan R Morrison[1], Robin G. Allaby[1], Logan Kistler[2]

[1]University of Warwick (United Kingdom), [2]National Museum of Natural History, Smithsonian Institute (United States)

Drought has a devastating effect on maize (*Zea mays*) crop yield and causes of ~70 percent of all yield losses in crop plants worldwide. Unpredictable climate change conditions have created an increased demand to understand key processes behind drought stress adaptability in maize. Methylation is an epigenetic mechanism that controls gene expression and plays an important defensive role when adapting to environmental stress. Here we aim to analyse epigenomic maize data from a controlled stress experiment to look for drought stress-specific methylation patterns. We hypothesise that during drought stress, methylation events may be observed near resistance-specific genes in only drought-tolerant maize. In addition, we predict that drought stress will cause more methylation events in drought-susceptible maize than in drought-tolerant. We performed whole genome bisulphite sequencing on leaf and root samples taken from drought-tolerant B76 and drought-susceptible B73 maize plants under control and experimental drought conditions. Using a newly developed strategy for assessing methylation changes in transposable element environments and established methods for analyzing differentially methylated regions, we were able to interrogate the methylation landscape before and after drought stress in genic, regulatory, and transposable fractions of the B73 and B76 genomes. It also revealed interesting methylation pattern differences between the two varieties.

POB-272

Amphioxus functional genomics reveals the evolution of vertebrate regulatory traits

Ignacio Maeso[1], Ferdinand Marletaz[2, 10], Panos N. Firbas[1], Juan J. Tena[1], Ozren Bogdanovic [3, 4, 5], Malcolm Perry[6, 7], Elisa de la Calle-Mustienes [1], Stephanie Bertrand[9], Boris Lenhard[6, 7], Peter W.H. Holland[2], Ryan Lister[5], Hector Escriva[9], Manuel Irimia[8], Jose Luis Gomez-Skarmeta[1]

[1]Centro Andaluz de Biologia del Desarrollo (CABD), CSIC-Universidad Pablo de Olavide-Junta de Andalucia, Seville (Spain), [2]Department of Zoology, University of Oxford (United Kingdom), [3]Genomics and Epigenetics Division, Garvan Institute of Medical Research, Sydney (Australia), [4]St Vincents Clinical School, Faculty of Medicine, University of New South Wales, Sydney (Australia), [5]The University of Western Australia, Crawley (Australia), [6]Institute of Clinical Sciences, Faculty of Medicine, Imperial College London (United Kingdom), [7]Computational Regulatory Genomics, MRC London Institute of Medical Sciences (United Kingdom), [8]Centre for Genomic Regulation (CRG), Barcelona Institute of Science and Technology (BIST) (Spain), [9]Sorbonne Universites, UPMC Univ Paris 06, CNRS, Biologie Integrative des Organismes Marins (BIOM), Observatoire Oceanologique (France), [10]Molecular Genetics Unit, Okinawa Institute of Science and Technology (Japan)

All chordates share a fundamental bodyplan that was greatly elaborated in vertebrates. Vertebrates also evolved highly distinctive genomes, sculpted by two whole genome duplications (WGD) and the acquisition of unique genomic traits. These traits include numerous features often associated with gene regulation, such as unusually large intergenic and intronic regions, a distinct set of highly conserved non-coding regions, and high global 5-methylcytosine (5mC) content and 5mC-dependent regulation of transcriptional enhancers. To investigate the evolution of these features and genome regulation in chordates, we characterized promoters, methylation, chromatin accessibility, histone modifications and transcriptomes in multiple tissues and throughout development of the cephalochordate amphioxus, a slow-evolving non-vertebrate chordate. These data revealed an intermediate stage in the evolution of differentially methylated enhancers, and high conservation of gene expression and its underlying *cis*-regulatory logic between amphioxus and vertebrates, maximally at a developmental phylotypic period. We also unraveled a strong increase in the complexity of vertebrate gene regulatory landscapes, specially among genes retained in several copies after the two vertebrate WGDs. Altogether, these data pave the way for a better understanding of the regulatory principles underlying key vertebrate innovations.

POB-273

Epigenetic mutations contribute to specialized metabolites variation in Arabidopsis thaliana accessions

Kazumasa Shirai[1], Fumio Matsuda[2, 3], Ryo Nakabayashi[2], Mitsuhiko P Sato[4], Masanori Okamoto[2, 5], Maho Tanaka[2], Akihiro Fujimoto[6, 7], Minami Shimizu[2], Kazuo Shinozaki[2], Motoaki Seki[2], Kazuki Saito[2, 8], Kousuke Hanada[1, 2]

[1]Kyushu Institute of Technology (Japan), [2]RIKEN (Japan), [3]Osaka University (Japan), [4]Kyushu University (Japan), [5]Utsunomiya University (Japan), [6]Kyoto University (Japan), [7]RIKEN (Japan), [8]Chiba University (Japan)

Plant specialized metabolites, so-called secondary metabolites, do not affect the growth and development in plant but are essential for surviving in species-specifc environments. The functional roles of specialized metabolites are associated with species-specific functions such as antibacterial action, defending plants against insects and providing flowers with colors. Therefore, it is well known that specialized metabolites are closely related to species-specific evolution in plant. Also, a large variations of specialized metabolites within a single species have been recently reported in several plant species. Our previous study showed that the metabolite variations within a single species contributed to local adaptation in *Arabidopsisthaliana* which is the model organism in plants (Shirai et at al., 2017, MBE). We furthermore found that genes associated with specialized metabolites tend to be controlled by the expressional variations in associated genes. Expressional variation can be caused by either DNA substitution mutation or DNA methylated mutation. Here, we identified single methylation polymorphism (SMP) sites among *A. thaliana*accessions. Then, we revealed that DNA methylated mutations contributed to the variations in gene expression causing the specialized metabolite variations in a higher rate than DNA substitution mutations.

POB-274

Epigenome-based Adaptive Evolution in Unicellular Bacteria: Evidence from Genomes, Methylomes, Transcriptomes and Phenotypes

Ichizo Kobayashi[1, 2, 3]

[1]Kyorin University, School of Medicine (Japan), [2]University of Paris-Saclay (France), [3]University of Tokyo (Japan)

The currently dominating model for adaptive evolution assumes selection from diverse genome sequences. A contrasting possibility is that extremely diverse *epigenomes*provide units of evolution just as in cell differentiation in multi-cellular organisms. Indeed there are increasing lines of evidence for trans-generation epigenetic inheritance in eukaryotes. We explore this *epigenome-based adaptive evolution* in unicellular bacteria, in which a somatic cell can be equated with a germ-line cell.

We focus on *Helicobacter pylori*, eubacterium responsible for stomach cancer. Each strain carries a large and unique repertoire of restriction-modification (RM) systems, which are bacterial immune systems distinguishing *self* and *non-self* by epigenetic sequence-specific DNA methylation. We earlier found their attack on host bacteria (autoimmunity) can increase their relative frequency (fitness) and proposed and demonstrated that they are selfish mobile elements. Target sequence recognition domains (TRDs) of several RM types move within a gene (locus), between genes (loci) and across taxonomic barriers, changing sequence-specificity in DNA methylation.

Many of *H. pylori* RM systems frequently change and generate millions of diversity in DNA methylome. Because DNA methylation locally affects gene expression, the ever-changing methylome may change transcriptome and adaptive phenotype sets. Our works involving gene knocking out, genome decoding, methylome decoding (PacBio), transcriptome analysis and phenotype analysis (motility etc.) support this concept. The emerging gene-regulation network with many DNA methyltransferases as hubs continues metamorphosing through changes in their sequence specificity pursuing adaptation: a FIREWORK model.

https://www.ncbi.nlm.nih.gov/books/NBK63963/

Ichizo Kobayashi Google Scholar Citations

POB-275

Role of DNA methylation in Paternal Genome Elimination, a peculiar reproductive strategy found in insects

Stevie Anne Bain[1], Dominik R Laetsch[1], Laura Ross[1]

[1]University of Edinburgh (United Kingdom)

Paternal genome elimination (PGE) is a genomic imprinting phenomenon found among several insect species, including the citrus mealybugs (*Planococcus citri*). In sons, the paternally inherited set of chromosomes is silenced and then eliminated from the germline. This process is likely under the control of a maternal factor, since mothers benefit from producing sons that only transmit their genetic material to offspring. However, it is unclear how the paternal genome in males is targeted for silencing and elimination. PGE appears to be regulated by epigenetic machinery similar to that in mammalian silencing and imprinting (histone and DNA methylation). However, the molecular details remain poorly understood and the extent to which these mechanisms play direct roles in the recognition, transcriptional suppression and germline elimination of paternally inherited alleles is inconclusive.

Here we study both sex-specific and parent-of-origin specific DNA methylation in *Planococcus citri* using Whole Genome Bisulfite Sequencing and other molecular techniques. We show how the epigenetic landscape of this species differs between the sexes by analysing both levels and patterns of DNA methylation across the genome. Analysis of hybrid data is used to investigate how the parental origins of chromosomes are identified, allowing specific elimination of the paternal set in males. Combining this information with transcriptome data gives insights into how these epigenetic marks are directly involved in gene silencing. Finally, quantitative expression analyses of the genes involved in evolutionarily conserved pathways of transcriptional repression further examine the role of DNA methylation in establishing and maintaining PGE throughout development.

POB-276

Effect of pleiotropy on regulatory sequence evolution different than on gene evolution

Peter Orchard[1], Zane Kliesmete[1], Beate Vieth[1], Ines Hellmann[1]

[1]LMU-Munich (Germany)

In contrast to the evolution protein coding genes, for regulatory regions it is easier to evolve tissue specific functions and thus have the freedom to evolve away from pleiotropy. Here, we analyse DNase Hypersensitive Regions (DHR) from nine fetal tissues to measure tissue-specificity and INSIGHT (Gronau et al. 2013) to obtain estimates of the amounts of weak and strong negative selection within groups of DHS sites.

Naively, we would expect that both types of selection would be additive, and thus increase with the number of tissues in which a DHR was used. We show that this is true for weak selection, but surprisingly not for the fraction of sites under strong selection:. Ubiquitous and tissue-specific DHR appear more conserved than sites with intermediate tissue specificity, leading us to suspect large differences in effect size amongst the tissues.

Indeed, we find that brain usage exerts the strongest negative selection. 30% of all sites in brain specific DHR evolve under strong negative selection, while this number drops to 18% if the DHR is used in at least one other tissue, implying that selection in brain acts differently from all other examined tissues. However, brain effects alone could not explain the non-additive effect of pleiotropy on strong negative selection. Usage in thymus has a similarly strong, but opposite effect on strong negative selection: thymus-specific DHR show no strong selection and only little weak selection, suggesting that thymus usage is incompatible with strong conservation.

POB-277

Genome-wide and single-base resolution DNA methylomes of the Sea Lamprey (Petromyzon marinus) Reveal Gradual Transition of the Genomic Methylation Pattern in Early Vertebrates

Zhixi Su[1], Zhao Zhang[1], Zhan Zhou[2], Gangbiao Liu[1], Yangyun Zou[1], James Lloyd P. B.[3], David W. McCauley[4], Weiming Li[5], Xun Gu[6, 1]

[1]Fudan University (China), [2]Zhejiang University (China), [3]University of California, Berkeley (United States), [4]University of Oklahoma (United States), [5]Michigan State University (United States), [6]Iowa State University (United States)

The mechanism by which the genome methylation pattern transited from a mosaic to a global pattern in early vertebrates remain largely elusive, partly owing to the lack of methylome data from the species diverged during this stage. Here, we used the whole-genome bisulfite-sequencing technology to investigate the genome-wide methylation in three tissues (heart, muscle, and sperm) from the sea lamprey, an extant Agarthan vertebrate. Analyses of methylation level and the extent of CpG dinucleotide depletion of gene-encoding, intergenic and promoter regions revealed a gradual increase in the methylation level from invertebrates to vertebrates, with the sea lamprey exhibiting an intermediate position. In addition, the methylation level of the majority of CpG_s was intermediate in each sea lamprey tissue. The methylation features in different genomic regions, such as the TSS flanking regions, exon-intron boundaries, transposons, as well as genes grouping with different expression levels, supported the gradual methylation transition hypothesis. We further analyzed the cell-to-cell DNA methylation variability, based on comparing the methylation status of adjacent CpG sites on individual sequencing reads, and found a high level of heterogeneity of methylation status between individual cells. BSP following with sequencing of 6 randomly selected regions also consistent with our findings. In this regard, we defined the genomic methylation pattern of sea lamprey as "global genomic DNA intermediate methylation". These findings demonstrate an intermediate genomic methylation pattern between invertebrates and jawed vertebrates, providing evidence that supports the hypothesis that methylation patterns underwent a gradual transition from invertebrates (mosaic) to vertebrates (global).

POB-278

Regulatory Differences in Natal Down Development between Altricial and Precocial Birds

Chih-Kuan Chen[1], Pei-Chi Su[2], I-Ming Liu[2], Hao-Fen Chuang[1], Wen-Hsiung Li[1], Chen Siang Ng[2]

[1]Academia Sinica (Taiwan), [2]National Tsing Hua University (Taiwan)

Birds can be classified as altricial or precocial. The hatchlings of altricial birds are almost naked and relatively immobile, whereas those of precocial birds are covered with natal down and quickly mobile. The precocial developmental mode is considered the ancestral state. Such evolutionary divergence is thought to reflect environmental adaptation, but the molecular basis of nakedness is unclear. We have characterized that altricial zebra finch has two types of feather formation (Type I in which the feather buds were suppressed, and Type II in which the feather buds develop in to natal down), whereas precocial chicken only has Type II feather formation. We identified that, in zebra finch, the FGF/MAPK signaling pathway is involved in natal down growth suppression. Nakedness of altricial birds should have evolved more than once, raising the question of whether convergent phenotypes can be made by similar genes or pathways. We extend the study to profile the expression patterns of the candidate genes in early feather development of pigeon (semialtricial), and budgerigar (altricial), using whole mount *in situ* hybridization. By comparing the expression patterns of *SHH, CTNNB1, SNAI1, and TWIST2* in those species, we found that the gene expression patterns are conserved among Type II feather formations, but are different among Type I feather formations in different species. Our preliminary data suggests that the suppression of natal down development might be from different molecular mechanisms among altricial birds and we plan to further characterize the transcriptomes between Type I and II feather formations in budgerigar.

POB-279

Assessing epigenetic effects of a lifestyle transition in human history

Olaf Thalmann[1], Kristian Hanghoej[2, 3], Karl-Heinz Herzig[1, 4], Jaroslaw Walkowiak[1], Ludovic Orlando[3], Mattias Jakobsson[5]

[1]Poznan University of Medical Sciences (Poland), [2]Natural History Museum of Denmark, University of Copenhagen (Denmark), [3]Universite de Toulouse, CNRS, Universite Paul Sabatier (France), [4] Institute of Biomedicine, Oulu University (Finland), [5]Uppsala University (Sweden)

The emergence of new molecular tools has allowed an intimate view into the genomes of any living organism including our own, and has started to uncover comprehensive epigenomic landscapes. This knowledge can help to directly link molecular footprints with the adaptation to lifestyle changes in the past. One epoch with important alterations in human lifestyle is the Neolithic, where human populations transitioned from hunting-gathering to farming, thus setting the stage for ever increasing populations and the emergence of novel diseases.

We assessed whether this transition was accompanied by shifts in epigenetic profiles by exploiting patterns of post-mortem DNA degradation within methylated and non-methylated CpGs of two high-quality genomes of humans living in close geographic and temporal proximity in ancient Scandinavia. The retrieved methylation patterns largely overlapped those described in modern humans, thereby substantiating the validity of our methodological approach. Aside from the average genome-wide coverage another, sample-specific bias was found to affect the power to reconstruct ancient methylation maps, strongly advocating the necessity of a standardized sampling whenever assessing population-wide changes in ancient methylomes. We further uncovered differentially methylated regions of functional relevance, providing a first glimpse into the epigenetic mechanisms accompanying this crucial transition in human history. Given the data at hand, we propose analytical guidelines of relevance for future paleo-epigenetic studies. We are confident that the ongoing extension to a population-wide sampling of Neolithic Scandinavians will allow generating a comprehensive picture of epigenetic shifts that accompanied the Neolithic Revolution, and deduce its impact on modern life.

POB-280

A novel model for the evolution of mammalian epigenomic traits

Noah Ethan Dukler[1, 2], Yi-Fei Huang[1], Adam Siepel[1, 2]

[1]Cold Spring Harbor Laboratory (United States), [2]Weill Cornell Medical College (United States)

With the proliferation of high-throughput functional genomics assays (ie. ChIP-seq, ATAC-seq, PRO-seq, etc.), there has been an explosion in the amount of available epigenetic data within and between multiple species. Comparative analysis of epigenomic data provides the opportunity to improve our understanding of the evolution of regulatory elements. A wide variety of phylogenetic tools (i.e. phastCons, GERP++, and dLess) have been developed to interrogate the forces that shape the evolution of genomic sequences and have been useful in identifying loci associated with disease, development, and molecular phenotypes. However, no similar methods exist for analyzing comparative epigenomic data. Here we propose a new phylogenetic hidden Markov model, epi-Phylo, for reconstructing the evolutionary histories of enhancers and promoters from comparative ChIP-seq data.

Epi-phylo combines a phylo-HMM with a negative binomial error model to jointly infer the location of regulatory elements and their evolutionary trajectories from noisy data. We apply epi-Phylo to existing datasets (Villar et al. 2015 & Berthelot et al. 2017) and show that sequence and epigenetic conservation are positively correlated. We also investigate whether specific regulatory pathways undergo accelerated turnover of regulatory elements and estimate the prevalence of compensatory turnover of regulatory sequences. In conclusion, epi-Phylo is a rigorous statistical phylogenetic model of epigenomic evolution, enabling the inference of explicit evolutionary histories for a variety of regulatory elements.

POB-281

Genomic Landscape of Methylation Islands in Hymenopteran Insects

Hyeonsoo Jeong[1], Xin Wu[1], Brandon Smith[1], Soojin V. Yi[1]

[1]Georgia Institute of Technology (United States)

DNA methylation is well characterized in mammals, but methylation in insects and other invertebrates has only recently been brought into the foreground of epigenomics. Recent genome-wide DNA methylation analyses of insect genomes have revealed an intriguing contrast from compared to those in mammals. In mammals, most CpGs are heavily methylated, with the exceptions of clusters of hypomethylated sites referred to as CpG islands. In contrast, DNA methylation in insects is localized to a small number of CpG sites. Here, we refer to clusters of methylated CpGs as 'methylation islands (MIs)', and characterize them in seven hymenopteran insect genomes with high quality bisulfite sequencing data. Methylation islands were primarily located within gene bodies. Methylated CpGs within MIs are highly evolutionarily conserved compared to methylated CpGs outside MIs or non-methylated CpGs. Additionally, genes harboring MIs exhibit consistently higher level of expression compared to those do not harbor MIs, even after considering their level of DNA methylation. These results provide insights into the relationship between DNA methylation, evolutionary conservation, and gene expression.

POB-282

Methylome and Transcriptome Reprogramming associated with Wolbachia in a parasitoid wasp

Xin Wu[1], Amelia Lindsey[2], Dan Sun[1], Paramita Chatterjee[1], Jack Werren[3], Richard Stouthamer[4], Soojin V Yi[1]

[1]Georgia Institute of Technlogogy (United States), [2]Indiana University Bloomington (United States), [3]University of Rochester (United States), [4]University of California, Riverside (United States)

We explore methylomic and transcriptomic responses to *Wolbachia*-induced parthenogenesis in the parasitoid wasp *Trichogramma pretiousum*. In this host-symbiont relationship, *Wolbachia* infects female wasps and arrests egg development in Meiosis I, enabling the egg to remain diploid and develop into viable adults, thus ensuring vertical transmission of *Wolbachia*. We have observed substantial genome-wide methylation changes due to *Wolbachia* infection in pathways related to early developmental processes such as meiosis, cell division, and egg development. Incorporating RNA-sequencing data, there is little overlap between differentially methylated genes (DMGs) and differentially expressed genes (DEGs) *per se*. DMGs and DEGs have contrasting characteristics in terms of gene body methylation and gene length. Interestingly, most of the DEGs were lineage specific genes that were not enriched for any Gene Ontology terms. Our results describe *Wolbachia*'s wide-ranging effects in its hosts. By linking the methylome and transcriptome profiles *Wolbachia*-infected wasps, we gain a clearer understanding of the molecular basis of parthenogesis and symbiont-mediated asexuality.

POB-283

Epigenetic ramifications of the trauma of enslavement, centuries of chattel slavery, and institutionalized racism

Fatimah Linda Collier Jackson[1]

[1]Howard University (United States)

The historic large-scale capture, forcible kidnapping, and subsequent forced labor exerted tremendous stress on the biology of enslaved Africans. This stress provided the substrate for natural selection in New World African populations. The conditions of enslavement, post-emancipation reconstruction, and state-sanctioned racism have had an enduring impact on New World Africans and their descendants. Using the US as a case study for chronic stress associated with enslavement, we observe elevated circulating cortisol levels, increased depression, and increased frequency of post-traumatic stress disorder in contemporary African Americans. Chronic food deprivation and food instability during the period of slavery and its aftermath are thought to have further exacerbated the long-term negative implications of enslavement and institutionalized racism. We also observe significance evidence of resilience and adaptive survival mechanisms emerging that ameliorated many of the disadvantageous consequences of sub-optimal environments. The epigenome allows monitoring of changes in DNA menthylation patterns (and concomitant changes in gene expression patterns) without alterations in nucleotide sequences associated with specific sources of trauma (e.g., confinement, forced labor, inadequate nutrition, physical abuse, psychological distress). We report on the developmental effects during critical periods in the lifecycle (e.g., prenatal, early childhood, adolescence, adulthood) and the epigenetic effects of this trauma on specific organs and organ systems (e.g., brain injury, damage to bone and muscle, nutritional deficiencies, confinement and restraint distress, substance abuse, and pharmacogenomics vulnerabilities.) Finally, we report on the evidence for resilience in immune function and adaptive psychological coping mechanisms in descendant populations.

POB-286

Understanding the molecular diversification of self recognition through ray-finned fish innate immune receptor families

Dustin Wcisel[1], Jeffrey Yoder[1], Alex Dornburg[2, 1]

[1]College of Veterinary Medicine North Carolina State University (United States), [2]North Carolina Museum of Natural Sciences (United States)

The use of molecular markers of self-identity as a basis for immunity was a major evolutionary innovation in the early history of vertebrates. The ability to discriminate between self and non-self has since spurred a coevolutionary competition between hosts and pathogens driving high levels of both inter- and intraspecific immune gene sequence diversity. Such diversification is essential for a species to survive new pathogens, yet the origin and evolutionary dynamics of vertebrate self recognition remain poorly understood. A powerful system for understanding the genetic and functional evolution of immune genes associated with self recognition are ray-finned fishes (Actinopterygii) which constitute over half of all living vertebrates. Like all vertebrates, fish possess certain core immune gene families, however they also encode a number of "fish-specific" immune gene families. Using a phylogenetic comparative framework, we integrate transcriptome and genomic sequence data from early diverging lineages of ray-finned fishes to establish the evolutionary origins of fish-specific immune receptor genes and test fundamental expectations of coevolution between markers of self and their candidate receptors. Our work provides a new perspective on the early history of the vertebrate immune system, overturning assumptions of teleost specific innovations, while revealing novel molecular innovations to pathogen resistance.

POB-288

Studying selection in 'real-time' by genotyping HLA immune genes from ancient DNA

Federica Pierini[1], Marcel Nutsua[2], Lisa Boehme[2], Almut Nebel[2], Ben Krause-Kyora[2], Tobias L. Lenz[1]

[1]Max Planck Institute for Evolutionary Biology (Germany), [2]Kiel University (Germany)

The highly polymorphic genes of the human leukocyte antigen (HLA) system play a key role in adaptive immunity. Past and ongoing pathogen-mediated selection is proposed to be one of the major factors affecting the genetic variability at those genes. Selection at the HLA genes is a dynamic process that involves parallel mechanisms acting at different time scales and creating an intriguing combination of shared polymorphism but distinct allele pools among populations and possibly even species.

The recent development of genomic tools for the analysis of ancient DNA (aDNA) provides a unique opportunity to unravel the selection processes shaping the human genome. In this light, the investigation of ancient HLA genes in historical populations could shed light on mechanisms of pathogen-mediated selection in humans. However, HLA genes exhibit exceptional genetic variability that defies standard sequencing and assembly approaches. To overcome this obstacle, a novel DNA capture approach, optimized for short aDNA fragments, in combination with an aDNA-optimized pipeline is here being applied to analyze HLA polymorphism in historical human populations.

We show the importance of a reliable HLA genotyping pipeline for ancient DNA. The pipeline has already been applied successfully to a dataset of ancient samples, linking HLA variability with susceptibility to leprosy, and can be further applied to explore HLA allele frequency changes through time when ancient temporal sample series are available.

POB-289

Variation of the MHC and other immune receptors in wild zebrafish

Jaanus Suurvaeli[1], Thomas Wiehe[1], Maria Leptin[1, 2]

[1]University of Cologne (Germany), [2]European Molecular Biology Laboratory (Germany)

The teleost *Danio rerio* (zebrafish) is a model organism widely used to study the mechanisms of disease and development. Even after ~60 years of research, surprisingly little is known about the evolution of the zebrafish immune system itself. In the last decade it has become increasingly apparent that its genome has undergone massive expansions of small Ig-based receptors, NACHT-LRR receptors (NLRs) and finTRIM-like genes. This process could even still be ongoing. In addition, class I and II antigen presenting genes are dispersed across the genome, with some regions considered to be the core MHC. The zebrafish MHC genes are associated with copy number variation and even rare null haplotypes have been described.

We have adapted a PacBio-based target enrichment protocol (RenSeq-PacBio) in order to characterize some of these immune gene families within different zebrafish populations. We have retrieved extensive sequence data, from both lab strains and wild populations, for exons of the class I and II MHC genes, for the Toll-Like Receptors, and for the ~600 copies of B30.2 domain. The B30.2 domain is commonly found in the butyrophilins and TRIMs of the MHC region; fish have attached it to not only finTRIM-like proteins but also to the NLRs. Finally, we are able to distinguish allelic variation from actual gene copies for the ~400 near-identical NACHT exons of the NLRs. Our dataset provides a high resolution overview of sequence variation within large immune gene families of wild zebrafish.

POB-290

Evolution of type I interferon and receptor genes in crocodilians

Katherine Brittain[1], Elizabeth A. Jones[1], David Ray A.[2], Jaime Gongora[1]

[1]The University of Sydney (Australia), [2]Texas Tech University (United States)

Crocodilians appear to have an effective innate immune system that allows them to cope with a variety of pathogenic challenges, although knowledge of the components that modulate the functions of the immune systems, such as interferons (IFNs), is limited. IFNs encode for molecules that induce a variety of cellular responses to viral and bacterial pathogens as part of both the innate and adaptive immune response in all vertebrates. However, their distribution and evolution in crocodilians remains unclear. To address this, we characterised type I IFN and their receptor genes in four published and 10 unpublished crocodilian genome sequences representing the families Alligatoridae (alligators), Crocodylidae (crocodiles) and Gavialidae (gharials). We identified four type I IFN genes and two receptor genes in these genome datasets. Comparative analysis of these genes showed highly conserved amino acid residues (>90% identity) between crocodilian species, for all genes. Type I IFNs showed highly conserved motifs and similar intron/exon patterns between crocodilians and other amniotic species. Selection tests showed these loci have evolved under purifying selection. Our analyses also suggest that some type I IFN genes emerged before crocodilian families diverged from the common ancestor, with some genes arising before the avian/crocodilian split while others appear to have emerged only in crocodilians as a result of gene duplication.

POB-291

Evolution and diversity of MHC class II DQB1 and DRB1 in wild pigs and peccaries

Carol Lee[1], Alvaro Perdomo[1, 2], Claire Rogel-Gaillard[3], Jaime Gongora[1]

[1]The University of Sydney (Australia), [2]University of Hohenheim (Germany), [3]Universite Paris-Saclay Jouy-en-Josas (France)

The major histocompatibility complex (MHC) plays an essential role in vertebrate adaptive immunity and is considered a key genomic model region for understanding gene family evolution and the co-evolution between host and pathogens. In Suidae and Tayassuidae (Suoidea), studies of MHC class II genes have focused on the domestic pig, but has been limited in wild species. Given the different evolutionary patterns observed between the class I and II MHC genes within domestic pigs, this provides an opportunity to further understand the evolution of class II genes across wild pigs and peccaries. To address this, we investigated the MHC class II DQB1 and DRB1 exon 2, encoding the antigen binding region, in 12 species of wild pigs and peccaries. Our preliminary analyses suggest a stronger balancing selection in DRB1 compared to DQB1. We also identified shared and unique alleles between pigs and peccaries indicating that some alleles have been retained since their divergence from the common ancestor ~35Ma, while other alleles may have emerged as species-specific adaptations. These findings provide an avenue to a more in-depth understanding of how these alleles are maintained between wild species of these families.

POB-292

Worldwide genetic variation of the IGHV and TRBV immune receptor gene families in humans

Jane A Yu[1], Shishi Luo[1, 2], Yun S Song[1, 2]

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

The immunoglobulin heavy variable (IGHV) and T cell beta variable (TRBV) loci are among the most complex and variable regions in the human genome. Generated through a process of gene duplication/deletion and diversification, these loci can vary extensively between individuals in copy number and contain genes that are highly similar, making their analysis technically challenging. Here, we present a comprehensive study of the functional gene segments in the IGHV and TRBV loci, quantifying their copy number and single nucleotide variation in a globally diverse sample of 109 (IGHV) and 286 (TRBV) humans from over a hundred populations. We find that the IGHV and TRBV gene families exhibit starkly different patterns of variation. In particular, with hundreds of copy number haplotypes (instances that have differences in the number of functional gene segments), the IGHV locus has undergone more frequent gene duplication/deletion compared to the TRBV locus, which has only a few copy number haplotypes. In contrast, the TRBV locus has a greater or at least equal propensity to mutate in non-duplicated genes, as evidenced by greater nucleotide variation, compared to the IGHV locus. Thus, despite common molecular and functional characteristics, the genes that comprise the IGHV and TRBV loci have evolved in strikingly different ways. As well as providing insight into the different evolutionary paths the IGHV and TRBV loci have taken, our results are also important to the adaptive immune repertoire sequencing community, where the lack of frequencies of common alleles and copy number variants is hampering existing analytical pipelines.

POB-294

Understanding how Japanese frogs are resistant to a deadly worldwide fungal disease through in silico analyses of major histocompatibility complex (MHC)

Quintin Lau[1], Takeshi Igawa[2], Tiffany Kosch[3], Lee Skerratt[3], Lee Berger[3], Alex Roberts[3], Yoko Satta[1]

[1]Sokendai (Japan), [2]Hiroshima University (Japan), [3]James Cook University (Japan)

Japanese frogs appear to be resistant to chytridiomycosis, a disease caused by the Bd fungus that is decimating amphibians worldwide, yet the genetic basis for resistance to chytridiomycosis is still unclear. One potential contributor to this resistance is variable MHC proteins, which can bind and present specific antigenic peptides to the immune system. However, the function and mechanism of frog MHC in the face of Bd is poorly understood and requires further study. We generated transcriptome data from spleen of 12 common species of Japanese endemic frogs, covering eight genera and four families (Ranidae, Bufonidae, Hylidae, and Rhacophoridae). From this, we collected coding sequence for MHC class I and class II in all species, including functionally important peptide-binding domains. We also collated published sequences of MHC from susceptible and resistant frogs across the world. We conducted investigation of MHC function using *in silico* supertyping analyses and computer-based peptide binding predictions. Supertyping analyses of MHC class II variants from Japanese frogs and other resistant frogs identified possible physiochemical properties of MHC that may be important for recognizing and binding chytridiomycosis-related antigens. We also present preliminary application of computer-based algorithms to predict what Bd peptides are bound by MHC proteins from resistant or susceptible frogs, which will further our understanding of MHC-chytridiomycosis dynamics.

POB-295

Survival of the Fastest not the Fittest: Modeling the effect of multilevel immune selection on B cell lineages in adaptive immune system population dynamics

Maksim Shestov[1, 4], Latifa F Jackson[2, 3], Jason H Moore[4]

[1]University of Pennsylvania (United States), [2]Howard University (United States), [3]Howard University (United States), [4]University of Pennsylvania (United States)

Survival of the fittest trait is a central tenet in evolutionary biology models. Nowhere is this more invoked than in evolution effects of immune response. Meanwhile adaptive immune systems are governed by molecular rules of clonal expansion. Studies on the immunogenic CDR region in maturing B cells have described the mechanistic underpinnings of affinity maturation in B cell lineages. The evolutionary fate of these lineages determines host immune fates. We sought to model and characterize the evolutionary outcomes of multi-level selection for an influenza-like infection. Affinity maturation can occur by A) CDR targeting with codon bias, B) uniform mutation with codon bias, C)targeting with no codon bias or D)uniform mutation with no codon bias B cell populations, a non-vectored SIR infection simulation was modified with the program Netlogo. Host populations (N=1500) were seeded in each immune scenario with a 0.01 infection probability and restricted movement. We show that individuals governed by a codon bias model have an evolutionary advantage over either no codon bias B cell populations or uniform mutation with no codon bias B cell populations in our disease model. Together these analyses show that neither connectivity nor clearance time can explain the increased effectiveness of the codon bias lineage in clearing infection. These findings suggest that it is not the individuals with the highest B cell affinity or fitness that will predominate over evolutionary time but instead those host individuals whose affinity maturation process is able to most rapidly clear infection in an influenza like model.

POB-296

Gene expression variability across cells and species shapes the innate immune response

Tzachi Hagai[1, 2], Sarah Teichmann[1, 2]

[1]EMBL - EBI (United Kingdom), [2]Wellcome Trust Sanger Institute (United Kingdom)

As the first line of defence against pathogens, cells mount an innate immune response, which is highly variable from cell to cell. The response must be potent yet carefully controlled to avoid self-damage. How these constraints have shaped the evolution of innate immunity remains poorly understood. Here, we characterise the transcriptional divergence of this programme between species and expression variability across cells. Using bulk and single-cell transcriptomics in primate and rodent fibroblasts challenged with an immune stimulus, we reveal a striking architecture of the innate immune response. Rapidly diverging genes, including cytokines and chemokines, vary across cells and have distinct promoter structures. Conversely, genes involved in response regulation, such as transcription factors and kinases, are transcriptionally conserved between species and display low cell-to-cell variability. We suggest that this unique expression pattern, observed across species and conditions, has evolved as a mechanism for fine-tuned regulation, to achieve an effective but balanced immune response.

POB-298

Testing Hypotheses on the Evolution of Resistance to Amphibian Chytridiomycosis

Minjie Fu [1], Bruce Waldman[1]

[1]Seoul National University (Republic of Korea)

Global amphibian population declines have been attributed to the spread of the chytrid fungus Batrachochytrium dendrobatidis (Bd), which results in morbidity and mortality in susceptible individuals. Asian amphibians, however, appear largely immune to the disease. These species present certain conformations of the MHC-IIβ domain that appear to decrease susceptibility to chytridiomycosis. Resistant species on other continents also bear alleles encoding this conformation, and differential survival of individuals with these resistance alleles has been demonstrated both in comparative and experimental studies. In Asia, Bd has co-evolved with its amphibian hosts over hundreds or possibly thousands of years, and we have uncovered the highest diversity of endemic lineages of the pathogen found anywhere in the world. These unique Korean Bd (KBd) lineages, distinct from global pandemic lineages (BdGPL), may have evolved to be less virulent to their amphibian hosts, explaining, at least in part, why population crashes attributable to this disease have not been observed in this part of the world. To test this hypothesis, we infected disease-free subjects of a known susceptible Australian species, Litoria caerulea, and Korean native species, Bufo gargarizans, Bombina orientalis, and Hyla japonica, to KBd, BdGPL, or a sham inoculate. All infected Litoria caerulea succumbed both to KBd and BdGPL, whereas Korean amphibians survived both treatments. Infection was further confirmed by time-course qPCR and histology. Thus, we conclude that KBd is just as virulent as BdGPL in susceptible species. This supports our previous work demonstrating the evolution of resistance to the disease based on adaptive immunity.

POB-299

The majority of novel protein coding genes identified through phylostratigraphy alone are either old genes or recent duplicates

Claudio Casola[1]

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

The evolution of novel protein coding genes from non-coding regions of the genome is one of the most compelling evidence of genetic innovations in nature. One popular approach to identify de novo genes is phylostratigraphy, which consists in determining the approximate time of origin (age) of a gene based on its distribution along a species phylogeny. Several studies have revealed significant flaws in determining the age of genes, including de novo genes, using phylostratigraphy alone. However, the rate and determinants of false positives in de novo gene surveys based on phylostratigraphy remains unknown. Here, I re-analyze the findings from two studies that identified tens to hundreds of rodent-specific de novo genes adopting an exclusively phylostratigrapic approach. Using a combination of synteny information and sequence similarity searches I show that about 75% of putative de novo genes in rodents represent either share homology with genes from other mammals or derive from gene duplication. Conversely, very few false positives were found in a third study on mouse de novo genes that integrated phylostratigraphy with sequence evolutionary history. Proteins encoded by validated de novo genes showed lower propensity to form aggregates compared to older genes, in line with the preadaptation hypothesis of novel genes birth. More than half validated de novo genes overlap within older genes, suggesting that open chromatin conformation may facilitate the emergence of novel genes. These results highlight the importance of both synteny data and extensive homology analyses to accurately identify de novo genes and infer their evolutionary dynamics.

POB-300

Count does not recover major events of gene flux in real biological data

Nils Kapust[1], Shijulal Nelson-Sathi[2], Barbara Schonfeld[3], Einat Hazkani-Covo[4], David Bryant[5], Peter J. Lockhart[6], Mayo Roettger[1], Joana C. Xavier[1], William F. Martin[1, 7]

[1]Heinrich Heine University Duesseldorf (Germany), [2]Rajiv Gandhi Centre for Biotechnology (India), [3]University of Tasmania (Australia), [4]The Open University of Israel (Israel), [5]University of Otago (New Zealand), [6]Massey University (New Zealand), [7]Universidade Nova de Lisboa (Portugal)

In prokaryotes, lateral gene transfer generates new combinations of genes among chromosomes via known mechanisms (transformation, transduction, conjugation and gene transfer agents) during evolution. In the eukaryotic host lineage, descended from archaea, lateral gene transfer occurs from organelles to the nucleus at endosymbiotic events. Current genome analyses focusing on gene distributions have shown evidence for sporadic, discontinuous events of gene transfer from bacteria to archaea during evolution. Other studies investigating prokaryote genome evolution have used traditional birth-and-death phylogenetic models to claim that gene transfer to archaea was continuous during evolution, rather than involving occasional periodic mass gene influx events. Here we test the ability of Count and its birth-and-death analysis to recover known events of mass acquisition and differential loss using plastid genomes and eukaryotic protein families that were acquired from plastids. Count showed a strong bias towards reconstructed histories with gene acquisitions distributed uniformly across the tree. Several different acquisitions by plastid DNA were sometimes inferred for the same protein family. That is, Count recovered gradual and continuous lateral gene transfer among lineages, although massive gains followed by gradual differential loss is the probable true evolutionary process that generated the gene distribution data.

POB-302

A new method for identifying and splitting fusion proteins in prokaryotes

Constantin Eiteneuer[1], Julius Kliss[1], Mayo Roettger[1], Madeline C. Weiss[1], Michael Knopp[1], William F. Martin[1, 2]

[1]Heinrich-Heine-Universitaet Duesseldorf (Germany), [2]Universidade Nova De Lisboa (Portugal)

Gene fusion is a major contributor to the formation of multi-domain proteins, an important process in protein evolution. When it comes to clustering, proteins of promiscuous domain architecture interfere with the identification of homologous protein families. Here we adapted a previous method based on BLAST hits (Enright et al. 1999) to identify fusion proteins and their respective point of fusion. We used 19 million protein sequences of all available complete prokaryotic genomes (5,655 strains) of the RefSeq 2016 database to identify fusion candidates. Via the intersection with an additional method based on domain structure (Henry et al. 2016) we were able to identify 525,755 fusion candidates. A novel approach to calculate the most likely point of fusion, which can increase the precision of the level of support for a fusion candidate was applied to this dataset. The method relies on mapping the local coverage for the component-fusion BLAST hits (which identified the protein as a fusion) to the fused protein, enabling a sliding-window based analysis to determine the point of fusion. We tested the effect of splitting all fusion proteins on the re-clustering of a previous set of protein families.

Enright AJ et al. Protein interaction maps for complete genomes based on gene fusion events. Nature. 1999; 402:86-90.

Henry CS. et al. Systematic identification and analysis of frequent gene fusion events in metabolic pathways. BMC Genomics. 2016; 17:473.

POB-303

Comparative analysis of fusion and non-fusion proteins in prokaryotes based on functional annotation

Julius Kliss[1], Constantin Eiteneuer[1], Mayo Roettger[1], Joana Xavier[1], Madeline C. Weiss[1], William F. Martin[1, 2]

[1]Heinrich-Heine-Universitaet Duesseldorf (Germany), [2]Universidade Nova de Lisboa (Portugal)

Fusion proteins are created by the combination of two or more individual genes. As the fusion of two proteins leads to a decreased regulatory effort via coexpression, an analysis of the functional roles of fusion proteins can give hints about their evolution. Furthermore, the analysis of fusion proteins can lead to a better understanding of possible difficulties in protein clustering and annotation. There have been several attempts to identify fusion proteins, however these were hindered by low specificity and large computational costs. Here we have developed a new approach based on Henry et al. (2016) that uses non-overlapping conserved domains representing complete genes in protein sequences and multiple additional filtering steps. This method allowed the analysis of 19 million prokaryotic protein sequences of all available complete genomes of the RefSeq 2016 database. The results of the identification approach were intersected with the results of another method (Enright et al. 1999) to increase specificity. The COG annotation reveals that the categories of energy production and conversion and signalling related proteins are fused more often than proteins involved in transcription and translation. A more precise functional annotation is achieved with the larger and more up-to- date KEGG orthology, supporting the results of the COG annotations. In combination with a method to split the segments of a fusion, we also determined the functional relationship of fusion segments with the original fusion protein and their proximity within pathways.

POB-304

Understanding de novo gene evolution from random sequences expressed in E. coli

Johana Fajardo C.[1], Diethard Tautz[1]

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

De novo gene birth refers to the emergence of new genes from previously non-coding sequences, such as introns or intergenic regions. It has been shown in recent years that this mechanism of gene birth, previously thought to be very unlikely, is frequently found in nature. Examples of de novo genes have been identified and characterized in several different species of organisms, from bacteria, to plants and metazoans. There are many unanswered questions about how these genes arise from practically random sequences of nucleotides in the genome, including the frequency at which they appear, how does selection act upon them and how are they integrated into, sometimes critical, pre-existing regulatory and metabolic networks. A publication last year (Neme et al, 2017) made a first attempt to quantitatively study the effects of 150bp-long random sequences expressed in bacteria in vitro. The surprising, and somewhat controversial, results showed that a large fraction of the full-length random sequences increased in frequency in the bacteria population, hinting at a positive effect on bacterial fitness caused by their expression. In our project, we are delving further into the data generated from this, and similar experiments expanding the study to include shorter sequences, and analyses of selection coefficients to understand their actual effect on fitness. As a result, we are getting a better understanding of the dynamics of selection and evolution of de novo genes, and will contribute to the interesting discussion initiated by this publication.

POB-307

Phylogenomic Analyses of Brachyura Illuminates Ancient Origin of Freshwater Crabs and Recent Origins of Hydrothermal Vent Crabs

Ling Ming Tsang[1], Ka Yan Ma[1], Jing Qin[1], Tin-Yam Chan[2], Peter Kee Lin Ng[3], Ka Hou Chu[1]

[1]The Chinese University of Hong Kong (Hong Kong), [2]National Taiwan Ocean University (Taiwan), [3]National University of Singapore (Singapore)

The evolutionary history of the true crabs (infraorder Brachyura) remains a major research focus among carcinologists. Recent molecular phylogenies reveal that the Brachyura underwent rapid radiation during the Cretaceous and Early Tertiary, resulting in many long and deep lower branches that are difficult to resolve. Here we attempt to reconstruct the phylogeny of Brachyura using a phylogenomic approach, by sequencing the transcriptomes of 45 species (including 35 brachyuran families and two anomuran outgroups) using the Illumina platform. In combination with 16 other brachyuran transcriptomes from public database, we analyzed a final dataset with more than 60 brachyuran species from over 40 families. Phylogenetic analyses based on ~400 putative homologous loci (250,000 bp) infer a well resolved, robust phylogeny for the Brachyura. The phylogenomic tree provides further support to previous hypotheses of the early divergence of primary freshwater crabs. The phylogeny, however, suggests the freshwater crabs are more closely related to Thoracotremata and therefore the monophyly of Heterotremata was rejected. Most of the interfamilial relationships are well resolved in the current phylogenomic tree and illuminates evolutionary history of the group. On the other hand, the hydrothermal vent associated crab families Bythograeidae and Xenograpsidae are inferred to exhibit recent origins. We show that phylogenomic analyses could serve as a powerful tool for resolving relationships among the brachyuran lineages and future study will focus on obtaining transcriptomes from more taxa.

POB-309

Inferring cell differentiation processes based on phylogenetic analysis of genome-wide epigenetic information

Kanako O Koyanagi[1]

[1]Hokkaido University (Japan)

How cells divide and differentiate is a fundamental question in organismal development. Cellular differentiation proceeds with changing gene expression patterns, and these changes are recorded in the genome as epigenetic modification. Although understanding these processes is important, experimentally monitoring these processes comprehensively for various differentiating cell types during development is laborious and sometimes impossible. Phylogenetic analysis is typically used to reconstruct evolutionary processes of genes and organisms based on inherent genetic and morphological characters. Similarly, if developmental changes are inherited through cell proliferation and differentiation, the differentiation process within an individual could be reconstructed by phylogenetic analysis of these developmental changes. This study aims to examine whether phylogenetic analysis of epigenomes of adult differentiated cells can infer cell differentiation processes of differentiating cells. For this purpose, mammalian hematopoietic system was used as a model case, where genome-wide DNA methylation and histone modification data are available. Based on these data, phylogenetic trees and ancestral states of internal nodes, which correspond to differentiating progenitor cells, were inferred by maximum parsimony method. As a result, it was shown that the known hierarchical differentiation process and the genome-wide epigenetic states of progenitor cells could be inferred, which supports the validity for using epigenomic information for inferring cell differentiation processes at least for mammalian hematopoiesis. Phylogenetic features of each epigenetic modification during the differentiation process will also be discussed.

POB-310

Characterization and Molecular Phylogeny of Korean Chrysanthemum Species

So Youn Won[1], Jae-A Jung[2], Jung Sun Kim[1], Sangho Kang[1]

[1]National Institute of Agricultural Sciences (Republic of Korea), [2]National Institute of Horticultural and Herbal Science (Republic of Korea)

Improvement of DNA sequencing technology has helped reveal the whole genome of nucleus and chloroplast in many plant species, which serves as an important resource for basic and applied research. For example, the DNA sequence comparison characterizes molecular phylogenetic relationships and evolutionary processes of individuals or species. Currently, the compared sequences are expanded from the canonical barcode sequences such as rbcL in chloroplast or internal transcribed spacer (ITS) in nucleus to the whole plastid sequences or hundreds of genes. . Chrysanthemum (Chrysanthemum morifolium) is a popular species used as an ornamental and herbal purpose with its diverse and beautiful flower and with its natural bio-active compounds. In Korea, wild Chrysanthemum species are naturally grown, but their classification and relationships are less identified and their genomic resources are not available yet. Because they are potentially used for genetic improvement of commercial chrysanthemum, we collected and characterized wild Chrysanthemum species. Also, we generated their genomic sequences based on Illumina HiSeq platform. The raw genomic sequences were preprocessed and used to obtain whole DNA sequences of chloroplast in a reference-guided manner. The plastid sequences were compared to identify the phylogenetic relationships among the Korean Chrysanthemum species. Here, we report the detailed procedure, and our chrysanthemum collection and their phylogenetic relationship.

POB-311

Cnidarian phylogenomics to understand the eumetazoan evolution

Mei-Fang Lin[1], James Reimer[2], Miyuki Kanda[3], Nana Arakaki[3], Hiroshi Watanabe[1]

[1]Okinawa Institute of Science and Technology (Japan), [2]University of the Ryukyus (Japan), [3]Okinawa Institute of Science and Technology (Japan)

Cnidarians are the pre-bilaterian animals that diverged close to the base of the metazoan radiation. With their privileged phylogenetic position, Cnidarians bear great potential to reveal fundamental questions in evolutionary biology. Based on the mitogenomic analyses, class Anthozoa has relatively low evolutionary rate which challenged their utility for phylogenetic and systematic purpose. To understand the genomic and biological nature of ancestral Cnidaria, there is a need to reconstruct the phylogenetic relationship among the Cnidaria by using phylogenomics approach for increasing the data coverage. Nowadays, several cnidarian genomic data are available for many of the orders, e.g. scleractinians, actiniarians, and corallimorpharians. However, there is a dearth of knowledge known about the Zoantharia, one of the basal Anthozoa based on mitochondrial DNA analysis. In this study, transcriptome from deep sequencing of Palythoa (Sphenopidae, Zoantharia, Cnidaria) was generated by using Illumina and PacBio platforms and applied for phylogenomic analysis. Our Palythoa transcriptome containing 166,394 contigs and 51,691 protein-coding genes is by far the most complete zoantharia transcriptome. The phylogenomics analysis will be conducted using a concatenated gene alignment and partitioned gene alignments from orthologues of several selected anthozoan genomes and transcriptomes. To increase the tree robustness, highly conservative orthologous with rational phylogenetic signal will be applied. The information of life history and body plan will be incorporated to the phylogenomic results and thus will improve our understanding of evolutionary relationship of cnidarian.

POB-312

Phylogenomic analysis suggests two independent adaptations to high-salt environments in the Archaea

Monique Aouad[1]

[1]Universite Lyon 1 (France)

The analysis of conserved core genes has allowed disentangling most of the ancient relationships in Archaea, an essential frame to fully apprehend their evolution and adaptation to a variety of environments. Some nodes remain nevertheless debated, such as the phylogenetic position of extreme halophilic lineages, the Halobacteria and the Nanohaloarchaea. The former were alternatively proposed as the sister-lineage of methanogens Class II, or Methanocellales, or Methanomicrobiales, while the latter were proposed as the sister-lineage of Halobacteria, or a member of the DPANN superphylum. Clarifying the relative phylogenetic positions of these two lineages will improve our knowledge of the deep evolutionary history of Archaea, and understand if adaptation to high-salt concentrations occurred once or twice independently in this Domain. Moreover, it will help deciphering the molecular adaptive processes that allowed the emergence of these extremophiles.

Using comparative genomics approaches, we have identified over 250 proteins carrying a reliable phylogenetic signal to address this issue. By combining different approaches limiting the impact of non-phylogenetic signal on phylogenetic inference (like for example the Slow-Fast method and the recoding of amino acids) we show that Nanohaloarchaea branch with Methanocellales, while Halobacteriales branch with Methanomicrobiales. This implies that adaption to high salinity emerged twice in Archaea, and that the phenotypic similarities observed in Nanohaloarchaea and Halobacteriales result from convergent evolution, possibly accompanied by horizontal gene transfers. Finally, our results suggest that the grouping of Nanohaloarchaea with other DPANN lineages is the consequence of a tree reconstruction artefact, challenging the existence of this candidate superphylum.

POB-313

Next generation sequencing for analyzing genetic diversity in cultivars of seven countries rice

Namkuk Kim[1], Sujeong Shin[1], Hyemin Park[1], Moonok Lee[1], Myungsil Jung[1], Heykyoung Kim[1]

[1]National Agricultural Product Quality Management Service(NAQS) (Republic of Korea)

In this study, we conducted whole genome re-sequencing of 400 cultivars originated from seven countries (Korea, China, Japan, Australia, America, Thailand and Vietnam) with Oryza sativa (Nipponbare/Japonica) genome as a reference. Almost 95% of the whole genome sequences of the reference genome could be covered by sequenced reads. Numerous single nucleotide polymorphisms (SNPs), insertions and deletions (indels) were detected among the cultivars and the reference genome. Phylogenetic analysis based on inter-varietal polymorphism showed that the 400 cultivars were very closely, but there were differences between Japonica and Indica cultivars. Among these variants (746,632 SNPs), we found 7,919 SNPs in CDS regions of 400 cultivars and 245 SNPs were identified as markers to discrimination cultivars of rice. The 72,695 SNPs were found as unique to Japonica and Indica rice. These variants data provide a foundation for deep exploration of rice diversity and gene-trait relationships and their use for future rice improvement.

POB-314

Gene distributions across taxonomic groups reveal recent lateral gene transfers among prokaryotes

Michael Knopp[1], Jessica Wimmer[1], William F. Martin[1, 2]

[1]Heinrich Heine University Duesseldorf (Germany), [2]Universidade Nova de Lisboa (Portugal)

Lateral gene transfer (LGT) among prokaryotes, which divide clonally, is one of the major sources of prokaryotic genome variability and the key to escape Muller's ratchet. Although the underlying mechanisms of LGT are known (transformation, conjugation, transduction, gene transfer agents and microtubules), quantifying the impact and frequency of prokaryote LGT on genome evolution still presents a challenge. We computed the local and global identities of all protein sequences of 5,655 prokaryotes (>19 million sequences) in a pairwise fashion, resulting in over ~3,6 x 10^{14} pairwise comparisons. From these global identities, over 400,000 protein families were constructed utilizing a Markov-chain clustering approach. Each family was classified via annotations and its protein distribution among genomes mapped. Genes that are specific to a given monophyletic taxonomic group probably arose there. They produce a unique pattern of gene distribution as they get transferred to a member of another monophyletic taxonomic group. Quantification of these patterns within the protein families was used to estimate such recent LGT-frequencies between all prokaryotic taxonomic groups. While LGTs could be detected between the majority of taxonomic groups, it was also possible to identify pairs of taxonomic groups with significantly higher LGT-frequencies. Furthermore, all protein families showing conspicuous protein distributions were analysed by their protein annotation, availability of members on plasmids and by calculation of phylogenetic trees for each family.

POB-315

Canary: a new approach to identify attraction artifacts in the analysis of single gene families

James Frederick Fleming[1], Roberto Feuda[3], Nicholas Roberts[2], Davide Pisani[1, 2]

[1]University of Bristol (United Kingdom), [2]University of Bristol (United Kingdom), [3]California Institute of Technology (United States)

Understanding the evolution of protein families is an important aspect of phylogenetics. However, confusion arises as single gene alignments contain limited phylogenetic information, and traditional approaches to tackle attraction artifacts are difficult to apply to the problem. Here we present a new approach - the Canary Sequence Method - that identifies and removes sequences that could cause attraction artifacts. We evidence the method's efficacy on simulations, a well-known case study and new experimental data in the metazoan opsins as we clarify the placement of cnidarian and ctenophoran sequences within the opsin family tree.

POB-317

Swift Neighbor-Joining method for massive sequence data

Naruya Saitou[1]

[1]National Institute of Genetics (Japan)

We are facing massive genome sequence data, and computationally very rapid methods for constructing gene phylogenies are required to process these genome data. Even our neighbor-joining method is slow for one million sequence data. Significant proportion of forthcoming genome data are human genomes, and gene genealogies of modern humans approximately follow evolutionary rate constancy or molecular clock. Under this condition, I developed new distance method to find neighbors very quickly. We first select global minimum pair of OTUs. This pair is neighbor candidate, and two more OTUs which are third and fourth closest to one of neighbor candidate pair OTUs are selected. These four OTUs are used for four-point metric to examine if neighbor candidate pair does cluster or not. If they cluster together, we determine them as neighbors, and suppress use of one OTU with longer exterior branch. If not, we suppress use of this pair and choose second minimum pair as neighbor candidate. By iterative searches, all neighbors are correctly discovered. This new method is named Swift Neighbor-Joining Method, and its algorithm will be published as part of Introduction to Evolutionary Genomics Second Edition (Saitou 2018, in press from Springer).

POB-318

Searching for binding pockets of FeS clusters in 5,655 prokaryotes

Giacomo Moggioli[1, 2], Nathalie Brenner[1], Joana Xavier[1], Maurizio Casiraghi[2], William F. Martin[1, 3]

[1]Heinrich Heine University, Duesseldorf (Germany), [2]Universita Degli Studi Milano-Bicocca, Milano (Italy), [3]Universidade Nova De Lisboa, Oeiras (Portugal)

Iron-sulfur (FeS) clusters are one of the most important electron carriers, involved in several biochemical reactions across all domains of life. FeS proteins bind and utilize FeS clusters in catalysis, and there is strong biochemical and phylogenetic evidence indicating they were among the very first catalysts and could have played an important role in the emergence of life. Here we used a bioinformatics approach to identify FeS proteins in a dataset of 5655 genomes from 212 Archaea and 5443 Bacteria, containing more than 19 million protein sequences. We manually collected 98 amino acidic motifs of binding pockets for five different types of FeS clusters from the literature and the Protein Data Bank (PDB) and 24 motifs from Prosite, which were then used to scan the collection of prokaryotic genomes. For the literature and PDB collection, 415,667 proteins match one or more binding motifs, and for the Prosite collection, 156,883 proteins were identified as matches. 102,858 proteins were common between both identification pipelines, and those were further explored regarding functional annotation and whether they belong to aerobic or anaerobic taxons. *Methanosarcinales*and other methanogenic organisms from Archaea and *Deltaproteobacteria* from Bacteria are among the groups that are significantly enriched in putative FeS proteins. The results shown here strongly indicate that the utilization of FeS clusters precedes the divergence of Bacteria and Archaea.

POB-319

Island biogeography of Candidia temminckii, a freshwater fish, suggests intraspecific replacements

Shoji Taniguchi[1], Johanna Bertl[2], Andreas Futschik[3], Hirohisa Kishino[1], Toshio Okazaki[1]

[1]The University of Tokyo (Japan), [2]Aarhus University (Denmark), [3]Johannes Kepler University Linz (Austria)

In contrast to other middle-latitudes such as Europe or North America, Japanese archipelago was not covered by ice-sheet during the ice age. The lack of ice-sheet enables a long-term biogeographic analysis in Japan. The primary freshwater fish community in Japan provides invaluable research materials. Because of intolerance to salt, primary freshwater fish disperse only through freshwater system. Phylogenetic analyses of *Candidia temminckii* revealed that the existence of three intraspecific clades in Japanese archipelago presumably diverged in Korean peninsula, and that Japanese population was formed through migration from Korea via land bridge caused by the lower sea level. We detected a notable phenomenon that today's Japanese distribution by clade is arranged in the sequence of migration, and the distributions of older clades are divided by newer clades. To interpret this pattern, we developed biogeographical model of migration, dispersal, and intraspecific replacements. By fitting the values of spatial autocorrelation and Templeton's statistics in nested clade phylogeographic analysis, the approximate Bayesian computation strongly supported the proposed hypothesis that the new migrants had higher fitness and replaced the indigenous populations at the time. To our knowledge, this is the first study to explicitly examine the intraspecific replacements in the framework of island biogeography. The effect of intraspecific replacements has been undervalued so far, but in fact, it might be an important factor in the formation of biota.

POB-320

Gene fusion as a phylogenetic marker - a study of Metazoan

Peter Mulhair[1], Mary O'Connell[1, 2], Raymond Moran[2], Christopher Creeevey[3], Davide Pisani[4, 5], Ian Carr[6]

[1]University of Leeds (United Kingdom), [2]Dublin City University (Ireland), [3]Aberystwyth University (United Kingdom), [4]University of Bristol (United Kingdom), [5]University of Bristol (United Kingdom), [6]University of Leeds (United Kingdom)

Molecular systematics has resolved key evolutionary relationships amongst Animalia. However, the composition of major clades such as the Spiralia and Panarthropoda and the relationships between lineages such as those at the root of the animal tree, between Porifera and Ctenophora, remain controversial. Recent studies have applied an array of different approaches to resolve these issues, including the use of alternative types of sequence data and the application of complex models of evolution. Rare Genomic Events (RGEs) such as insertion-deletion events and miRNAs, have proven powerful in the resolution of highly debated relationships such as the root of the placental mammal tree and the placement of Tardigrada. Here we have investigated the properties of a specific type of RGE - gene fusion. We identified gene fusion events in the metazoa using a sequence similarity network-based approach. We used 63 metazoan genomes and identified a total of 168,787 gene fusion families that are distributed across the species sampled. We have determined that there are relatively low rates of secondary loss in the fusion gene set. We will use well-resolved nodes to test whether the presence/absence pattern of gene fusions in our dataset can recapitulate the known topology in these regions. Finally we will present our results from applying these data to resolving contentious nodes in the animal tree of life. We propose that gene fusion may have useful phylogenetic signal to contribute to resolving contentious nodes in the animal tree of life.

POB-321

Phylogenetic history of ERM proteins in metazoan tree of life

Victoria Shabardina [1], Yukie Kashima[2]

[1] University of Muenster (Germany), [2]University of Tokyo (Japan)

ERM (Ezrin-Radixin-Moesin) proteins serve as a linker between cell membrane and cytoskeleton in the multicellular organisms and consequently participate in division, mobility and adhesion of the cells. This protein family is particularly interesting for evolution studies due to the high conservation of their amino acid sequence (ERM protein homologs from human and fruit fly share 76 per cent sequence similarity) and as a marker of the emergence of multicellularity. Ezrin, radixin and moesin mostly exhibit redundant functions in cells, although there are several examples of ERM tissue specific expression and functioning. This fact has been a puzzle for molecular biologists for a long time; as well it suggests specific evolutionary history of the protein family. In this first massive, grand-scale study of ERM proteins we reconstructed the ERM phylogeny tree using 257 protein sequences from the set of 82 organisms representing all taxonomy groups of metazoan. Several phylogeny methods, including maximum likelihood and Bayesian inference methods, were used to reconstruct the optimal phylogeny relationships. One of the important conclusions is that ezrin appears to be the most recently diverged paralog. Moreover, the results improve our understanding of the ways the functionally conserved proteins evolve by and what is the connection of this process with complexity of the organisms. Phylogenetic analysis of ezrin, radixin and moesin elucidates the mystery of non-redundant redundancy in their functions in the cells.

POB-322

Charting the gene set of the last universal common ancestor

Madeline C. Weiss[1], William F. Martin[1, 2]

[1]Heinrich Heine University Duesseldorf (Germany), [2]Universidade Nova de Lisboa (Portugal)

We have been investigating the nature of the last universal common ancestor, LUCA using genomes. Our question has been which genes LUCA contained as inferred from the gene collection present in 134 archaeal and 1847 bacterial genomes. Our approach was not to look for genes that are universally distributed, but instead to look among all genes to find those that trace to the ancestor of archaea and bacteria by phylogenetic criteria. From 6103411 genes we found 11093 gene families (clusters) that occur in archaea and bacteria. Most of those 11093 generate trees indicating that the transdomain distribution was due to lateral gene transfer (LGT). When we exclude cases of LGT, what remain are 355 genes that trace to LUCA. In contrast to genes that are universal across genomes, these genes contain information about microbial physiology. They allowed us to make a number of inferences about where LUCA lived and what it lived from (Weiss et al. 2016). The results reconstructed LUCA as a strictly anaerobic proto-organism that was half alive and that lived from gasses: CO_2, H_2, N_2, CO, and H_2S. We have now expanded this analysis to encompass 212 archaeal and 5443 bacterial organisms. Based on phylogenetic criteria we have investigated the larger dataset in which fused genes were separated and new protein families were generated.

Weiss MC, Sousa FL, Mrnjavac N, Neukirchen S, Roettger M, Nelson-Sathi S, Martin WF: The physiology and habitat of the last universal common ancestor. Nat Microbiol 1: 16116 (2016).

POB-324

Phylogenomic analyses of testate (shell-building) amoebae in New England bogs and fens

Alistaire D. Ruggiero[1], Angela O'Donnell[1], Agnes Weiner[1], Naomi Ostriker[1], Evie Hoffman[1], Laura A. Katz[1, 2]

[1]Smith College, Northampton, MA 01063 (United States), [2]University of Massachusetts, Amherst, Amherst, MA 01003 (United States)

Though microbes constitute the majority of life on Earth, little is known about the biodiversity and biogeography of many microbial lineages, including the testate (shell-building) amoebae that are the focus of this project. Testate amoebae act as bioindicators of past and current climates, and are abundant in threatened environments, like bogs and fens. While testate amoebae in the clade Arcellinida (Tubulinea, Amoebozoa) are traditionally classified according to the shape and composition of their tests (shells), molecular data show these traits to be deceptive. Here, we use single-cell 'omics of diverse Arcellinida, coupled with the taxon-rich phylogenomic pipeline available in the Katzlab, to evaluate phylogenetic relationships and species boundaries. Previous work uncovered evidence of non-monohyly of genera such as *Hyalosphenia*. To expand on this work, we performed whole transcriptome amplifications on *Hyalosphenia* isolated from *Sphagnum* moss samples from Hawley Bog (Massachusetts). Using our custom phylogenomic pipeline that consists of a ~13,000 gene phylogenomic database as well as ~1,000 taxa that include bacteria, archaea, and eukaryotes, we constructed singe gene trees to infer evolutionary relationships and to look for evidence of sexual reproduction between cells. With these trees, we further unravel the relationship between *Hyalosphenia papilio* and *Hyalosphenia elegans*. More broadly, our findings contribute to the elucidation of the evolution of testate amoebae.

POB-325

The Likelihood Decay Index: Branch support for the phylogenomics era.

Chris Creevey[1], Mark Wilkinson[3], James McInerney[2]

[1]Queens University Belfast (United Kingdom), [2]University of Nottingham (United Kingdom), [3]Natural History Museum, London. (United Kingdom)

Bootstrap proportion (BP) support remains a commonly used metric of the reliability of genome-scale phylogenetic analyses because sampling error decreases as the length of sequences increase resulting in a trend where BP support approaches 100%. However, not all conflicting phylogenetic signal is due to sampling error; processes such as incomplete lineage sorting and horizontal gene transfer can result in valid alternative genetic histories.

Despite this, with long-enough alignments, 100% BP can be achieved even if 49% of the data supports an alternative topology. The heterogeneous nature of the underlying support for branches with 100% BP requires a novel approach and a change in our notion of "support".

To address this, we suggest a likelihood decay support value. Based on the premise of Bremmer support, it is the difference in likelihoods of the optimal trees that do or do not include a given split. Likelihood decay represents a novel way to assess support which discriminates between different internal branches and is insensitive to alignment length. We demonstrate these properties with simulations and investigate the phylogenetic support in "solved" phylogenomic studies where 100% BP support has been obtained. The likelihood decay index has been implemented in "Machete", freely available at https://github.com/ChrisCreevey/machete.

POB-327

Evolutionary metrics

Satoshi Oota[1]

[1]RIKEN (Japan)

Reconstruction of phylogenetic trees is equivalent to map the multi-dimensional evolutionary distance space to the two-dimensional tree graph space. It is well known that the pair-wise evolutionary distance space often violates Euclidian metrics. This is one of the main causes of the misinference in the phylogenetic tree reconstruction. Meanwhile, virtually all existing phylogenetic reconstruction methods follow the minimum evolution (ME). The ThreeTree concept is a new kind of approach to reconstruct phylogenetic trees based on the optimal metricity of the evolutionary distance space rather than the ME principle. This method can neatly find evolutionarily plausible topologies even from phylogenetically ill-conditioned data.

POB-328

Inferring incomplete lineage sorting, duplications, losses and transfers with reconciliations

Yao-ban Chan[1], Celine Scornavacca[2], Vincent Ranwez[3]

[1]The University of Melbourne (Australia), [2]Universite de Montpellier (France), [3]Montpellier SupAgro (France)

Gene trees and species trees can be discordant due to several processes. Reconciliations explain these discrepancies by mapping the gene tree into the species tree and explicitly accounting for these processes. Standard models of reconciliations consider macro-evolutionary events at the genetic level: gene duplications, losses and transfer. Another common source of gene tree-species tree discordance is incomplete lineage sorting (ILS), whereby gene divergences corresponding to speciations occur "out of order". However, ILS is seldom considered in reconciliation models. In this talk, we present a unified formal IDTL reconciliation model which includes all of these processes. We show how to properly cost ILS under this model, and then outline a fixed-parameter tractable (FPT) algorithm that calculates the most parsimonious IDTL reconciliation, with guaranteed time-consistency of transfer events. Provided that the number of branches in contiguous regions of the species tree in which ILS is allowed is bounded by a constant, this algorithm is linear in the number of genes and quadratic in the number of species. This provides a formal foundation to the inference of ILS in a reconciliation framework.

POB-329

Evolutionary dynamics of Papaveraceae plastid genomes and contrasting patterns of organization and nucleotide substitution rates

Seongjun Park[1], SeonJoo Park[1]

[1]Yeungnam University (Republic of Korea)

Papaveraceae (poppy family) comprises approximately 775 species in 42 genera distributed throughout the world, which is classified into two subfamilies: Fumarioideae (DC) Endl. (including *Pteridophyllum* and *Hypecoum*) and Papaveroideae Eaton. However, little is known about plastid genome (plastome) diversity and evolution within the family Papaveraceae. Only two complete plastomes are available, from the subfamily Papaveroideae, and neither of these are members of the subfamily Fumarioideae. To better understand the plastome evolution in the poppy family, we sequenced the plastome sequences of 12 species from two subfamilies. Comparative genomic analyses revealed contrasting patterns of plastome evolution and nucleotide substitution rates within the family Papaveraceae. The subfamily Fumarioideae plastomes exhibits multiple genomic changes, including inversions, operon disruption, gene relocation and duplication, high substitution rates and inverted repeat (IR) shifts. In contrast, the subfamily Papaveroideae plastomes are highly conserved with gene content and order identical to the ancestral organization of angiosperms except for the genus *Eomecon*. Ancestral plastome reconstruction suggests that several rearrangements occurred in the ancestor of the subfamily Fumarioideae followed by independent rearrangements in each genus. A series of IR boundary shifts and inversions played a critical role in shaping the plastome organization and accelerated substitution rates of the subfamily Fumarioideae.

POB-330

Understanding the effect of calibrations in molecular clock dating

Alan James Beavan[1], Mark Beaumont[1], Davide Pisani[1, 2]

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

Applying a timescale to evolution continues to be a controversial endeavour of evolutionary biologists. It is however, essential in order to understand the history of biodiversity on this planet. One of the principle sources of controversy is the extent of the importance of calibrations, which are known constraints on divergence times usually from the fossil record. Here I investigate the importance of calibrating molecular clocks using evolutionary simulations. Phylogenies with highly variable evolutionary rates, which require relaxing of the clock to correctly obtain dates for, will be the main subject of investigation. I will also be presenting the effects of increased rates of evolution during speciation on molecular clock estimates. It is expected that systematic increases in evolutionary rate that cannot be identified from examination of contemporary sequences, such as this, would increase the age of groups estimated using the molecular clock. The use of precise and accurate calibrations ought to be able to account for this. This work will lead to a greater understanding of the accuracy of the molecular clock and reveal situations where it may not be as informative as we thought. This work will also serve to form hypotheses and about genome evolution during speciation, which will be tested empirically in the future.

POB-331

Computing likelihoods of allele frequencues on phylogenetic trees using diffusion models

David James Bryant[1], Stephanus Marnus Stoltz[1]

[1]University of Otago (New Zealand)

In this talk we present an efficient and flexible method for computing likelihoods for phenotypic traits on a phylogeny. We consider a particular family of continuous trait models, where
the trait is the ancestral allele frequencies and the transition densities are determined implicitly via a diffusion. The diffusion processes considered are of Wright-Fisher type in one spatial dimension with the dual operator restricted to zero-flux boundary conditions. The numerical integration of the diffusion processes are approached via fitting Chebyshev basis functions rather than evaluation of the partial likelihoods at specific grid points. The result is higher efficiency, perhaps at the cost of simplicity.

POB-332

Local demographic patterns buried in the present-day mtDNA pool: A study of Finns

Sanni Oversti[1], Paivi Onkamo[1, 2], Monika Stoljarova[3], Bruce Budowle[4, 5], Antti Sajantila[6], Jukka Palo[6, 7]

[1]University of Helsinki (Finland), [2]University of Turku (Finland), [3]Tallinn University of Technology (Estonia), [4]University of North Texas Health Science Center (United States), [5]King Abdulaziz University (Saudi Arabia), [6]University of Helsinki (Finland), [7]National Institute for Health and Welfare (Finland)

By utilizing mitochondrial DNA (mtDNA) we are able to trace maternal lineages back in time. A wealth of population genetics studies has shown that the mitochondrial gene pool in modern Europeans is a mixture of Mesolithic hunter-gatherer associated haplogroups (U and V) and Neolithic associated farmer haplogroups (H, J, K and T). Also, the demographic histories deduced from mtDNA data seem relatively homologous for many European populations. However, a lot of this population level research has utilized only the mtDNA control region, containing only a fraction of the information of the mitochondrial genome.

Here we have reassessed the Finnish population history using complete mtDNA genome sequences. When solely the control region is considered, Finnish mtDNA diversity and haplogroup distribution seem similar to that of other European populations. A different picture emerges from the 843 full mitochondrial genome sequences from modern Finns analyzed here. Based on this fine resolution comparison, a significant proportion of subhaplogroups in Finland, up to one third, can be considered as Finn-characteristic, i.e. these lineages are rather common in Finland but they are virtually absent from other populations. Bayesian phylogenetic analyses suggest that most of these lineages date back around 3,000-5,000 years, which would be temporally coincidence with the arrival and especially the spreading of the agriculture and Corded Ware culture in Finland. In addition, when comparing the effective population sizes for Finn-characteristic and for other haplogroups, signatures of two remarkably different demographic histories can be detected.

POB-333

A novel method for detection of syntenic regions between genomes reveals the extent of DNA transfer between plasmids and chromosomes in prokaryotes

Ahmad Samer Kadib Alban[1], Tal Dagan[1], Giddy Landan[1]

[1]Chrisian Albrechts University of Kiel (Germany)

Plasmids are extra chromosomal genetic elements that replicate autonomously. Plasmids are abundantly found in prokaryotes and recognized as major drivers of lateral gene transfer in prokaryote evolution. The research of DNA transfer between plasmids and chromosomes so far focussed on the transfer of genes encoding for beneficial traits, e.g., resistance mechanisms to antibiotics or heavy metals. Nonetheless, the overall extent and polarity of DNA transfer between plasmids and chromosomes is poorly understood. Here we present a novel comparative genomics approach to detect homologous regions between plasmids and chromosomes. Our method combines local sequence similarity as detected by BLAST and short identical sequences as identified by MUMMER into a unified framework of syntenic regions between two genomes. This approach proves efficient in uncovering different levels of sequence similarity that would be ignored otherwise. Thus, it enables detecting syntenic regions in the presence of genomic rearrangements, duplications and insertions/deletions. We applied our approach to 1400 prokaryote species including 3270 plasmid-chromosome pairs that co-inhabit the same cell. This uncovered 65191 syntenic regions in a total length of 105Gbp with a median of 6 syntenic regions per plasmid-chromosome pair. The median syntenic regions length is 1.5Kbp. The detected DNA segments resulting from lateral DNA transfer account for overall 7.9% of the chromosomal DNA with a median of 2.8%, and overall 6.73% plasmid DNA with a median of 1.88%. Our approach uncovers frequent DNA transfer between plasmids and chromosomes that is highly variable among different prokaryotic lineages.

POB-334

Evolutionary history of short specific sequences in the three domains of life

Nicole Gruenheit[1], Michael Knopp[1], Nils Kapust[1], Peter Lockhart[2], William Martin[1]

[1]HHU Duesseldorf (Germany), [2]Massey University (New Zealand)

Ever since the sequencing of the first genomes, the importance of k-mers -- DNA sequences of length k -- has become more and more apparent. One startling observation, which so far has mainly been used in metagenomics tools, was that there are k-mers which are specific for each bacterial species. Although databases have been created for a large number of bacteria, several questions still remain unanswered one of which is: What are those short DNA sequences, that only occur in one bacterial species, although bacteria are known to frequently exchange genes through lateral gene transfer? Also, all previous analyses focussed on bacteria but do not consider the other domains of life. Therefore, we created a k-mer database containing the k-mer profiles of all currently sequenced bacteria, archaea, and eukaryotes. Using these profiles, we firstly identified all k-mers specific to a certain species or any higher taxonomic order and assigned them to genes, promoter regions, or intergenic regions without any function. Finally, we computed the similarities and distances between the k-mers and used phylogenetic trees to analyse and visualise groups within the k-mer space, which revealed unexpected patterns in the k-mer distributions and their relationships especially when pangenomes and pathogens were considered.

POB-335

CO-dependent CO2 fixation: unique, ancient and ancestral in biochemistry

Joana C. Xavier[1], Martina Preiner[1], William F. Martin[1, 2]

[1]Heinrich Heine University Duesseldorf (Germany), [2]Universidade Nova de Lisboa (Portugal)

Carbon can enter metabolism via six known pathways of CO_2 fixation, however only one is linear and exergonic - the Wood-Ljungdahl (WL) pathway. The WL pathway is present in a diverse group of anaerobic prokaryotes and there is strong evidence pointing to its ancestry, all the way to the last universal common ancestor. Carbon monoxide (CO) is a central compound to the WL pathway in the conversion of inorganic to organic carbon but also as a possible energy source. Here we scan the major databases of biochemical reactions for the involvement of CO and CO_2 in prokaryotic metabolism and show that CO is quite unique to the WL pathway with very few other cases where it occurs as a by-product, whereas CO_2 is ubiquitously present across metabolic maps. We show the significant enrichment of CO2 in specific pathways and analyse the reactions utilizing CO_2 and CO for their taxonomic distribution, reversibility and dependence on different organic and inorganic catalysts. An in-depth phylogenomic analysis of the special enzyme responsible for the (reversible) interconversion between CO_2 and CO, carbon monoxide dehydrogenase (CODH), was also performed. A manual collection of 18 CODH sequences from different archaea and bacteria with alternative metabolic capacities was blasted against 5,655 prokaryotic genomes, showing a specific and unique distribution with some interesting displacements. This distribution is complemented with that of acetyl-CoA synthase, which converts CO, a methyl group and CoA to acetyl-CoA, the first organic carbon compound entering metabolic networks fed by the WL pathway.

POB-337

Population-level processes impact the inference of macroevolutionary regimes from molecular phylogenies when concatenation is employed

Carlos Schrago[1], Anieli Pereira[1]

[1]Federal University of Rio de Janeiro (Brazil)

Interest in methods that estimate speciation and extinction rates from molecular phylogenies has increased over the last decade. The application of such methods requires reliable estimates of tree topology and node ages, which are frequently obtained using standard phylogenetic inference combining concatenated loci and molecular dating. However, this practice disregards population-level processes that generate gene tree/species tree discordance. We evaluated the impact of employing concatenation and coalescent-based phylogeny inference in recovering the correct macroevolutionary regime using simulated data based on the well-established diversification rate shift of delphinids in Cetacea. We found that under scenarios of strong incomplete lineage sorting, macroevolutionary analysis of phylogenies inferred by concatenating loci failed to recover the delphinid diversification shift, while the coalescent-based tree consistently retrieved the correct rate regime. We suggest that ignoring microevolutionary processes reduces the power of methods that estimate macroevolutionary regimes from molecular data.

POB-338

An appraisal of the relationships of Sigmodontinae (Cricetidae): a phylogenomic approach resolve tribal relationships.

Guillermo D'Elia[1], Andres Parada[1]

[1]Universidad Austral de Chile (Chile)

Cricetid rodents of the subfamily Sigmodontinae constitute a large radiations of New World mammals. Currently more than 400 living species are recognized. These species are allocated in ca. 90 living genera, which in turn are grouped into 10 tribes. With few exceptions, recent phylogenetic analyses have stabilized the number and contents of the distinct tribes. Conversely, relationships among tribes remain mostly unresolved or poorly supported. Here, we present results of analyses based on over 10 thousand orthologous genes, gathered via transcriptome sequencing, which were analyzed with distinct methods (e.g., ML, coalescent based species trees, molecular clock) and datasets (e.g., matrices with 75% and 100% gene occupancy). Results portray a resolved and dated scheme of sigmodontine tribal relationships. For instance, within a robust Oryzomyalia, Oryzomyini is sister to a clade formed by the remaining tribes. In addition, we also assessed relationships among the five subfamilies of the family Cricetidae, where New World endemic subfamilies (i.e., Tylomyinae, Neotominae, Sigmodontinae) form a robust clade. Financial support: FONDECYT 1141055 and 3150604.

POB-339

Expansion and evolution of Terpene synthase gene families in stout camphor tree

Han-Yu Wang[1], Chih-Yao Hsu[1], Chung-Shien Wu[1], Ling-Ni Wang[1], Isheng Tsai[1], Shu-Miaw Chaw [1]

[1] Academia Sinica (Taiwan)

Cinnamomum kanehirae (Lauracae), the stout camphor tree (SCT), is a large evergreen tree endemic to Taiwan. SCT is economically important for its bacterial resistant wood that were once commonly used in making furniture and sculptures. SCT contains camphor oils that are natural odoriferous terpenoids. Terpenoids are the most abundant and structurally diverse group of natural products in plants. They play a vital role in interactions with other plants and organisms, and their ecological environments, therefore, they are indispensable for plant growth and development. Terpene synthases (TPSs) are the key enzymes involved in the biosynthesis of terpenoids. They were divided into five different subfamilies. In this study, we analyzed the draft genome of SCT with a particular focus on its TPS gene family. We used HMMER to explore TPS homologs, and identified 100 putative TPS genes. This is so far the largest TPS group in known plant genomes. These putative TPS genes vary in size from 200 to 1,500 nucleotides. We also sampled TPS sequences of other seed plant representatives to construct a TPS gene tree. As expected, the TPS genes of SCT constitute five major known subfamilies. However, the TPS-a and TPS-b subfamilies (involved in secondary metabolism), as well as TPS-c and TPS-e (involved in primary metabolism) are remarkably expanded in SCT. This finding suggests that duplications of TPS genes might have repeatedly occurred in SCT after its divergence from other eudicots.

POB-341

Analysis of Plasmid Gene Network

Ignacio Riquelme Medina[1], James O McInerney[2]

[1]University of Manchester (United Kingdom), [2]University of Nottingham (United Kingdom)

Plasmids are small DNA molecules that are present in organisms like bacteria or fungi, that can be transmitted between individuals of the same generation (horizontal gene transfer) and generally give advantageous traits to the individual like antibiotic resistance or toxin production. Because plasmids are not part of the genomic DNA and play a big role in the horizontal gene transfer their evolution is expected to be different to the genomic DNA. In this project we intend to examine the evolution of the plasmids using unipartite and bipartite networks, checking how they differ from genomic networks and what characterizes them as well as trying to identify the factors that play a role in shaping these networks by using models. Finally we will also look out for communities of genes in the plasmids that always occur together and communities that avoid each other.

POB-342

Hybrids of Paradise: A genomic perspective on intergeneric gene flow among Birds-of-Paradise

Mozes Pil Kyu Blom[1], Stefan Prost[2], Les Christidis[3], Brett Benz[4], Valentina Peona[5], Alexander Suh[5], Martin Irestedt[1]

[1]Swedish Museum of Natural History (Sweden), [2]Stanford University (United States), [3]University of Melbourne (Australia), [4]American Museum of Natural History (United States), [5]Uppsala University (Sweden)

It is now well known that species boundaries often remain semipermeable for prolonged periods of time. Interspecific gene flow mostly occurs between sister-species but instances of hybridization are occasionally also recorded among distantly related clades. However, the evolutionary implications of incidental gene flow at a macroevolutionary scale remain largely unclear. Here, we address this issue by focusing on the genomic signature of interspecific hybridization in Birds-of-Paradise (*Paradisaeidae*). Birds-of-Paradise are among the most prominent examples of sexual-selection but, paradoxically, intergeneric hybrids with intermediate phenotypes are frequently observed. To characterize gene flow among taxa, we assembled a high-quality genome using a combination of long-read sequencing and scaffolding approaches (i.e. PacBio, Hi-C and 10X) and re-sequenced the genomes for 39 other species of *Paradisaeidae*. We inferred a summary-coalescent species tree using 10,000 independent sequence windows and quantified the extent of topological congruence across the genome. We inferred a species tree that was overall well supported, but also identified species that were difficult to place due to extensive discordance in coalescent patterns between loci. To investigate whether this is caused by past hybridization or incomplete lineage sorting, we evaluated the relative frequency of competing topologies, quantified the distribution of coalescent times and used contemporary multi-species network approaches to model reticulation and deep coalescence simultaneously. We uncovered multiple instances of introgressive hybridization between species that have been phylogenetically distinct for millions of years and discuss the role of interspecific gene flow in promoting evolutionary diversification within this extraordinary radiation of Indo-Australian birds.

POB-344

Inferring Prokaryotic Evolution using a Deterministic Model of Speciation

Ashley Ann Superson[1], Michael Ryan Culver[2], Anna Maria Spagnuolo[2], Fabia Ursula Battistuzzi[1]

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

Determining accurate speciation rates is challenging because of their dependency on accurate phylogenies and timing. In Prokaryotes, this is particularly problematic given the absence of additional evidence (e.g., fossil record) to help narrow down uncertainties in their evolutionary patterns. A few studies that have provided a statistical framework for analyzing diversification in microbes have shown constant speciation rates for some datasets and non-constant for others. The models applied in these studies are computationally intense, thus limiting the number of hypotheses that can be tested to identify causes of discrepancies in speciation patterns. To address this issue, we developed a simple deterministic model of diversification that does not take into account extinction rates that are often not known, especially for prokaryotes. Our model uses ordinary differential equations where time elapses between nodes become the regions over which the model operates, from which rates of speciation can be identified from an empirical dataset. We test multiple hypotheses that could affect evolutionary patterns including planetary-scale environmental changes and phylogenetic profiles produced by different taxon sampling. Preliminary results indicate constant speciation rates for Prokaryotes, no appreciable effect of environmental changes on speciation rates, and show differences in speciation curves given different taxon sampling scenarios. Future work will be focused on assessing the accuracy of our model using simulated data produced under different evolutionary patterns.

POB-345

Cis-trans evolution of chloroplast development regulator genes in plants

Yao-Ming Chang[1], Hsin-Hung Lin[1], Wen-Hsiung Li[1, 2]

[1]Academia Sinica (Taiwan), [2]University of Chicago (United States)

Chloroplast, the factory of photosynthesis, is essential for plants. Golden2-like (GLK) genes are the well-known regulators for chloroplast development. In previous studies, GLK duplicated genes had been found in most plant species from mosses to flowering plants. In Arabidopsis and rice, two GLK duplicates were found functionally redundant but only one was highly expressed in leaves. In tomato, two duplicates were also functionally redundant but both were highly expressed in different tissues. In maize, two duplicates were found to regulate the different chloroplast morphologies in mesophyll and bundle sheath cells. How these different expression patterns of GLK genes in different species have evolved is not clear. Previous studies proposed that GLK duplicates in dicots and monocots were independently evolved, using protein sequence analysis. In this study, we found that the protein sequences of GLK genes in dicots and monocots are too conserved to reconstruct their true phylogeny. Therefore, we developed a new approach to analyze their promoter sequences and reconstructed a more realistic phylogeny. Our result showed that the two GLK duplicates in Arabidopsis and tomato were orthologous to one duplicate (GLK2) in monocots, overturning the current view of independent evolution. In addition, we proposed a hypothesis that GLK1 genes in monocots changed their cis-regulation for highly expressed in mesophyll cells and GLK2 gene in maize changed its trans-regulation for C4-specific chloroplast development in bundle sheath cells. This study provides a new way to reconstruct the phylogeny of functionally conserved genes and sheds light on the GLK evolution.

POB-346

Investigating Signals of Local Adaptation on the Human X Chromosome

Jonathan Rice[1], Emilia Huerta-Sanchez[1]

[1]University of California Merced (United States)

Previous studies of local adaptation have used genotyping arrays to investigate genomic variation of X-linked SNPs. With the advent of whole genome sequencing (WGS), we are now able to better identify regions of the genome that differ between populations. In this study we take advantage of WGS data from the 1000 Genomes Project to find regions of local adaptation on the X chromosome and the autosomes. Using a matrix of all pairwise F_{ST} values computed genome-wide, we first use a phylogenetic approach to investigate the patterns of genetic variation in the X chromosome and the autosomes. By building trees from this matrix, we found that most populations group with others from the same continent. However, admixed populations tend to be found between populations that contribute to their mixed ancestry. We also found that F_{ST} values are larger on the X chromosome compared to the autosomes and attribute this to the smaller effective population size of the X chromosome. To identify signals of local adaptation, we examined all possible population pairs and computed pairwise F_{ST} and Maximum Allele Frequency Difference (MAXAFD) across all SNPs on the X chromosome. Then by isolating SNPs with lower F_{ST} and higher MAXAFD values than expected, we were able to identify 3 X-linked genes with large differences between closely related populations - especially among African population pairs. An identical study of the autosomes found 22 genes with signals of local adaptation mainly across South Asian populations.

POB-347

Integration of phylogenomics and 3D protein modeling reveal phylogenetic inertia in ryanodine receptor ligan peptides of scorpion venom

Carlos Eduardo Santibanez-Lopez[1], Ricardo Kriebel[2], Jesus Ballesteros[1], Prashant P Sharma[1]

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

Scorpions, an iconic lineage of arthropods, have evolved venoms diverse toxins with a plethora of biological targets, but characterizing the evolution of this molecular diversity has been limited by the lack of a comprehensive phylogenetic hypothesis of scorpion relationships. Here, we generated the most comprehensive scorpion phylogenomic tree to date, providing an evolutionary framework to pinpoint the origins of calcins, a type of inhibitor cystine knot (ICK) peptides that bind to ryanodine receptors in skeletal and cardiac muscle in mammals. We surveyed scorpion genomes and transcriptomes to compile a comprehensive sampling of the ICK peptide homologs across the breath of scorpion diversity, utilized molecular 3D modeling coupled with geometric morphometrics to characterize evolutionary changes in the shape of these peptides, and applied phylogenetic comparative methods to describe the selective pressure and evolutionary trends in these molecules. We show that calcins are phylogenetically restricted to the parvorder Iurida, one of the two basal branches of scorpions, but not to its sister group, parvorder Buthida, which bear the related lambda potassium channel toxins (LKTx). Our results demonstrate phylogenetic inertia of both calcins and LKTx genes and support their characterization as the first synapomorphies (shared derived traits) of the parvorders.

POB-350

Phylogenomics resolves New World primates phylogeny

Horacio Schneider[1], Jeferson Carneiro[1], Iracilda Sampaio[1]

[1]Universidade Federal do Para (Brazil)

Presently there is a good consensus about the phylogenetic tree of the New World primates. All the nodes are very well resolved except the one that places *Aotus* in relation to *Cebus*, *Saimiri* and the callitrichines. Aiming to resolve the relationships among them we analyzed sequence data generated by the ENCODE initiative. 40 genomic regions from Prosimians (*Otolemur* and *Microcebus*), New World primates (*Aotus*, *Callithrix*, *Callicebus* and *Saimiri*), and Old World primates (*Papio*, *Nomascus*, *Homo* and *Pan*) were downloaded from the GenBank. The sequences were inspected using Repeat Masker, aligned with MAFFT and cleaned with GBLOCK. Maximum parsimony and likelihood analyses were conducted in PAUP* and IQ-TREE, respectively. Bayesian inference and divergence times were estimated in MrBayes and BEAST, respectively. The phylogenetic analyses using 40 concatenated encode regions, covering 5,462,230 bp, grouped significantly *Aotus* and *Callithrix* as a sister clade. This arrangement was supported by 9 encode regions, while 2 regions produced a discordant arrangement with *Aotus* and *Saimiri* as sister groups. The remaining 29 regions showed an unresolved trichotomy *Aotus*-*Saimiri*-*Callithrix*. Divergence time estimates showed an age of 20.46 for the emergence of *Saimiri* and of 19.78 Ma for the split between *Aotus* and *Callithrix*. Therefore, as the diversification between *Saimiri*/*Aotus*/*Callithrix* occurred in a short period of time, the two discordant gene trees may be explained by incomplete lineage sorting and hybridization.

POB-351

PhyCAP: Phylogenomic-noise Cleaning Approach by PCA

Eisuke Iwamoto[1], Koichiro Tamura[1, 2]

[1]Tokyo Metropolitan University (Japan), [2]Tokyo Metropolitan University (Japan)

Recent phylogenomic analyses using genome-scale data owing to high-throughput sequencing technology showed that stochastic error is no longer a problem in phylogenetic tree reconstruction. However, it has been exposed at the same time that as another problem, systematic biases such as rate and/or base composition heterogeneity among sites and/or lineages distort accurate tree reconstruction. Although several approaches have been proposed to resolve this problem, most of them are applicable to limited cases. Furthermore, they often take a forbidden computation time, when applied to a genome-scale large dataset. Therefore, we have developed a new method, PhyCAP, which estimates the extent of systematic bias in each gene sequence within a multi-gene dataset using principal component analysis (PCA) of branch lengths of initial tree reconstructed by fast tree making method like NJ method. Then, according to the estimated extent of systematic bias, whether a gene is included in or excluded from tree reconstruction is determined. We evaluated the efficiency of PhyCAP comparing to four other methods (nhPhyml, PartitionFinder, TIGER and LS^3) using three empirical datasets (corbiculate bees, birds and bony vertebrates), in addition to computer simulated datasets. It turned out that PhyCAP was more effective than other methods in most cases, even if it required much shorter computation time than other methods.

POB-352

Testing the accuracy of mitochondrial genomes reconstruction from transcriptomes: an insight from Reticulitermes termites mitogenomics.

Giobbe Forni[1], Guglielmo Puccio[1], Barbara Mantovani[1], Andrea Luchetti[1]

[1]University of Bologna (Italy)

Mitochondrial genomes are of primary importance in the field of molecular phylogenetics. Since the last thirty years, there has been a rapid accumulation of sequenced insects mitogenomes: over 600 molecules are currently available for different insects orders. In addition to this large body of data, it is possible to further broaden the taxonomic sampling by mining published transcriptomes, usually generated for nuclear gene expression studies. We developed a pipeline for mitogenomes assembly in order to test if and how the transcriptomes quality may affect the process and whether evolutionarily distant reference genomes could be used for mitochondrial reads recovery. The pipeline includes a combination of reference-based iterative and baiting approaches, which was tested on *Reticulitermes* termites, due to their well-known phylogeny and the number of available transcriptomes and mitogenomes. Independently from the dataset size and sequencing layout or the presence of contaminants, we assembled complete molecules from six published transcriptomes and succesfully annotated all of the 13 PCG, two mitochondrial rRNA, 22 tRNAs and partial control region. These molecules resulted consistent to conspecific mitogenomes obtained with genome sequencing approaches (Sanger, NGS) and phylogenetic analyses recovered the expected branching pattern. Moreover, we also found out that the use of reference genomes from distantly-related species efficiently guides the assembly: this will likely help in those wide insects clades where only few mitogenomes have been sequenced. Further data mining is successfully ongoing for other insects taxa, with the help of a public available python script which completely automates the process.

POB-353

Phylogenomic analyses of transcriptome data from individual foramifera

Evie Grey Elizabeth Hoffman[1], Alistaire Ruggiero[1], Naomi Ostriker[1], Agnes Weiner[1], Laura Katz[1]

[1]Smith College (United States)

Though microbes constitute the majority of life on Earth, little is known about the biodiversity of many microbial lineages, including the Foraminifera that are the focus of this project. Foraminifera act as bioindicators of past and current climates, and are found in both marine and (more rarely) freshwater. Foraminifera use many strategies to build their tests, such as constructing a test from organic materials, or agglutinating (sticking) materials from their environment to form their test. Here, we use single-cell transcriptome data, coupled with the taxon-rich phylogenomic pipeline available in the Katzlab, to investigate metabolic pathways involved in test building in foraminifera. We have already completed high throughput sequencing to characterize expressed genes from diverse forams. Using our custom phylogenomic pipeline that consists of a ~13,000 gene trees as well as ~1,000 taxa that include bacteria, archaea, and eukaryotes, we are constructing single gene trees to infer evolutionary relationships and to look at patterns of duplication/loss for key pathways. With these trees, we are investigating candidate genes involved in pathways that may contribute to the generation of adhesions and the deposition of test (shell) material. More broadly, our findings contribute to the elucidation of the evolution of testate protist lineages.

POB-354

Disentangling dietary transitions in termite evolution by transcriptome- and mitochondrial genome-based phylogenies

Ales Bucek[1, 2], Jan Sobotnik[2], David Sillam-Dusses[3, 4], Nathan Lo[5], Thomas Bourguignon[1, 2]

[1]Okinawa Institute of Science and Technology Graduate University (Japan), [2]Czech University of Life Sciences (Czech Republic), [3]Sorbonne Universites (France), [4]Universite Paris 13 (France), [5]University of Sydney (Australia)

Termites are the oldest social insect lineage. They are one of the few animal groups able to feed on lignocellulose, the most abundant biomolecule on Earth. These characteristics largely contributed to their success, and allowed termites to become one of the most abundant animal group in tropical ecosystems. To understand the key evolutionary transitions that shaped modern termites, including the evolution of sociality and the loss of symbiotic gut protozoa, a robust molecular phylogenetic tree is required.

We sequenced 43 termite transcriptomes, and obtained genomic and transcriptomic data for another 17 termite species and seven Dictyoptera outgroups. We carried out maximum likelihood and summary coalescent-based phylogeny inference using 4065 predicted orthologous protein-coding genes. Our analyses confirm the branching pattern of most termite lineages, including the polyphyly of Rhinotermitidae and Termitinae, and the paraphyly of *Heterotermes* to *Coptotermes*. Surprisingly, our analyses also indicate sister relationships between the bacterial comb-builders Sphaerotermitinae and the fungus-grower Macrotermitinae. This tree topology implies that the cellulolytic protozoa, unique to lower termites and lost in Termitidae, were not replaced by fungal gardens, as previously suggested, but by bacteria. Finally, we compared the transcriptome-based phylogenetic trees (T-trees) and mitochondrial genome-based phylogenies (mtG-trees), and found that T-trees often resolved branches which were poorly supported in mtG-trees. T-trees also provided alternative topology for few highly supported branches in mtG-trees and the coalescent-based T-trees indicated that the species tree inference was influenced by hybridization and/or incomplete lineage sorting which was presumably unrecognized in previous termite molecular phylogenies.

POB-356

Using Alignment Uncertainty to improve Phylogenetic Bootstrap Reliability

Evan W. Floden[1, 2], Kuei Yuan Lan[3], Javier Herrero[4], Olivier Gascuel[5], Cedric Notredame[1, 2], Jia-Ming Chang[3]

[1]The Barcelona Institute of Science and Technology (Spain), [2]Universitat Pompeu Fabra (UPF) (Spain), [3]National Chengchi University (Taiwan), [4]UCL Cancer Institute (United Kingdom), [5]Institut Pasteur (France)

Most evolutionary analyses are based on pre-estimated multiple sequence alignment. From a computational point of view, it is as complex to estimate a correct alignment, as it is to derive a correct tree from that alignment. Wong et al. established the existence of an uncertainty induced by multiple sequence alignment when reconstructing phylogenies. They were able to show that in many cases different aligners produce different phylogenies, with no simple objective criterion sufficient to distinguish among these alternatives. We demonstrate that incorporating MSA induced uncertainty into bootstrap sampling can significantly increase correlation between clade correctness and its corresponding boot-strap value. We named this concatenation and bootstrapping method, Weighted Partial Super Multiple Sequence Alignment (wpSMSA). We show on three simulated datasets of 16, 32 and 64 tips that our method improves the predictive power of bootstrap values. We also used as a benchmark an empirical collection of 853 1-to-1 orthologous genes from seven yeast species and found wpSMSA to significantly improve discrimination capacity between topologically correct and incorrect trees. For reduced trees by 50% and 95% bootstrap thresholds, wpSMSA comes out the lowest Type I error (less false positive). Bootstrap values of wpSMSA are comparable to similar readouts estimated using a single method. The automated generation of replicates has been implemented in the T-Coffee package, which is available as open source freeware available from www.tcoffee.org.

POB-358

Adaptive evolution after losing an essential gene in yeast

Shang-Lin Chang[1], Hsuan-Kai Wang[2], Luh Tung[1], Tien-Hsien Chang[1]

[1]Academia Sinica (Taiwan), [2]College of Life Science (Taiwan)

Essential genes are conventionally defined as those that are indispensable to cell viability. Recent studies, however, suggest that gene essentiality is dependent on both environmental and genetic contexts and is therefore a quantitative, rather than a binary, trait. For example, background-specific genetic modifiers can govern gene essentiality and ~9% of essential gene deletions in the budding yeast can be suppressed by adaptive mutations. In addition, orthologous genes can acquire opposing statuses of essentiality in different species in the course of evolution. Thus, it is hypothetically possible that an essential gene loss may not always be fatefully detrimental; instead it may pave the way toward genome evolution. Should such a hypothesis hold true, the often-incurred fitness tradeoff in the event of an essential-gene loss rescued by secondary mutations would have to be ameliorated through system re-optimization. We explored this issue in the context of pre-mRNA splicing by experimentally evolving yeast cells harboring a permanent loss of the splicing factor Prp28, in the presence of a genetic modifier. Here we show that cellular fitness can be restored by adaptive mutations that modify the transcriptional machinery. We found that altering the SAGA (Spt-Ada-Gcn5 acetyltransferase) co-activator complex improves splicing in the Prp28-less cells. Detailed analysis revealed that transcriptional attenuation accounts for the splicing improvement, which in turn boosts cellular fitness. Our results thus argue that transcription and splicing are exquisitely coordinated in yeast and that the intrinsic interconnectivity within a biological system can be exploited for adaptive evolution.

POB-359

An experimental phylogeny to benchmark ancestral sequence reconstruction

Ryan N. Randall[1], Caelan E. Radford[1], Kelsey A. Roof[1], Divya K. Natarajan[1], Eric A. Gaucher[1, 2]

[1]Georgia Institute Of Technology (United States), [2]Georgia Institute Of Technology (United States)

Ancestral sequence reconstruction (ASR) is a still-burgeoning method that has revealed many key mechanisms of molecular evolution. One criticism of the approach is the inability to validate its algorithms within a biological context as opposed to a computer simulation. Here we build an experimental phylogeny using the gene of a single red fluorescent protein to address this criticism. The evolved phylogeny consists of 19 operational taxonomic units (leaves) and 17 ancestral bifurcations (nodes) that display a wide variety of fluorescent phenotypes. The 19 leaves then serve as 'modern' sequences that we subject to ASR analyses using various algorithms and to benchmark against the known ancestral genotypes and ancestral phenotypes. We confirm computer simulations that show all algorithms infer ancient sequences with high accuracy, yet we also reveal wide variation in the phenotypes encoded by incorrectly inferred sequences. Specifically, Bayesian methods incorporating rate variation significantly outperform the maximum parsimony criterion in phenotypic accuracy. Subsampling of extant sequences had minor effect on the inference of ancestral sequences.

POB-360

Mapping resistance to eight different viruses in the genetic model organism Drosophila melanogaster

Rodrigo Cogni[1], Francis Jiggins[2]

[1]Universidade de Sao Paulo (Brazil), [2]Cambridge University (United Kingdom)

Natural enemies are a main selection pressure driving the evolution and coevolution of organisms. We are using recent advances in Drosophila genetics to address fundamental questions about the evolution and genetics of pathogen defence. Variation in susceptibility to infectious disease often has a substantial genetic component in natural populations. We are using genome-wide association studies (GWAS) in Drosophila melanogaster to identify the genetic basis of variation in susceptibility to viral infection. We assayed resistance to eight different viruses on a natural population and found extensive genetic variation in resistance to all viruses. In general, we found more genetic variation for viruses that naturally infect D. melanogaster, and less genetic variation for viruses that were isolated from other insects or viruses that are avirulent. We found that resistance to different viruses are not correlated. For some of the viruses we found a few major effect loci that explains most of the genetic variation. These major effect loci affecting resistance are virus specific. These major effect loci also show contrasting patterns of evidence for recent natural selection. Finally, by correlating virus resistance with lifespan and fecundity data, we showed that viruses resistance does not present major fitness costs.

POB-361

Mistranslation influences the bacterial stress response

Laasya Samhita[1], Parth K Raval[1], Deepa Agashe[1]

[1]National Centre for Biological Sciences (India)

Mistranslation is typically deleterious, and cells are thought to have evolved various mechanisms to minimise the error rate. However, translation errors have also been shown to give the bacterial cell an advantage under specific environments, usually mediated by an altered proteome. Most such demonstrations have involved engineered systems using mutant tRNAs or tRNA synthetase enzymes. Here we investigate the consequence of lowering initiator tRNA (tRNAi) levels in *E.coli* , a condition that occurs naturally upon starvation and one that has already been shown to lead to mistranslation. We find that a mutant with lower tRNAi shows heightened resistance to DNA damage in particular exhibiting greatly enhanced resistance and persistence to ciprofloxacin. This condition could be partially mimicked by inhibiting protein synthesis using the antibiotic kasugamycin, suggesting a link between protein synthesis inhibition and response to DNA damage. Further investigation revealed an elegant strategy where lowering tRNAi levels reduces the threshold at which SOS induction occurs in response to DNA damage. While it is well established that resistance to ciprofloxacin is mediated by the SOS response, we find that SOS induction itself is influenced by the level of mistranslation in the cell. Overall, a unique link between protein synthesis and DNA damage appears to be exploited by the cell in responding to certain stressful conditions.

POB-363

A population genetics perspective on measures of intolerance to mutation

Zachary L Fuller[1], Jeremy J Berg[1], Hakhamanesh Mostafavi[1], Guy Sella[1], Molly Przeworski[1]

[1]Columbia University (United States)

To better predict disease mutations in patients, human geneticists have introduced a number of measures that aim to identify genes (or domains) in which mutations are likely to be highly deleterious. The idea is that, since deleterious mutations are kept at low frequency by natural selection, genes with fewer loss-of-function variants than expected under neutrality are likely to be under purifying selection. In one widely-used version of this approach, a depletion in loss-of-function variants is further used to classify genes as haploinsufficient (i.e., with a dominance coefficient $h > 0$) or recessive ($h=0$) in their effect on fitness. Applied to known Mendelian disease genes, this measure appears to reliably distinguish severe haploinsufficient from recessive cases. While measures of intolerance to mutation are not articulated in terms of an explicit population genetics model, doing so makes it clear these approaches cannot actually distinguish haplosufficient and haploinsufficient effects on fitness but are instead informative about hs jointly (where s is the selection coefficient). Thus, the apparent success of such measures applied to Mendelian disease genes appears to reflect the large selection coefficient (s) of Mendelian disease genes, rather than any insight into dominance (h) and may not carry over to other genes. We develop an analogous method that is explicitly based on a population genetic model, as well as a model of local mutation rates, and provides more direct interpretability. Using our approach, we delimit the parameter space under which such methods are expected to be informative about dominance effects.

POB-364

Fitness effects of new mutations in Chlamydomonas

Peter Keightley[1], Katharina Boendel[1], Toby Samuels[1], Susanne Kraemer[1], Rob Ness[2], Colegrave Nick[1]

[1]University of Edinburgh (United Kingdom), [2]University of Toronto (Canada)

We conducted mutation accumulation (MA) experiments for c.1,000 generations, starting from several different strains of the single-celled alga Chlamydomonas reinhardtii, and measured growth rate and competitive fitness of evolved lines and their unmutated ancestors. MA lines of most strains declined in fitness, on average. We sequenced the genomes of 85 MA lines, and identified a total of new 6,843 mutations. There was substantial variation in the mutation rate among strains, and evidence for context-dependent heterogeneity in the mutation rate and clusters of mutations. Fitness and the total numbers of accumulated mutations and coding mutations in the genome were significantly negatively correlated. We fitted a model of the relationship between mutation number and fitness, and found that models containing one to two mutational effect categories (one neutral and one deleterious category) fitted the data satisfactorily in most cases. We have crossed the MA lines to their respective unmutated ancestors to generate populations of backcross genotypes segregating for the accumulating mutations, and have allowed populations to evolve in competition. Mutations are expected to initially segregate at a frequency of one-half. Changes in allelic frequency of the mutations (which are assayed by deep sequencing) can be directly related to the distribution of fitness effects.

POB-366

Exploring the mutational landscape of TEM-1 beta-lactamase reveals mechanism of bacterial death

Andre Birgy[1, 2], Herve Jacquier[1, 3], Melanie Magnan[1], Sebastien Fleurier[4], Karine Panigoni[1], Julie Lasvergnas[1], Audrey Chapron[1], Ivan Matic[4, 5], Olivier Tenaillon[1]

[1]Universite Paris Diderot, Sorbonne Paris Cite, Paris (France), [2] Hopital Robert-Debre, AP-HP, Paris (France), [3]Groupe Hospitalier Lariboisiere-Fernand Widal, AP-HP, Paris (France), [4]Universite Paris Descartes, Sorbonne Paris Cite, Faculte de Medecine Paris Descartes, 75014 (France), [5]Centre National de la Recherche Scientifique, 75016 Paris (France)

Beta-lactamase TEM-1 is an antibiotic resistance enzyme that combine a medical importance, a fast evolution in the wild and is easy amenable to manipulation in the laboratory. As such, it has become models of biochemistry and also models for the study of protein evolution. The characterization of the distribution of mutational effect within a protein shed light on the molecular mechanisms and the constraints influencing the evolution of proteins.

Using an exhaustive mutagenesis approach followed by an experimental evolution under antibiotic selection coupled with high-throughput sequencing, we were able to determine and describe precisely the distribution of the effects of mutations in the TEM-1 protein. Three categories of mutants have been identified as having different behaviors in terms of survival kinetics when facing the antibiotic. Phenotypic studies have allowed us to propose a scenario involving a progressive saturation of the main penicillin-binding proteins as a function of the hydrolytic activity of the beta-lactamase mutant. Finally, a qualitative biochemical model compatible with this kinetic is proposed.

The coupling of this quantitative and mechanistic evolutive approache makes it possible both to advance in understanding the constraints underlying the evolution of proteins but also to plunge into the heart of resistance to antibiotics, and mechanisms of bacterial death.

POB-367

Protein stability potentially governing protein evolution

Ryo Kurahashi[1], Shunichi Tanaka[2], Satoshi Sano[1], Hiroyoshi Matsumura[3], Kazufumi Takano[1]

[1]Kyoto Prefectural University (Japan), [2]Ritsumeikan University (Japan), [3]Ritsumeikan University (Japan)

Evolution is unique to organisms. Regardless of being various creatures in the present day, the cause of evolution is common; replication error, gene duplication and transposon. Genetic mutations directly effect to derived proteins, the best samples for studying evolution. Although these mutations might allow protein to improve the activity or gain new function, most mutations unfortunately destabilize proteins. Thus, protein stability is a significant factor in evolution, however, its role remains unclear. Here, we performed random mutation experiments focused on protein stability to elucidate the role of protein stability in evolutionary processes. In the first round used an esterase from Sulfolobus tokodaii, approximately 20% of the variants displayed higher activity than wild-type protein (i.e., 20% evolvability). During evolutionary processes, the evolvability of each library depended on the stability of parent proteins, indicating that protein evolution is potentially governed by protein stability. Furthermore, decreased activity could be recovered during evolution by maintaining the stability of variants. This evolutionary process supports the nearly neutral theory of molecular evolution where mutation are slightly deleterious for activity but rarely fatal for stability. In the light of fitness landscape, these results show that protein stability is essential to exploring sequence space for improvement of activity like a stick for mountain- climbing. In order to apply these findings to protein engineering, we attempt to establish novel approach using a screening with stability as a selection pressure. This work allows us to observe processes of exploring sequence space over several generations.

POB-368

Detection and charaterization of deleterious variants in traditional chicken breeds

Chiara Bortoluzzi[1], Martijn Derks[1], Steffen Weigend[2], Martien AM Groenen[1], Hendrik-Jan Megens[1]

[1]Wageningen University & Research (Netherlands), [2]Institute of Farm Animal Genetics, Neustadt- Mariensee (Germany)

Deleterious variants are constantly generated by mutation, but are not always purged from the population. Understanding the extent and the reasons why deleterious variants persist in populations are of interest for conservation purposes, because these variants reduce reproductive fitness and genetic diversity when homozygous, which becomes especially relevant when populations decline.

Traditional chicken breeds offer a powerful model to address the role of demographic discontinuities and selection on deleterious variation, because of their intense artificial selection, small effective population size, and substantial genetic variation. Here, we describe and characterize the genome-wide distribution of low frequency deleterious and loss-of-function (LoF) variants in protein-coding genes of 76 whole-genome sequences from traditional breeds that represent different management, demographic, and selection histories, and conservation status. On average, we observed 4774 missense variants, of which 8.42% were deleterious and 70% tolerated. We identified 1203 stop-gain, 1648 frame-shift, 687 start-lost, and 715 stop-lost variants. A Gene ontology (GO) overrepresentation test showed that genes with deleterious and LoF variants were involved in the regulation of development processes, reproductive system development, ovarian follicle development, and oviposition. Regions of the genome under selection enriched for missense and LoF variants were involved in body weight and skeletal integrity.

We show that deleterious and LoF variants affect gene products involved in vital development and reproductive functions. Our findings confirm the importance of next-generation sequencing information in the design of breeding schemes and conservation programs.

POB-369

Exploring the biochemical changes driven by protein evolution among phosphoenolpyruvate carboxylases (PEPC), the key carbon fixation enzymes of a C4 grass family.

Chatchawal Phansopa[1], Jim Reid[1], Pascal-Antoine Christin[2]

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

In photosynthesis, carbon fixation in C_3 plants is largely limited by the inefficiency of Rubisco, particularly those in warm and arid climates. C_4 plants have overcome this by compartmentalising CO_2 fixation in mesophyll cells while decarboxylating C_4-acids in bundle sheath cells, hence minimising wasteful photorespiration by concentrating carbon around Rubisco. This study focuses on the biochemical properties of PEPC, the enzyme that carboxylate phosphoenolpyruvate (PEP) and bicarbonate to oxaloacetate in C_4 *Alloteropsis* species. Based on the previous transcriptomic analyses by our group, the cognate genes of various C_4 PEPC were amplified, cloned and expressed as recombinant proteins in bacteria. The enzymes were purified and assayed via established spectrophotometric methods to ascertain their Michaelis-Menten kinetics for PEP and bicarbonate, and sensitivity to known allosteric effectors, namely glucose-6-phosphate (activator) and malate (inhibitor). We analysed, with respect to their animo acid changes, the kinetic parameters of PEPC expressed by the native *ppc_1P3* and *ppc_1P6* genes and those with laterally acquired modifications from other lineages of the subtribe *Melinidinae*, *Andropogonae* and *Cenchrinae*. By generating via site-directed mutagenesis a library of C_4 PEPC with specific substitutions at key amino acid residues, we also explored how the mechanisms of PEPC might have driven protein evolution and considered the wider implications for the host.

POB-370

The Genomics of Adaptive Divergence with Gene Flow by Means of Experimental Evolution

Sergio Tusso[1, 2], Bart P.S. Nieuwenhuis[1, 2], Simone Immler[1, 3], Jochen B.W. Wolf[1, 2]

[1]Uppsala University (Sweden), [2]Ludwig Maximilians Universitat (Germany), [3]University of East Anglia (United Kingdom)

The rapid progress of genomic tools now allows investigating the molecular underpinnings of speciation and adaptation at base-pair resolution. Nonetheless, it remains a challenge to infer evolutionary processes from a single temporal snapshot of genetic variation incorporating both intrinsic genomic properties and external factors such as selective pressures, population structure and/or demographic history. Here, we study the build-up of genomic divergence as a function of controlled evolutionary histories in an experimental setup. 132 independent populations of sexually reproducing fission yeast were subjected to divergent ecological contrasts across a set of three different speciation modes ranging from allopatry to full sympatry. Genomic changes were tracked over more than 10 months of evolution (53 sexual and more than 1500 asexual generations). Fitness measurements show clear adaptation to the ecological conditions, but large differences between treatments with a marked effect of gene flow. This project yields fundamental insight into fitness effects of adaptive genetic variation and illustrates the importance of gene flow for adaptation and genetic divergence within and between populations.

POB-371

Genomic evidence for mitonuclear incompatibilities in interpopulation hybrids of the copepod Tigriopus californicus

Ronald S Burton[1], Thiago G Lima[1], Ricardo J Pereira[1, 2]

[1]University of California, San Diego (United States), [2]Ludwig Maximilians Universitat Munchen (Germany)

Mitochondrial function, essential to cellular energy production, requires the intricate interaction between genes located in both the nuclear and mitochondrial genomes. In allopatric populations, rapid evolution of mtDNA can lead to population-specific mitonuclear coevolution. Hybridization between populations may then disrupt coadaptation and result in reduced fitness. In the copepod *Tigriopus californicus*, hybrid breakdown in life history traits is consistent with this model mitonuclear incompatibility. The genomic architecture of mitonuclear incompatibilities remains unknown - how many genomic regions are involved and are those regions the same in independently evolving populations? Here we present results of two experiments addressing those questions. In both, reciprocal crosses produce hybrid populations with the same complement of nuclear alleles but alternate mtDNA haplotypes. After one or more generations, populations are sequenced to determine allelic frequencies at ~2 million SNPs that distinguish the parental populations. In one experiment, three pairs of populations were reciprocally crossed; sequencing of the F2 generation revealed multiple genomic regions where allelic frequency divergence between reciprocals suggests mitonuclear incompatibilities; the chromosomal regions identified involved 4-5 chromosomes and were highly cross-specific. In the second experiment, replicate populations from reciprocal crosses were maintained for nine generations. In hybrid populations that recovered fitnesses similar to parentals, changes in allelic frequencies again showed strong evidence for mitonuclear incompatibilities. Results from replicate populations suggest that fitness recovery may repeatedly involve some of the same genes. Combined, these results indicate that mitonuclear incompatibilities have a complex architecture that differs between populations, potentially generating genome-wide barriers to gene flow.

POB-372

Signature of positive selection in Langshan chicken comparison with the Indonesian breed and red jungle-fowl

Tatsuhiko Goto[1, 2], Raman A Lawal[1], John E Pool[3], Dong-Dong Wu[4], Ya-Ping Zhang[4, 5], Paul M Hocking[6], David W Burt[6, 7], Olivier Hanotte[1, 8]

[1]The University of Nottingham (United Kingdom), [2]Obihiro University of Agriculture and Veterinary Medicine (Japan), [3]University of Wisconsin-Madison (United States), [4]Chinese Academy of Sciences (China), [5]Yunnan University (China), [6]University of Edinburgh (United Kingdom), [7]The University of Queensland (Australia), [8]International Livestock Research Institute (ILRI) (Ethiopia)

We report here the population genomic analysis of full genome sequences of two chicken breeds (Langshan n = 8 and Kedu Hitam n = 10) and their wild ancestor (Red Junglefowl n = 5) for the identification of candidate regions under positive selection in Langshan, a breed characterized by an extreme phenotype (dark brown eggshell coloration). With an average of 6.9-fold genome coverage per bird 32,562,920 bi-allelic single nucleotide polymorphisms (SNPs) on chromosomes 1-28 were detected in the dataset (GATK Best Practices). To detect signatures of positive selection, the genetic differentiation index (*Fst*) in 10 kb sliding 20-kb windows (total 92,168 windows) were calculated among three populations, and population branch excess (PBE) of Langshan were computed for each windows. Genomic regions with higher PBE values were considered are likely candidate regions where Langshan-specific positive selection occurred. We set a significant threshold PBE > 0.8 and found 36 significant windows above the threshold. These windows define 20 candidate regions on chromosomes 1-6. Of these 20 regions, 13 (chromosomes 2, 3, 4, 5, and 6) overlap with seven eggshell color quantitative trait locus (QTL) regions, which were identified by four different QTL mapping studies. These results illustrate the usefulness of comparative population genomics approaches to understand the genetic basis of phenotypic diversity in chicken.

POB-373

Adaptation to new environment: the pH challenge

Rita Di Martino[1], Sara 1 Mitri

[1]University of Lausanne - UNIL (Switzerland)

Perturbations of environmental parameters such as pH or carbon sources induce microbes to develop various strategies to maintain high fitness in novel conditions. In this study, we focus on an *Agrobacterium tumefaciens* strain isolated from a four-species community growing in a metal-working fluid (MWF), a coolant and lubricating agent. To untangle the metabolic pathways of *A. tumefaciens* in this environment, we designed a minimal medium in which different concentrations of carbon sources commonly present in the MWF were added separately. Among these tested compounds, citric acid got our attention: after several days of incubation, *A. tumefaciens* switched from no growth to a sudden 20-fold increase compared to the starting inoculum. Growth occurred only at the highest-tested citric acid concentration simultaneously among all the replicates. Moreover, this specific concentration has a very acidic pH, and we verified that cells collected from the stationary phase could grow instantly in the same fresh medium but not in a neutralized citric acid medium. Our hypothesis is that the exposure to the acidic condition triggered an adaptive response. We are evaluating what has changed both at the genomic and transcriptomic level. Our aim is to understand the impact of pH on metabolic functions of *A. tumefaciens* and its ability to adapt to extreme environments. The result of this experiment will give us more insight on how bacteria react to external variations, highlighting which are the pathways responsible for fitness preservation in a new environment.

POB-374

Evolution and maintenance of non-male producing phenotypes in Daphnia

Christoph Haag[1], Yan Galimov[4], Celine Reisser[2], Cecile Molinier[1], Peter Fields[3]

[1]CNRS, Univ Montpellier, Univ Paul Valery Montpellier 3, EPHE, IRD (France), [2]IFREMER Centre du Pacifique, Tahiti (French Polynesia), [3]Univ. Basel (Switzerland), [4]Russian Academy of Sciences, Moscow (Russian Federation)

In *Daphnia*, parthenogenetic reproduction alternates with sexual reproduction, and sex is usually determined by the environment (ESD). However, some clones never produce males, neither in nature nor when stimulated with a male-inducing hormone. These non-male producing ("NMP") phenotypes are genetically determined females and participate in sexual reproduction only as females, fertilized by males from other clones (clones with ESD, called "MP" because they can produce males). We found that the NMP phenotype is determined by a large, W-like chromosome region. Several candidate genes involved in sex determination occur in this region, which is also enriched for genes that are differentially expressed between MP and NMP, even under control conditions when both produce females. Surprisingly, however, when exposed to the male-inducing hormone, gene expression changes more strongly in NMP than in MP individuals. This suggests that the NMP phenotype is actively regulated to prevent male production. Investigating the potential fitness effects of NMP, we found that, when rare, NMP phenotypes have an advantage due to obligate outcrossing. However, despite strong inbreeding depression, rates of inbreeding in MP phenotypes are too low to explain the frequency of NMP in this population (about 20%). Additional factors that may help to explain the relatively high frequency are partial male-specialization of MP clones and maladaptive (too early) onset of male production. Together with negative frequency-dependent selection on sex ratios when NMP phenotypes are common, our results suggest that the MP/NMP polymorphism is maintained by balancing selection acting on several genes in the W-like region.

POB-377

The architecture of adaptation: a master mutation or a mass of mutations?

Sophie Archambeault[1, 2, 3], Luis Baertschi[1], Catherine Peichel[1, 2, 3]

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

Genomic mapping of the loci of phenotypic evolution in multiple organisms has revealed genomic 'hotspots', or regions of the genome that control more traits than expected. This clustering of traits has important implications for the speed of adaptation, and could be due to pleiotropic effects of a single mutation or tight genetic linkage of multiple causative mutations. The threespine stickleback (*Gasterosteus aculeatus*) is a powerful model for the study of adaptive evolution because the marine ecotype has repeatedly adapted to freshwater environments across the northern hemisphere in the last 12,000 years. This adaptive process has resulted in the repeated fixation of a 16 kilobase genomic hotspot on chromosome IV that affects multiple traits, including defensive armor, sensory hair cells, and schooling behavior. We have performed fine-scale association mapping in a fully interbreeding, polymorphic population of freshwater stickleback in order to disentangle the relationships between genotype and phenotype in this genomic hotspot. Our findings suggest that the coordinated phenotypic changes are due to both pleiotropic effects of a single mutation and tight linkage of multiple causative mutations.

POB-378

Adaptations for dim light vision in cichlids: evidence from opsin gene family evolution and protein function

Frances E Hauser[1], Katriina L Ilves[4], Ryan K Schott[1], Gianni M Castiglione[1, 2], Hernan Lopez-Fernandez[5], Belinda SW Chang[1, 2, 3]

[1]University of Toronto (Canada), [2]University of Toronto (Canada), [3]University of Toronto (Canada), [4]Pace University (United States), [5]University of Michigan (United States)

The genetic underpinnings of sensory systems can offer important insight into the molecular basis of adaptation. Investigations of the visual system are particularly useful in establishing a relationship between genotype and phenotype, since vision allows for the near instantaneous relay of information between organisms and their external world. In cichlid fishes, extensive variation in opsin genes, which mediate the first step in vision, has likely contributed to their explosive diversification in African lakes. In contrast, the visual systems of cichlids that diversified in South and Central America has been hypothesized to be dim light adapted, in part due to the frequently low-light environments these fish encounter in the Neotropical rivers. Integrating phylogenomic, molecular evolutionary, and protein experimental approaches, we unravel the relationship between opsin gene family evolution, genetic variation, and opsin protein function in response to different visual environments during the diversification of Neotropical cichlids. First, we demonstrate that relaxed selection on the ultraviolet opsin in cichlids inhabiting dim environments has driven several independent instances of gene loss and pseudogenization in this opsin. Second, we identify and structurally model parallel amino acid substitutions conferring dim light adaptation in both cone and rod opsin proteins. Finally, we test the effect of a novel rhodopsin mutation in vitro, finding functional effects consistent with a protein phenotype optimized for dim light conditions in South American cichlids. Together, this work sheds light on the selective forces mediating sensory gene family reduction, as well as the molecular mechanisms driving opsin protein adaptation.

POB-379

Revealing SNPs of 0.1% effect in a diverse yeast cross through Barcoded Bulk QTL mapping

Katherine R Lawrence[1], Artur Rego-Costa[2], Michael M Desai[2], Alex N Nguyen Ba[2]

[1]Massachusetts Institute of Technology (United States), [2]Harvard University (United States)

Biologists have long sought to elucidate the genetic basis of complex traits, whether for human disease risk, model organism genetics, or agricultural outcomes. Studies of the genetic basis underlying quantitative traits have consistently observed missing heritability: not all of the genetic variance in these traits can be explained by the quantitative trait loci (QTLs) detected. Potential sources of this additional variance include numerous undetected small-effect loci as well as epistasis between QTLs, among others. In yeast crosses, QTL mapping sensitivity has been limited to fitness effect sizes of order 1%, due primarily to trade-offs between the number of segregants one can assay and the resolution of those genotype and phenotype measurements. Other studies in such crosses have indicated that up to hundreds of QTLs at sub-1% fitness effects may be significantly contributing to phenotypic variance. Here we demonstrate a QTL mapping analysis of a pool of yeast segregants that is 100-fold larger than previous work, allowing detection of QTLs with ten-fold smaller effect sizes and, in many cases, identification of their locations down to single nucleotides. Collecting high-resolution genotype and phenotype data on this scale is achievable and cost-effective due to a suite of novel techniques: robotic liquid handling, lineage barcoding, combinatorial indexing, custom enzyme purification, and bulk fitness assays. We compare the contributions to missing heritability of small-effect QTLs as well as epistatic interactions among QTLs across different complex traits, illuminating this regime of genetic architecture underlying phenotypic diversity.

POB-381

The joint fitness landscape of two genetically interacting genes

Xukang Shen[1], Jianzhi Zhang[1]

[1]University of Michigan (United States)

Fitness landscapes describe the genotype-fitness relationship and determine evolutionary trajectories. With the advent of high-throughput DNA sequencing and other genomic tools, empirical fitness landscapes of individual genes containing tens of thousands of mutational variants are being constructed. These studies showed that the intragenic fitness landscape is rugged due to abundant epistasis between mutations. However, epistasis between genes is currently known at a large scale only from double gene deletion studies. Because the epistasis between two genes may depend on the specific mutations in the two genes, the current data provide at most an incomplete picture of intergenic epistasis. We thus study the joint fitness landscape of two genes known to interact on the basis of double gene deletion studies: RPS25A and RPL41A. RPS25A encodes a small subunit ribosomal protein essential for binding of internal ribosome entry site RNAs to the 40S ribosomal head, while RPL41 encodes a large subunit ribosomal protein that interacts with the kinase CKII and stimulates the phosphorylation of DNA topoisomerase II-alpha by CKII. We first delete these two genes and confirm that their null mutations interact genetically. We then chemically synthesize these two genes with mutation and generate yeast strains that carry variants of the mutated genes. We compete all yeast strains together and measure their relative fitness using barcode sequence where the RPS25A and RPL41A gene sequences serve as the barcode of each strain. Data are being collected and analyzed at this time, and we expect to report our findings at the SMBE meeting.

POB-382

From DNA to ecosystems: using a model microbial community to study adaptation

Sandeep Venkataram[1], Jacob Robertson[1], Sergey Kryazhimskiy[1]

[1]University of California, San Diego (United States)

Recent discoveries have revealed the major role microbial communities play in modulating a range of biological processes, from human health and nutrition to global nutrient cycling. However, the evolution of such communities, and the consequences of evolution on community ecology, are poorly understood.

Here we study adaptation using a facultatively mutualistic model community consisting of the yeast *Saccharomyces cerevisiae* and the alga *Chlamydomonas reinhardtii*. In this community, the algae fix nitrogen, which is a limiting resource for the yeast, while the CO2 released by yeast during metabolism enhances algal photosynthesis.

We first seek to understand the stability of the community as a function of abiotic and biotic parameters. We hypothesize that sufficiently high fixed nitrogen concentrations or initial yeast densities would lead the yeast to outcompete rather than coexist with the algae. By studying community structure along a range of these parameters, we determine the conditions in which the community is no longer stable, and thus generate a quantitative description of mutualism breakdown in this system.

We will then utilize DNA barcode technology to study the evolutionary dynamics of yeast in the presence and absence of algae using otherwise identical growth conditions. This approach will let us directly test the effect of biotic interactions on adaptive dynamics and allow us to isolate clones containing single adaptive mutations for further physiological and ecological analyses. Overall, we hope to establish a powerful system for studying the interrelationship between ecology and evolution at many biological scales.

POB-383

Molecular and phenotypic characterization of *roo* elements inserted in the promoter of a *Drosophila* stress-response gene

Miriam Merenciano[1], Josefa Gonzalez[1]

[1]CSIC-UPF (Spain)

Transposable elements can produce a broad range of structural variants, and they can also contain cis-regulatory elements that may change the expression of nearby genes. We previously discovered nine independent *roo* solo-LTR insertions located in the 807 bp promoter region of *CG18446*, which overlaps with the first intron of *CG46338*. One of the identified insertions was associated with a cold-resistance phenotype. In this work, we further analysed 234 strains from 14 worldwide natural populations. In total, we found 20 independent *roo* solo-LTR elements inserted in the *CG18446* promoter region. To investigate whether other *roo* elements cluster in gene promoter regions, we analyzed all *roo* insertions in the *D. melanogaster* genome. We only found four genomic regions in which two *roo* insertions were present, suggesting that the previously described insertion cluster is unique. We also studied whether the two *roo*insertions found at higher population frequencies, *FBti0019985* and *roo*$_{-90}$ have any molecular or phenotypic effect beyond the already described cold-resistance phenotype. We found that *FBti0019985* flies are associated with resistance to ethanol exposure and to *Pseudomonas entomophila* infection. These functional changes are associated with changes in *CG18446* expression. Furthermore, in vivo enhancer assays suggest that *FBti0019985* contains cis-regulatory elements that could be responsible for the gene expression changes under infection. However, *roo*$_{-90}$ outbred flies did not show phenotypic differences in any condition tested. Overall, our results showed that at least one of the 20 insertions in the insertion cluster described are associated with changes in fitness-related traits.

POB-384

The impact of protein architecture on adaptive evolution

Ana Filipa Moutinho[1], Julien Yann Dutheil[1]

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

The flexibility of proteins in providing diverse and efficient solutions has enabled survival in constantly changing environments. Recent studies proposed that the distribution of fitness effects of new amino-acid mutations is influenced by mutation rate, recombination and effective population size. However, there is a lack of understanding on the extent at which different gene categories and structural variation among protein coding regions play a role in adaptive evolution. Here, we studied molecular evolution on a finer scale through the analysis of the impact of genetic variants in the different conformations of protein structure, aiming to understand how protein biophysics and protein evolvability influence fitness and adaptive evolution. By using 198 *Drosophila melanogaster* genomes, we measured the distribution of fitness effects of amino-acid substitutions on a gene and site basis, across different structural motifs. We found significant differences on the rate of adaptive non-synonymous substitutions among different structural properties, such as protein size and secondary structure organization. This suggests that adaptive mutations are constrained by biophysical properties of proteins, and are more likely to occur in certain regions of the molecules. Moreover, we observed that the cellular localization of proteins also plays a key role in adaptive evolution, with proteins bound to the plasma membrane showing the highest rate of adaptation, two-fold higher than nuclear and secreted proteins. These results demonstrate the broad impact of protein structure on the rate of adaptation and provide crucial knowledge on how, and to which extent, individual fitness is affected by protein biophysics.

POB-386

Positive selection and red-shifting substitutions in the rhodopsin gene of a globally distributed family of fishes making evolutionary transitions into freshwater.

Alexander Van Nynatten[1, 3], Gianni M Castiglione[2], Nathan Lovejoy R[1, 2, 3], Belinda Chang SW[1, 2]

[1]University of Toronto (Canada), [2]University of Toronto (Canada), [3]University of Toronto Scarborough (Canada)

Rhodopsin, the light-sensitive visual pigment expressed in rod photoreceptors, is exquisitely well adapted for vision in dimly-lit environments. However, rhodopsin's ability to detect light diminishes underwater where the visible spectrum narrows with depth, unless its sensitivity coincides with the wavelengths of light available. In deep-sea fishes, specific amino acid substitutions in the opsin protein component of rhodopsin blue-shift its peak spectral sensitivity to match the wavelengths of light penetrating deepest in off-shore marine waters. In contrast, coastal regions and large rivers are more turbid and tannin stained; red-shifting the underwater light environment. The molecular mechanisms involved in adaptation to red-shifted waters has not yet been well established. We compared rates of molecular evolution in rhodopsin sequences of a globally distributed family of fishes inhabiting off-shore, coastal and freshwater habitats. We find positive selection in rhodopsin, consistent with the diverse array of visual environments these fishes inhabit. Models comparing phylogenetic partitions, separating marine and freshwater lineages, indicate dN/dS is highest along the transitional branch from marine to freshwater in South America. This branch represents the most substantial red-shift in environmental light conditions and included among the positively selected sites are red-shifting substitutions F261Y and S124A. Recreating these and other substitutions on the branch in-vitro indicates the freshwater lineages have a red-shifted rhodopsin pigment compared to the most recent marine ancestor. This red-shifted rhodopsin pigment is more sensitive to the wavelengths of light penetrating deepest in Amazonian rivers, providing a clear adaptive advantage to this ancestrally marine clade of freshwater fishes.

POB-387

Applying phylogenetic tree-based approach to genome-wide association studies in Mycobacterium tuberculosis

Valentina Burskaia[1], Gennady Fedonin[2], Georgii Bazykin[1, 2], Alexey Neverov[3]

[1]Skolkovo Institute of Science and Technology (Russian Federation), [2] Institute for Information Transmission Problems (Russian Federation), [3]Central Research Institute for Epidemiology (Russian Federation)

Phylogenetic approach could significantly improve Genome-Wide Association Studies (GWAS) in microbial organisms, especially in cases of recurrent antibiotic resistance evolution. Phylogenetic approach allows to describe simultaneous appearance and coexistence of phenotype and genotype traits on independent branches of the tree. We use TreeWAS software as well as our original algorithms to find dependencies between genotype and phenotype traits. 5000 full genomes of Mycobacterium tuberculosis are used; the alignment contains 200 000 SNPs; 13 types of drugs are analyzed. Mycobacterium tuberculosis is almost ideal model organism for our goal: as there is almost no recombination in it, independent appearance of similar mutations in different lineages is nearly impossible (And so we never deal with hemiplasious allels).

POB-389

Selective landscapes of gene expression and other quantitative molecular traits in mammals

Marco Mariotti[1], Siming Ma[1], Toni Gabaldon[2], Vadim N Gladyshev[1]

[1]Brigham and Women's Hospital, Harvard Medical School, Boston, MA (United States), [2]Centre for Genomic Regulation, Universitat Pompeu Fabra, Barcelona (Spain)

Application of modern "omics" approaches across many species can help uncover the selective forces acting on molecular traits. We employed RNAseq and metabolite profiling to quantify gene expression and metabolite abundance across multiple tissues (brain, kidney, liver, heart) of dozens of species, encompassing all major mammalian lineages (primates, rodents, carnivores, and others).

This plethora of comparative data allows to study gene evolution combining phylogenetic histories with RNA levels and other relevant measures. However, there is a challenge: the inherent complexity of omics, with thousands of features quantified in each experiment, is aggravated by the phylogenetic dimension. To facilitate our analyses, we have developed a framework bringing together data science and evolutionary biology. Treedex, the tree-data explorer (http://treedex.org/), is designed to provide access to state-of-the-art evolutionary methods (R and Python packages) within an interactive visualization environment with phylogenetic perspective.

We focused our studies on gene expression. Since many species have no available genome, we predicted coding sequences in each transcriptome de novo, then applied phylogenomics to reconstruct gene histories and define orthology. We then evaluate the selective landscapes of RNA levels. For each one-to-one orthologous group, we fit and compare quantitative models representing scenarios of neutrality (Brownian motion), stabilizing selection (single-optimum Ornstein-Uhlenbeck) and directional selection (multiple-optima Ornstein-Uhlenbeck). Our results show evidence of extensive stabilizing selection, while directional selection is also observed in lineage- and gene-specific patterns. Our study highlights both the general features of gene expression evolution in the mammalian lineage, and the peculiarities of its various branches.

POB-392

Transcriptional and translational effects of adaptive synonymous mutations in *Pseudomonas fluorescens*

Nicholas McCloskey[1], Rees Kassen[1], Aaron Hinz[1]

[1]University of Ottawa (Canada)

Synonymous mutations have traditionally been thought to have no significant effect on fitness. However, a growing body of recent research has shown that this is not always the case. In an experimentally evolved population of *Pseudomonas fluorescens* grown in minimal glucose media, synonymous mutations arose in a glucose transport gene that resulted in beneficial fitness effects comparable to those of non-synonymous mutations. We found that the increase in fitness was a direct result of increased gene expression; however, the precise mechanism was unclear. Synonymous mutations have been shown to affect gene expression on transcriptional and translational levels through changes in mRNA secondary structure and codon usage.

Our study investigates the underlying mechanisms in which these evolved synonymous mutations lead to increased gene expression. In addition to the evolved mutations, we created a library of 45 strains with single synonymous mutations within the glucose transport gene and found a positive correlation between fitness and gene expression. To determine whether these mutations affect transcript levels, translational efficiency or a combination of both, we systematically incorporated transcriptional and translational fusions of a yellow fluorescent protein within the glucose transport operon. Additionally, through manipulation of the local genetic sequences and supplementation of rare tRNAs, we investigated the specific molecular requirements necessary for the increased expression. Our study aims to provide direct evidence of the adaptive mechanisms of beneficial synonymous mutations in an experimentally evolved setting.

POB-394

Systematic screening of compensatory mutations across a protein binding interface.

David Ding[1], Debora Marks[1], Mike Laub[2]

[1]Harvard Medical School (United States), [2]MIT (United States)

All proteins function by interacting with other biomolecules. To evolve new functions, proteins must change these interactions. This project aims to understand how single mutational steps contribute to new function without deleterious effects such as spurious crosstalk or loss of all interactions.

The system we use to study mutational trajectories is the bacterial toxin ParE and its antitoxin ParD. This system is advantageous because there are many paralogous copies, which have evolved specific interactions. Notably, a small number of mutations can alter the specific binding partner of an antitoxin. Also, it is possible to assay the effect of combinations of mutations in a high-throughput manner *in vivo*.

Specifically, we are studying how mutations in one binding partner constrain future mutational steps in the other binding partner. We are screening every mutation in the toxin for rescue of deleterious mutations in the antitoxin.

Further, we are comparing our experimental results with probabilistic models of coevolution. Specifically, we compare pairwise models of protein sequences with higher-order models. In the end, we also believe that this experimental dissection will help us build better statistical models, enabling us to improve predictions of the effect of mutations on protein interactions in general.

POB-396

Deep Mutational Scanning to Probe the Sequence-Function Relationships of Metallo-Beta-Lactamases

John Chen[1], Ray Socha[1], Nobuhiko Tokuriki[1]

[1]University of British Columbia (Canada)

Despite a wealth of specific knowledge on the properties of enzymes operating in a wide range of reactions, the fundamental questions of how the enzyme sequence renders to its function remains yet unanswered. Understanding of the enzyme sequence-function relationship can yield lucrative opportunities such as enhancing protein modelling techniques, which could bolster rational design of enzymes, aid prediction of mutation disease phenotypes based, and provide insight into evolutionary potential of bacterial drug resistance genes.

To study protein sequence-function relationships our lab engages in an approach known as deep mutational scanning (DMS), which aims to directly generate and characterize the effects of a comprehensive set of mutations on an enzyme. The model proteins we have chosen are VIM2, an antibiotic degrading enzymes from the metallo-beta-lactamase(MBL) superfamily. Each gene is mutated into a library of all possible single amino acid variants and placed under functional selection in Escherichia coli. Deep sequencing of the purified libraries post-selection provides enrichment levels for each variant, which is inferred to be proportional to its level of function.

I will discuss the results of screening the VIM2 variants against three different classes of beta-lactam antibiotics. We can gauge the functional capacity of all variants against each antibiotic to reveal individual sequence-function patterns, while comparing effects across different antibiotics can also inform us of trade-off phenotypic effects. We will also use additional biochemical and bioinformatics information to help inform the underlying mechanism of these behaviors, such as comparing functionality with Rosetta stability predictions for all variants.

POB-397

Gene, environment and cellular interactions underlying behavioral variance and their relation to fitness during experimental evolution

Luke Noble[1], Thiago Guzella[2], Matthew Rockman[1], Henrique Teotonio[2]

[1]New York University (United States), [2]ENS (United States)

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

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

POB-399

Directed evolution of ion-selectivity in bacterial and archaeal flagellar motors

Jessica Clark[1], Gonzalo Peralta[1], Josh M Machet[1], Mark M Tanaka[1], Matthew A B Baker[1]

[1]University (Australia)

The bacterial flagella motor is a crucial component of the bacterial chemotaxis pathway. This pathway produces a characteristic behaviour pattern consisting of long 'runs' interspersed with 'tumbles', which result in randomly changed direction. The conservation of the motor is indicative of its evolutionary benefit, allowing organisms to discover and exploit different environments. Recently, many advances have been made into understanding the process of chemotaxis, with a significant proportion of the work being focused on ion powered bacterial motors, usually those found in *Escherichia coli* (Colin and Sourjik, 2017). Previous work was performed retrospectively through statistical phylogenetics, but recent work in directed evolution has characterised evolutionary changes (Ni et. al., 2017). Here we used a combination of experimental evolution techniques, including error-prone PCR and site directed mutagenesis to induce and select for changes in the ion selectivity of the bacterial flagella motor. To do this we sectioned swim plates, and fluidics to select upmotile subpopulations directly from liquid media. Finally, we directly compared motility in evolved strains, *E. coli* strains, and the ATP-powered archaeal flagellar motor of *Haloferax volcanii* in order to examine convergent evolutionary phenotypes.

Colin, R. and Sourjik, V., 2017, 'Emergent properties of bacterial chemotaxis pathway'. Curr Opin Microbiol, 39, pp.24-33.

Ni, B., Ghosh, B., Paldy, F., Colin, R., Heimerl, T. and Sourjik, V., 2017, 'Evolutionary Remodeling of Bacterial Motility Checkpoint Control'. Cell Rep, 18(4), pp.866-877.

POB-401

The birth and the death of premature start codons in human genome is limited by selection

Vitaly Segodin[3], Svetlana Iarovenko[5], Stepan Denisov[1, 2, 3], Ruslan Soldatov[4]

[1]Skolkovo Institute of Science and Technology (Russian Federation), [2]IITP RAS (Russian Federation), [3]HSE University (Russian Federation), [4]Harvard Medical School (United States), [5]Lomonosov Moscow State University (Russian Federation)

Regulation of gene expression on the level of translation is less studied than the transcriptional regulation. Ribosome profiling technology reveals a class of short open reading frames located within 5'UTR - upsteam ORFs (uORFs). However, functional importance and evolution of these elements are poorly understood. We studied all short open reading frames having their start codon (uAUG) within 5'UTR. We divided them into the following types: uORF (uAUG and the stop codon located in 5'UTR), iORF (uAUG and the start codon of the main CDS are in the same frame and there is no stop codon between them), oORF (there is no stop codon between uAUG and the start codon of the main CDS, and they are not in the same frame). We studied cases of birth of new uORF, iORF and oORF in the human lineage after its divergence with Macaca mulatta. We estimated the birth rates of these sequences and compared them to a neutral control. We also calculated the death rate of these elements and evaluated selection acting on already existing uORFs, iORFs and oORFs. We found that uAUG, on average, appears as a slightly deleterious mutation, but already fixed uAUGs are mostly protected by the negative selection. Interestingly, iORFs die faster and get born slower than uORFs and oORFs. This implies that the extension of an N-terminus of a protein can lead to harmful effects, in agreement with previous studies. We also studied how ribosome reinitiation probability, Kozak context and other factors affect selection.

POB-403

Quantifying the effects of pleiotropic mutations on the robustness of signaling pathways

Purnima S. Kompella[1], Sergio G. Peisajovich[4], Alan M. Moses[1, 2, 3]

[1]University of Toronto (Canada), [2]University of Toronto (Canada), [3]University of Toronto (Canada), [4]Illumina (United States)

Cells receive stimuli from their surroundings and process them into physiological responses through signal transduction pathways. Point mutations are an important source of genetic variation and understanding how pathways respond to mutations can reveal how cells process information. Pathway robustness to mutations is thought to facilitate adaptation to future environmental or genetic changes. The mutational robustness of different signaling pathways can be studied quantitatively by examining the effects of mutations on shared components. Here, we measured the effects of over 100,000 random variants of a yeast gene, Ste50, on mating and HOG pathway responses by flow cytometry. These two pathways exhibit different dynamics in response to environmental stimuli. Through the action of network features such as feedback loops, the HOG pathway is more tightly regulated than the mating pathway. Therefore, we hypothesized that the HOG pathway would be more robust to mutations. We find that random mutations in Ste50 have a larger effect on the mating pathway than on the HOG pathway. Using our experimental system, we are also able to quantify, in parallel, the effects of mutations in Ste50 on insulation of pathway response. Furthermore, we assess the contributions of specific feedback loops to mutational robustness. Our analysis reveals quantitative differences in the mutational robustness of different signaling pathways and improves our understanding of the phenotypic significance of genetic variation in pleiotropic components of signaling pathways.

POB-405

The Physiological and Genomic Impact Intense Selection for Starvation Resistance in Drosophila melanogaster

Mark A Phillips[1], James Kezos N[2], Michael R Rose[1]

[1]University of California, Irvine (United States), [2]Sanford Burnham Prebys Medical Discovery Institute (United States)

An intense selection regime for increased starvation resistance was imposed on ten, large, outbred experimental Drosophila populations. We observed the response of starvation and desiccation resistance, metabolic reserves, and heart robustness via electrical pacing. As expected, significant increases in starvation resistance and lipid content were found in our selected populations relative to controls. We also found that selection for starvation resistance indirectly improved desiccation resistance, water content, and glycogen content among these populations. The rate of cardiac arrests in our ten selected populations was doubled when compared to controls. Finally, age-specific mortality rates were increased at early adult ages by selection. Next, we investigated the genetic basis of these physiological changes by comparing genomic data from our selected populations to their controls. We identified approximate twelve hundred differentiated SNPs between the two, which corresponded to one hundred and forty-five genes. The associated functions of our candidate genes covered a variety of biological processes. However, we found significant enrichment for genes associated with metabolic processes, and many of the genes we identified have known human orthologs associated with cardiovascular, renal, and metabolic disorders. Combined with this list of relevant candidate genes, the cardiac dysfunction, increased adult mortality, and elevated lipid levels in our selected populations make them a potentially useful model system for heart disease and obesity-related disorders.

POB-406

The transcriptomic landscape of yaks reveals molecular pathways for high altitude adaptation

Xuebin Qi[1], Qu Zhang[1, 9], Yaoxi He[1, 8], Lixin Yang[1], Linping Yang[2], Zhengheng Liu[2], Shiming Liu[3], Tianyi Wu[3], Chaoying Cui[4], Ouzhuluobu NA[4], Jianlin Han[5], Shengguo Zhao[6], Chunnian Liang[7], Bing Su[1]

[1]Kunming Institute of Zoology, Chinese Academy of Sciences (China), [2]Animal Husbandry, Veterinary and Forestry Bureau of Maqu County (China), [3]High Altitude Medical Research Institute (China), [4]Tibetan University (China), [5]Gansu Agricultural University (China), [6]Lanzhou Animal Husbandry and Veterinary Drug Institute, Chinese Academy of Agricultural Sciences (China), [7]Institute of Animal Sciences, Chinese Academy of Agricultural Sciences (CAAS) (China), [8]University of Chinese Academy of Sciences (China), [9]Perspective Sciences (China)

Yak is one of the largest native mammalian species at the Qinghai-Tibetan Plateau, and yet the genetic mechanism underlying its adaptation to high altitude environments remains elusive. We conducted a cross-tissue, cross-altitude and cross-species study to characterize the transcriptomic landscape of domestic yaks. The generated multi-tissue transcriptomic data greatly improved the current yak annotation by identifying tens of thousands novel transcripts. We found that among the eight tested tissues (lung, heart, kidney, liver, spleen, muscle, testis and brain), lung is the key organ showing adaptive transcriptional changes, and more than 90% of the cross-altitude differentially expressed genes (DEGs) in lung displayed a non-linear expression change with an optimal expression in samples from around 4,000 meters above the sea level, the average elevation of the Qinghai-Tibetan Plateau, to which yak has been adapted. In contrast, the majority of the cross-altitude DEGs in the other tissues showed a linear expression change, reflecting a pattern of acclimatization. Further, in comparison with lowland cattle and sheep, we identified 1,704 genes with yak-specific expression changes. The functional enrichment analysis revealed four pathways: PI3K-Akt signaling pathway, HIF-1 signaling pathway, focal adhesion pathway and ECM-receptor interaction pathway. These pathways are enriched with hypoxia-related DEGs and provide possible mechanistic explanation of genetic adaptation to high altitude in yak.

POB-407

Signals of selection in immune response genes of an admixed southern African population

Caitlin Uren[1], Eileen G Hoal[1], Gerard Tromp[1], Paul D van Helden[1], Brenna M Henn[2], Marlo Moller[1]

[1]Stellenbosch University (South Africa), [2]UC Davis Genome Center (United States)

The recent availability of exome sequence data and improved statistical analyses has facilitated investigations into the extent of selective pressure due to pathogens in numerous populations, even those that are admixed. However, there have been very few studies investigating this in southern African populations where it is hypothesised that the selective pressure due to tuberculosis and smallpox was vast. Here, we perform a positive selection scan using the population branch statistic to identify signals of selection associated with viral or bacterial immune response in the highly admixed South African Coloured (SAC) population. Using ancestral populations from the 1000 Genomes Project for comparison, we found SAC-specific signals of selection in genes associated with immune response to infectious disease. This study not only confirms the role that natural selection plays in shaping human immunity, it also highlights particular novel pathways associated with immune response to pathogens, that could be investigated further, particularly with respect to tuberculosis susceptibility in southern Africa.

POB-408

Founder-specific inbreeding depression affects racing performance in Thoroughbred horses

Evelyn Todd[1], Natasha Hamilton[1], Simon Ho[1], Peter Thomson[1], Brandon Velie[2], Rachel Ang[1]

[1]University of Sydney (Australia), [2]Swedish University of Agricultural Sciences (Sweden)

Thoroughbred horse racing plays an integral economic and historic role in society. Uniquely, the Thoroughbred horse population has been closed and selectively bred for over 300 years. In this study we utilized the extensive pedigree and phenotypic information available to examine the effects of selective breeding on mutational load. By analysing the relationship between inbreeding and racing performance of 135,996 individuals, our study shows that genetic load has persisted in the Thoroughbred population even after many generations of selection for athletic performance. However, we also found evidence for selection increasing the frequency of some favourable genes that might have improved racing performance over time. The Thoroughbred population originated from a narrow foundation bottleneck, resulting in over 80% of inbreeding in contemporary individuals being attributed back to a small number of ancestors from the foundation of the breed. We found that higher rates of inbreeding to some of these ancestors has had variable effects on the racing performance of contemporary Thoroughbreds, demonstrating that an understanding of the distribution of genetic load in a population is important in making decisions to improve its overall phenotypic value. By analysing one of the largest domestic animal populations in the world, out study provides unique insights into inbreeding that are not possible when analysing smaller populations where such comprehensive records are not available.

POB-409

The genomic basis of athletic performance and navigation in racing pigeons.

Malgorzata Anna Gazda[1, 2], Pedro Andrade[1, 2], Sandra Afonso[1], Jolita Dilyte[1], John Archer[1], Ricardo Lopes[1], Rui Faria[1, 3], Miguel Carneiro[1, 2]

[1]CIBIO (Portugal), [2]University of Porto (Portugal), [3]University of Sheffield (Poland)

Racing pigeons have been selectively bred to find their way home quickly over what are often extremely long distances. This breed is of substantial commercial value and is also an excellent avian model to gain empirical insights into the evolution of traits associated with flying performance and spatial orientation. Here, we investigate the molecular basis of the superior athletic and navigational capabilities of racing pigeons using whole-genome and RNA sequencing data. We inferred multiple signatures of positive selection distributed across the genome of racing pigeons. The strongest signature overlapped the CASK gene, a gene implicated in the formation of neuromuscular junctions. However, no diagnostic alleles were found between racing pigeons and other breeds, and only a small proportion of highly differentiated variants were exclusively detected in racing pigeons. We can thus conclude that very few individual genetic changes, if any, are either strictly necessary or sufficient for superior athletics and navigation. Gene expression analysis between racing and non-racing breeds revealed modest differences in muscle (213) and brain (29). These transcripts, however, showed only slightly elevated levels of genetic differentiation between the two groups, suggesting that most differential expression is not causative but likely a consequence of alterations in regulatory networks. Our results show that the unique suite of traits that enable fast flight, long endurance, and accurate navigation in racing pigeons, do not result from few loci acting as master switches but likely from a polygenic architecture that leveraged standing genetic variation available at the onset of the breed formation

POB-410

Signatures of local adaptation in human zinc transporters genes

Ana Roca-Umbert[1], Rocio Caro-Consuegra[1], Gabriel Felipe Rodriguez-Lozano[1], Nino Spataro[2], Elena Bosch[1]

[1]Institute of Evolutionary Biology (UPF-CSIC) (Spain), [2]Centre for Genomic Regulation (CRG) (Spain)

Zinc is an essential micronutrient involved in many different biological functions such as the immune system. In humans, zinc homeostasis is accomplished by the joint action of 24 zinc transporter genes (ZTGs) to ensure the molecular and cellular functions that depend on this micronutient. Until recently zinc intake in human populations directly correlated with the zinc content of the soil in which their crops were grown. Thus, zinc soil content may have been an important driver of local genetic adaptations. In the present study, we explore the complete set of 24 ZTGs for signatures of adaptation by using the 1000 Genomes sequencing data (Phase 3), which comprises 26 worldwide populations from 5 main geographical regions including zinc-deficient soils. Classical hard sweeps were analyzed per population and geographical region through the use of specific statistics such as Tajima's D, Fay and Wu's H test, iHS, XP-EHH and F_{ST}. Moreover, we also investigated polygenic selection by aggregating potential signals of adaptation across the complete set of ZTGs which were compared to random regions of our genome. Notably, at least three ZTGs showed significant signals for positive selection in Africa (*SLC39A4*) and East Asia (*SLC30A9* and *SLC39A8*). Additionally, we detected a general pattern of unusual concerted shifts in allele frequencies for the whole set of ZTGs, especially when comparing African to non-African populations. Finally, a number of putative adaptive SNPs are suggested for functional follow-up to help to pinpoint the true variants behind the detected signatures of selection.

POB-411

Evolution of alpha satellites in the human genome

Manatsu Hamazaki[1], Hideki Innan[1]

[1]SOKENDAI (The Graduate University for Advanced Studies) (Japan)

Approximately a half of the human genome consists of repetitive sequences, a part of which constitute tandem arrays and have some functions, such as the centromere and the telomere. It is well accepted that unequal crossing over and gene conversion play important roles to shape the configuration of repeats, but the evolutionary mechanisms in functional repetitive regions are not fully understood. How does selection operate to maintain sequence similarity between repeat units? How is the number of a repetitive sequence is determined? In order to address these questions, we focus on human centromeric regions, which comprise hierarchal tandem repeat arrays. The minimum repeat unit is a ~170 bp long sequence called an alpha satellite. Several alpha satellites form a several megabase array, called higher-order repeats (HORs), which is the repeat unit of the centromeric region. We here report detailed analyses of nucleotide sequence and copy number variation between human individuals and discuss the potential roles of selection behind the observed infraspecific variation.

POB-413

The polygenic basis of an ancient divergence in yeast thermotolerance

Carly Weiss[1], Jeremy Roop[1], Rylee Hackley[1, 2], Julie Chuong[2], Igor Grigoriev[1, 3], Adam Arkin[4, 5], Jeffrey Skerker[4, 5], Rachel Brem[1, 2]

[1]University of California Berkeley (United States), [2]Buck Institute for Research on Aging (United States), [3]US Department of Energy Joint Genome Institute (United States), [4]University of California Berkeley (United States), [5]Lawrence Berkeley National Laboratory (United States)

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

POB-415

The spectrum of loss of function tolerance in the human genome

Konrad Karczewski[1, 2], Laurent Francioli[1, 2], Kaitlin Samocha[1, 2], Beryl Cummings[1, 2], Daniel Birnbaum[1, 2], Mark Daly[1, 2], Daniel MacArthur[1, 2]

[1]Massachusetts General Hospital (United States), [2]Broad Institute (United States)

Deciphering the function and essentiality of genes in the genome is a central problem in human genetics. Large-scale exome and genome sequencing panels, such as the Exome Aggregation Consortium (ExAC) and the Genome Aggregation Database (gnomAD), have provided a glimpse into standing genetic variation, including loss-of-function (LoF) variation. The presence of LoF variants at high rates suggests a gene's redundancy, while a significant depletion suggests strong selective pressures against these variants, and thus, the gene's essentiality. Understanding where each human gene lies along the spectrum between these extremes is important for prioritizing candidate disease genes and the development of inhibitory therapeutics.

Using a mutational model, which accurately fits observed neutral synonymous variation, to generate expected numbers of variants for each gene in the genome, we previously found that in a dataset of 60706 exomes from ExAC, 3230 genes were found to be significantly depleted (constrained) for LoF variation, even in the heterozygous state. Here, we apply an expanded constraint model to high-confidence LoF variants in 123136 exomes from gnomAD. With this increased power, we identify approximately 15% more constrained genes, capturing more known disease genes and putative essential genes. Furthermore, we incorporate allele frequency information to distinguish weak selection against heterozygous variation from strong selection against homozygous variation. Finally, we investigate the extent of population-specific constraint across major continental groups. These results show how large-scale datasets can reveal natural selection on each gene, even before the responsible phenotypic consequences are discovered, improving interpretation of disease variants and therapeutic targets.

POB-418

Evolution and Molecular Mechanisms of Photoreceptor Transmutation

Ryan K Schott[1, 2], Belinda SW Chang[2]

[1]National Museum of Natural History, Smithsonian Institution (United States), [2]University of Toronto (Canada)

Vertebrates typically utilize a duplex retina that contains rod photoreceptors for dim-light vision and cone photoreceptors for bright-light vision. Snakes and geckos, however, have a predominance of simplex retinas that contain only one photoreceptor type. Evolutionary transitions between photoreceptors, termed photoreceptor transmutations, have been shown to explain the evolution of these simplex retinas, but little work has focused on the molecular evolutionary underpinnings of the observed morphological changes. To address this, we sequenced new visual gene data using transcriptomic and targeted capture approaches and used this data to analyze gene loss and shifts in selection pressures on phototransduction genes. Neither geckos nor snakes showed evidence of gene loss associated with photoreceptor transmutation, but both showed loss associated with ancestral shifts to nocturnality. Surprisingly, we found that geckos still express most rod phototransduction genes despite the ancestral loss of rods and a functional rod opsin. This implies that rods genes are expressed in cones, which may contribute to their observed rod-like physiology. Additionally, we found that phototransduction genes in both geckos and snakes were under significantly different selective pressures compared to other reptiles, but similar selective pressures as each other. In snakes with photoreceptor transmutation, we found widespread positive selection on cone-specific genes that may indicate adaptation during the evolution of rod-like cones. Together these results demonstrate that evolution of complex systems can occur through multiple mechanisms simultaneously, and that both environmental constraints and historical contingency can play important roles in the formation of new cell types with convergent functionality.

POB-419

Identifying host factors that modulate the outcome of influenza infection in humans

Mary O'Neill[1, 2, 3], Maxime Rotival[1, 2, 3], Helene Quach[1, 2, 3], Lluis Quintana-Murci[1, 2, 3]

[1]Institut Pasteur (France), [2]CNRS (France), [3]Institut Pasteur (France)

There exists considerable individual and population variation in susceptibility to infectious diseases, but the role host genetics plays in driving the outcome following pathogen exposure remains elusive. Here, we utilize a systems immunology approach - combining genomic, transcriptomic, and epigenomic data, as well as demographic and biological resources for 200 individuals - to investigate host determinants of susceptibility to influenza A virus (IAV) infection. We uncover high inter-individual variability in RNA sequence reads of viral origin (1-13.5%) following challenge of primary monocytes to a live IAV strain, suggesting differences in the ability of IAV to enter and/or replicate among host cells. Interestingly, the average percent of viral-mapping reads differs between African- (4.9%) and European-descent (6.8%) samples (p-value<1.9E-8), a finding that is replicated at the transcriptional (IAV mRNAs) and translational (IAV protein) levels. We find that host gene expression can be predictive of viral mRNA levels (6h post-infection), with 3% and 12% of active genes correlating with the phenotype in the basal (unstimulated) or following TLR7/8 (antiviral) stimulation, respectively. To quantify the contribution of host factors to the inter- and intra-population variability in the outcome that follows influenza infection, we aim to integrate identified host molecular phenotypes, and eQTLs of these loci, into a predictive model of viral mRNA levels. This work contributes to our growing understanding of human genetic factors driving individual and population variation in immune responses, and highlights the value of studying intermediate phenotypes to dissect complex traits such as disease susceptibility.

POB-420

Cross-species phylogenetic genome-phenome approach to understand the evolution of ageing in mammals

Xavier Farre[1], Gerard Muntane[1, 2], Arcadi Navarro[1, 3, 4]

[1]Universitat Pompeu Fabra (Spain), [2]Universitat Rovira i Virgili, Biomedical Network Research Centre on Mental Health (CIBERSAM) (Spain), [3]Universitat Pompeu Fabra (Spain), [4]The Barcelona Institute of Science and Technology (Spain)

Although well studied, the nature of the aging process remains unclear. Aging can be broadly defined as the inevitable, irreversible and time-dependent functional decline process that affects most living organisms paired with a decrease in fertility and an increase in mortality. However, there is no simple explanation as to why some species do not display the increasing mortality or decreasing fertility commonly linked with aging, which indicates that aging patterns are broad but not universal.

Most studies on aging have so far focused on single species, however recent comparisons of genomic, transcriptomic, and metabolomic data across lineages with different lifespans are unveiling molecular signatures associated with longevity. Here, we examine the relationship between genomic variation and maximum lifespan (MLS) across mammal species, by using the power of cross-species phylogenetic comparative methods to find signals of convergent evolution and to finally identify genes, pathways and/or genetic variants related to senescence.

We have analysed protein multiple sequence alignments for 62 mammals to find parallel substitutions that were recurrent in species with increased MLS across the mammal tree correcting by their body weight.

We also checked whether the same patterns of lifespan were conserved in non-mammal species containing the same amino acid distribution, using species from the Aves, Reptilia and Teleostei amongst others.

We identified a set of genes containing parallel substitutions which showed an enrichment in inflammatory response, blood coagulation and immune system pathways, which have already been previously linked with aging using other experimental methods.

POB-422

Detecting the signature of epistatic selection in subdivided populations

Champak Beeravolu Reddy[1], Frederic Guillaume[1], Katalin Csillery[1, 2]

[1]University of Zurich (Switzerland), [2]Swiss Federal Research Institute WSL (Switzerland)

There is increasing evidence that adaptation at quantitative or complex phenotypic traits likely occurs via subtle, potentially correlated, allele frequency shifts at multiple loci simultaneously. However, detecting the signature of such polygenic adaptation remains difficult with most currently available methods. Epistasis, the non-additive effect of gene interactions, can play a central role in polygenic adaptation when it creates specific combinations of alleles that are favored by selection thereby resisting the randomizing effect of recombination. Theory also predicts that directional epistasis may increase the additive genetic variance of a population subject to directional selection during adaptation. Here, we study the multilinear epistatic model of Hansen and Wagner (2001) as implemented in Nemo, an individual-based, forward time simulator. Our approach explicitly accounts for pairwise epistasis between loci affecting a quantitative trait and is able to deliver genomic data for a large variety of demographic scenarios. We investigate the potential to detect signals of polygenic adaptation and epistatic selection (selection for coadapted gene combinations) under scenarios of divergent and homogenizing selection between two populations using population and quantitative genetic approaches. Our population genetic approach is based on Ohta (1982), who decomposed linkage disequilibrium (LD) into its within and between population components to detect epistatic selection relative to the background LD due to physical linkage, drift, and population structure. On the other hand the quantitative genetic approach is based on detecting the excess covariance of allelic effects. We illustrate the potential of these approaches in a two population model with migration and mutation.

POB-423

A new inference method for ongoing selective sweep

Naoko T Fujito[1], Yoko Satta[1], Toshiyuki Hayakawa[2, 3], Naoyuki Takahata[1]

[1]SOKENDAI (The Graduate University for Advanced Studies) (Japan), [2]Kyushu University (Japan), [3]Kyushu University (Japan)

A simple method was developed to detect signatures of ongoing selective sweep in single nucleotide polymorphism (SNP) data. Based largely on the traditional site frequency spectrum (SFS), this method additionally incorporates linkage disequilibrium (LD) between pairs of SNP sites and uniquely represents both SFS and LD information as hierarchical "barcodes." This barcode representation allows the identification of a hitchhiking genomic region surrounding a putative target site of positive selection, or a core site. Sweep signals at linked neutral sites are then measured by the proportion (F_c) of derived alleles within the hitchhiking region that are linked in the derived allele group defined at the core site. Coalescent simulators with and without positive selection were used to assess the false positive and negative rates of the $F_c$statistic. To demonstrate the power, the method was further applied to three kinds of genomic data: (1) established cases of positive selection in human populations such as *LCT*, *OCA2*, *EDAR*, *SLC24A5*, and *ASPM* loci, (2) a technically challenging case of very partial soft selective sweep such as the *ST8SIA2* locus, and (3) our current interest of the 108 published schizophrenia-associated loci, which were identified by GWAS, in the African, European and East Asian meta-populations. Overall, the new method is powerful and can be used to identify SNP sites responsible for ongoing selective sweep.

POB-424

Addiction, Immunity, and Infectious Disease in a Global Population

Latifa Jackson[2, 1], Maksim Shestov[3]

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

Understanding complex disease related traits like substance addiction has been hampered by the lack of functional insights in to the human genome. Nonetheless, immune function degradation in long term addicts occurs despite variation in proximate metabolic processing mechanisms. This observation led to the hypothesis that environmental factors may determine allele frequency differences for immune function and addiction hotspots in certain ethnic populations living predominantly in tropical and subtropical environments. We identified addiction (N=587 genes) and immune (N=2640 genes) curated by NCBI. These genes were projected onto the genome to identify cluster genomic regions for immunity and addiction. Hotspots were defined as regions of the genome with more than 15 genes within a 1.5Mb linear genomic window. When gene lists were mapped, three hotspots were located on chromosomes 11, 17, and 19._ Human polymorphism data was sorted by ecoregion then surveyed from the Human Genome Diversity Panel (N=1054 individuals/51 populations), the HapMap Project (N=1148 individuals/11 populations) and the 1000 Genomes dataset (N=1092 individuals)._ Our analyses suggest that when tropical living human populations are grouped into tropical versus non-tropical living groups, significant differences in allele frequencies are identified for five functionally important variants on chr11. These polymorphisms were then assessed against known infectious disease distributions and found to be related to Hepatitis B disease patterns in central Chinese and a tropical sub-Saharan African population. These results suggest that diseases such as Hepatitis B may contribute to the allele frequency differences seen in variants putatively involved in addiction phenotypes.

POB-426

Inference of Microevolutionary Dynamics for Quantitative Molecular Phenotypes

Shadi Zabad[3], Alan Moses[1, 2, 3]

[1]University of Toronto (Canada), [2]University of Toronto (Canada), [3]University of Toronto (Canada)

Understanding the long-term effects of microevolutionary processes on quantitative traits has been a central aim of evolutionary quantitative genetics. Following the foundational work of Felsenstein (1973) and Lande (1976), these effects and their complex interactions have been studied under the framework of stochastic diffusion models. Two models in particular attracted the attention of the evolutionary modeling community: Brownian motion and the Ornstein-Uhlenbeck (OU) process. In studies of phenotypic evolution, the former has come to be treated as the de facto "null" model of neutral evolution whereas the latter is used to infer stabilizing selection and adaptive shifts. However, the interpretation of the two models has proved challenging and the assumptions underlying them have been questioned. In this work, we examine the reliability of this modeling approach in the context of quantitative molecular phenotypes. Our analyses indicate that the inferences drawn from these models are often unreliable. We investigate a number of factors, both statistical and biological, that could explain the limitations of this traditional modeling approach. As an alternative, we propose an inference pipeline that enables us to discover a rich class of stochastic models corresponding to various microevolutionary processes.

POB-429

A genetic handicap approach: how to estimate the genome-wide burden of slightly-deleterious variants

Konstantin Popadin[1], Alexandre Reymond[1]

[1]University of Lausane (Switzerland)

All organisms harbor numerous slightly-deleterious (SD) variants in their genomes. However, the composition of this burden, as well as its total effect on fitness, are poorly known due to extreme complexity of the system: thousands of SD variants with very small individual phenotypic effects can interact with each other in non-additive ways, making reconstruction of fitness from genome data almost impossible. Taking into account widespread negative epistatic interactions between SD variants (doi: 10.1126/science.aah5238) we formulated a genetic handicap approach, claiming that the healthy (fit) carrier of severely-deleterious variant (hereafter genetic handicap) is expected to have decreased burden of SD variants genome-wide as compared to controls (organisms without severely-deleterious variants). If so, the genetic handicap can be used in evolutionary and population studies as a marker of potentially fitter genomes. Recently, considering trisomy 21 as a genetic handicap in humans (doi: 10.1101/gr.228411.117) we got the first empirical evidence, supporting our approach: live-born Down Syndrome patients have decreased burden of SD variants compared to control individuals. Here we investigate the genetic handicap approach addressing both theoretical, empirical and experimental levels: (i) Under which types of selection (e.g. hard / soft; different strength and direction of epistasis) does genetic handicap approach works? (ii) Which metrics if any (e.g. polygenic risk scores, transcriptional risk scores) help to uncover the genetic handicap effects on genomic and transcriptomic population data? (iii) How can we design experiments with populations of model organisms to estimate their mutation burden?