POA-003

Dicyemid mesozoan genome reveals adaptations to the parasitic lifestyle

Tsai-Ming Lu[1], Hidetaka Furuya[2], Miyuki Kanda[3], Noriyuki Satoh[1]

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

Parasitism has independently occurred more than 200 times across 15 animal phyla, yet remains a topic of debate that how free-living ancestors evolved to parasitic organisms. Dicyemid mesozoans are microscopic endoparasites inhabiting the renal sacs of some cephalopods. They possess simplified body organization without differentiated organs and have long fascinated biologists because of their incompletely known lifecycles. Obtaining genomic data from enigmatic parasites would be essential to better comprehend parasitism evolution. Here we decoded the genome of *Dicyema japonicum*which is approximately 68 Mbp with an extraordinarily shortened intron size of 38.2 bp on average. Comparing among bilaterians, *D. japonicum* retained fewer genes in most KEGG pathways, and four parasite species from different phyla showed a convergent gene number reduction in the metabolism pathways. In contrast, *D. japonicum*exhibited multi-copy gene clusters associated with endocytosis and membrane trafficking, perhaps reflecting its specialized nutrient-uptake strategy. Up-regulated transcripts at dispersal larvae stage were over-represented on gene ontology terms of motor activity and response to the stimulus. Taken together with immunostaining signals of neurotransmitters and neuropeptides occurring on apical cells, dicyemids may have potential mechanisms to sense environmental cues and could actively approach new hosts. In summary, the dicyemid genome provides a resource to uncover mysterious lifecycle of dicyemids, as well as studying comparative genomics to gain insights into the parasitism evolution. Furthermore, genomes of parasites may adapt through eliminating genes which are not necessary for parasitic lifestyle or increasing gene copies corresponding to lineage-specific biological processes.

POA-004

The cycad coralloid root contains a diverse endophytic bacterial community including cyanobacteria encoding specific biosynthetic gene clusters

Angelica Cibrian-Jaramillo[1], Francisco Barona-Gomez[1], Antonio Corona-Gomez[1], Karina Gutierrez-Garcia[1], Pablo Cruz-Morales[1], Pablo Suarez-Moo[1], Nelly Selem-Mojica[1], Miguel A. Perez-Farrera[2]

[1]CINVESTAV (Mexico), [2]Universidad de Ciencias y Artes del Estado de Chiapas (Mexico)

Cycads are the only early seed plants that have evolved a specialized coralloid root to host endophytic bacteria that fix nitrogen for the plant. To provide evolutionary insights into this million-year old symbiosis, we investigate the phylogenetic and functional diversity of its endophytic bacterial community, based on the (meta)genomic characterization of the coralloid root of several *Dioon* species collected from natural populations. We employed a co-culture-based metagenomics experimental strategy, termed EcoMining, to reveal both predominant and rare bacteria that were analyzed through phylogenomics and detailed metabolic annotation. Most of the characterized bacteria were identified as diazotrophic plant endophytes belonging to at least 18 different bacterial families, although there seems to be a 'core' of only a few genera. The draft genomes of several Cyanobacteria strains were obtained, and after whole-genome inferences they were found to form a monophyletic group, suggesting a level of specialization characteristic of co-evolved symbiotic relationships. In agreement with their large size, the draft genomes of these organisms were found to encode for biosynthetic gene clusters predicted to direct the synthesis of specialized metabolites present only in these symbionts. Overall, we provide a new notion of the composition and evolution of the cycad coralloid root that contributes to studies on phylogenetic and functional patterns in plant-bacteria symbiotic systems.

POA-007

The coming and going of mutualistic symbionts: The ins and outs of co-obligate endosymbiont replacement in Cinara aphids

Alejandro Manzano-Marin[1], Armelle Coeur d'acier[1], Anne-Laure Clamens[1], Celine Orvain[2], Corinne Cruaud[2], Valerie Barbe[2], Emmanuelle Jousselin[1]

[1]INRA, CIRAD, IRD, Montpellier SupAgro, Univ. Montpellier (France), [2]Institut de Biologie Francois Jacob-Genoscope (France)

Nutritional-based mutualistic associations between microorganisms and insects with unbalanced diets is pervasive, with several insects having developed obligate associations with either bacterial or fungal symbionts. The last common ancestor of *Cinara* aphids (Hemiptera: Aphididae) harboured two obligate nutritional endosymbionts: *Buchnera* (the primary obligate endosymbiont of most aphids) and most probably a *Serratia symbiotica*. The latter shows a very dynamic pattern of replacement by different bacterial taxa, thus offering a special opportunity to study the turnover of obligate mutualistic associations. We have assembled the endosymbionts' genomes from over 60 species of *Cinara*. We found that most secondary co-obligate endosymbionts belong to taxa commonly associated to aphids as facultative endosymbionts, hinting at the origin of these lineages. Also, we observed common patterns of genome reduction and convergence in metabolic pathway retention. Within a group of *Erwinia*-associated aphids, horizontal transfer of B-vitamin-biosynthetic genes between endosymbiotic lineages has played an important role in the establishment of the new mutualistic associations. Finally, in the species *Cinara strobi*, we found that while the species has already acquired a new co-obligate symbiont, it retains the former *S. symbiotica* symbiont. It keeps a large but highly degenerated genome and has become unable to biosynthetise all B vitamins and essential amino acids, rendering it unable to fulfil its mutualistic role. In brief, the results provided by analysing the genomes of these obligate di- and tri-symbiotic consortia are illuminating the different processes involved in the evolution (from birth to death) of new obligate mutualistic associations between bacteria and insect hosts.

POA-009

Population genomics and co-evolutionary dynamics of Wolbachia-host symbiotic interaction in different host species

Kun D. Huang[1, 2], Matthias Scholz[2, 1], Davide Albanese[2], Claudio Donati[2], Nicola Segata[1], Omar Rota-Stabelli[2]

[1]University of Trento (Italy), [2]Fondazione Edmund Mach (FEM) (Italy)

Wolbachia are intracellular endosymbionts capable of manipulating the physiology of their arthropod and nematode hosts. Previous studies showed that although *Wolbachia*are generally maternal inherited, horizontal transfer between unrelated host species is common: this has provided a milestone in understanding co-evolution of *Wolbachia*and their hosts. Very little is known about evolutionary dynamics within the same host species: this is because tracking short-term events requires a large amount of data and multiple *Wolbachia* genomes from the same host species. To tackle this issue, here we used data from a collection of over 1000 *Wolbachia* genomes assembled metagenomically from publicly available sequencing projects to build 11 host-specific population datasets for many insect or nematode species. We first compared phylogenies of *Wolbachia* genomes with those of their host mitochondrial genomes. We observe conflicting topology in most of the populations: this indicates that horizontal transfer is a common inheriting mechanism also within individuals of the same species. We then used population genomics to investigate selective pressure and found that bottleneck occurred in different *Wolbachia* populations. We finally inferred divergence times for the *Wolbachia* strains in the 11 populations, using Bayesian relaxed clocks and multiple sequentially Markovian coalescent (MSMC) methods. We found several instances of chronological discrepancies between *Wolbachia* and their hosts. Our first comprehensive population genomics study of *Wolbachia* gives a fresh data-driven insight into the mechanisms behind symbiotic interaction at the host population level, and increases our understanding of *Wolbachia* complex evolutionary biology.

POA-010

Global shifts in gene expression profiles accompanied with environmental changes in cnidarian-dinoflagellate endosymbiosis

Yuu Ishii[1], Shinichiro Maruyama[1], Yusuke Aihara[2], Takeshi Yamaguchi[3], Katsushi Yamaguchi[4], Shuji Shigenobu[4], Hiroki Takahashi[3], Masakado Kawata[1], Naoto Ueno[3], Jun Minagawa[2]

[1]Tohoku University (Japan), [2]National Institute for Basic Biology (Japan), [3]National Institute for Basic Biology (Japan), [4]National Institute for Basic Biology (Japan)

Coral reef ecosystems rely on stable symbiotic relationships between cnidarian animals and the dinoflagellate alga *Symbiodinium* spp. Recent studies have shown that elevated seawater temperature can cause collapse of endosymbiosis by expulsion of the symbiotic algae from cnidarians, which is known as 'bleaching', and subsequent mass mortality. Toward further understanding of mechanisms underlying the collapse of the endosymbiosis between cnidarian hosts and symbiotic algae, we conducted RNAseq analyses using the symbiotic and apo-symbiotic forms of the model sea anemone *Exaiptasia pallida*, incubated under normal and high temperature conditions. We detected a number of symbiosis-dependent, high temperature-responsive differentially expressed genes, which were considered to be related to the process of endosymbiosis collapse. Considering heat stress as a trigger of bleaching, we defined the DEGs of which the expression levels were changed in the same direction in temperature elevation ('symbiotic-high' relative to 'symbiotic-normal') and de-symbiotization ('apo-symbiotic-normal' relative to 'symbiotic-normal'), and here we called them heat-induced bleaching-associated (HIBA) genes. We performed GO term enrichment analysis on the HIBA and non-HIBA groups in the symbiosis-dependent temperature-responsive DEGs. Using the HIBA genes, the terms related to carbohydrate and protein metabolisms were enriched, while in non-HIBA genes the enriched terms were 'transmembrane transport' and 'proteinaceous extracellular matrix'. Many of the HIBA genes were annotated to be enzymes catalyzing a number of complex forms of carbohydrates, suggesting that the genes may encode the enzymes catabolizing carbohydrates transferred from symbionts to host cnidarians under normal condition and that their expressions were suppressed under heat-stress condition.

POA-015

Identification of fungal and algal genes involved in the symbiosis of lichen Usnea hakonensis

Mieko Kono[1], Yoshiaki Kon[2], Yoshihito Ohmura[3], Yoko Satta[1], Yohey Terai[1]

[1]SOKENDAI (The Graduate University for Advanced Studies) (Japan), [2]Tokyo Metropolitan Hitotsubashi High School (Japan), [3]National Museum of Nature and Science (Japan)

Lichens are symbiotic organisms that consist of fungi, photosynthetic organisms (algae and/or cyanobacteria), and bacteria. They are adapted to diverse environments including extreme conditions where each constituent organism could not survive alone. Although lichens are considered as one of the most successful symbiotic organisms, genes that function in lichen symbiosis remain largely unknown. In this study, we aimed to identify the genes involved in symbiotic interaction between the fungus and the alga, the core interaction of symbiosis, by comparing gene expression between non-symbiotic and symbiotic states of lichen *Usnea hakonensis*. The symbiosis of *U. hakonensis* can be resynthesized in a laboratory condition by co-culturing fungal and algal isolated cultures. We searched for the genes differentially expressed between the non-symbiotic state (isolated cultures) and the symbiotic state (resynthesized lichens) as symbiosis-related genes. First, we determined the genomes of the fungus and the alga by using high-throughput sequencing techniques. Then, we performed RNA-seq to quantify the gene expression in the non-symbiotic and symbiotic states. In total, approximately 500 fungal and 400 algal genes showed higher expression in the symbiotic state. According to the predicted functions of these genes, it was inferred that in the symbiotic state, photosynthetic products are transported from the alga to the fungus, while nutrients such as nitrogen and phosphate are transported from the fungus to the alga. The results implied that the expression of genes involved in bidirectional transport of nutrients between the fungus and the alga may be the prerequisite for lichen symbiosis.

POA-018

Unveiling the architecture and evolution of microbial genomes from their homologous sequences

Kaoru Yoshida[1]

[1]Sony Computer Science Laboratories, Inc. (Japan)

Homologous sequences are accumulated in the evolution of a genome through duplication, conversion and recombination events, while individual homologous sequences may be divided into multiple fragments or structured with other neighbors to migrate together.

We previously introduced a new method for genome analysis, which uses only sequence information of a single genome, autonomously collects homologous sequences and their fragments distributed in the genome, assembles them into domains, hierarchically classifies the domains into families, and visualizes them on the genome. For over 1500 microbial strains, their genomes were individually analyzed with this method. Together with homologous domains, including paralogs and transposons, evolutionary relationships among endogenous molecules in individual genomes were revealed.

Recently, we have extended the method to globalize the homologous domain families (HDFs) found in individual genomes so that they can be compared among different genomes. In the case of Cyanobacteria composed of 60 strains, total 12263 HDFs were found, of which 308 HDFs were shared by multiple strains and the rest were intragenomic. Relationships among the strains sharing HDFs are shown in the form of a network.

While conventional molecular phylogenetic approaches generally focus on mutation and vertical gene transfer, our approach focuses on gene duplication, recombination, transposition and horizontal gene transfer and also takes plastic domains, rather than raw sequences, as homologous units. Therefore, it will be useful for elucidating the evolutionary dynamics of genome architectures as well as capturing critical genomic regions related with interactions of different organisms.

POA-019

Coevolution of Termites and Their Microbiomes?

Xianfa Xie[1], Alonzo Anderson[1], Latoya Wran[1], Myrna Serrano[2], Gregory Buck[2]

[1]Virginia State University (United States), [2]Virginia Commonwealth University (United States)

Termite gut microbiomes serve critically important function in breaking down the cellulosic plant materials that termites feed on. Whether this itself would create a co-evolutionary relationship between termites and their microbiomes remains to be debated. While some studies suggest that termite gut microbiomes are vertically transferred from adult to young termites, thus creating a suitable condition for co-evolution, some other studies suggest that the termite microbiomes could be shaped by the environment. We try to address this question by studying the microbiomes, particularly in relation to cellulose degradation, in different termite colonies as well as in their living environment in soil, and the results from this study are surprising.

POA-020

Coprolites reveal microbial symbionts facilitated the adaptation to environment of the extinct cave goat Myotragus balearicus

Yichen Liu[1], Luis Arriola[1], Jamie Wood[3], Josep Alcover Antoni [2], Joan Pons[2], Bastien Llamas[1], Alan Cooper[1], Laura Weyrich[1], Pere Bover [1]

[1]The University of Adelaide (Australia), [2]Mediterranean Institute for Advanced Studies (IMEDEA, CSIC-UIB) (Spain), [3]Landcare Research (New Zealand)

The gut microbiome performs various functions (providing nutrients, degrading toxins, shaping the host immune system, *etc*.) and shares its evolutionary history with the host. However, the evolutionary role of the gut microbiome remains further explored, especially in ancient non-human mammals. Here, we used *Myotragus balearicus*, an extinct cave goat that was widely distributed on the Balearic Islands and included toxic plants in its diet, as a model species to investigate the role of the gut microbiome in animal adaptation. We reconstructed the species and functions present within the *M. balearicus* gut microbiome using eight feces remains (coprolites) and reveal microbial symbionts that are very likely critical to the adaptation of *M. balearicus* to the native vegetation. First, we observed that key genes involved in the detoxification of xenobiotic toxins (*e.g., cyt* P450) are enriched in coprolite metagenomes, implying the role of the gut microbiome of *M. balearicus* in the degradation of toxic plants. We also identified that *Romboutsia ilealis*, a probiotic bacterium that can improve the gut function and alleviate the inflammatory responses, is highly abundant in well-preserved coprolites. The draft genome of the ancient *R. ilealis* (80.5% coverage, 12.1X) was obtained and immunomodulatory genes were confirmed. Interestingly, the phylogenetic relationship of the *Romboutsia* species across multiple mammals mirrors that of their hosts, suggesting an extremely deep co-evolutionary relationship. This study demonstrates the strength of this approach to reveal physiological information about extinct species and explore deep co-evolutionary relationships utilized by mammals in adaptive processes.

POA-021

Polarella genomics: understanding cold adaptation and evolutionary transition to symbiosis in dinoflagellates

Timothy Gordon Stephens[1], Debashish Bhattacharya[2], Mark A Ragan[1], Cheong Xin Chan[1, 3]

[1]The University of Queensland (Australia), [2]Rutgers University (United States), [3]The University of Queensland (Australia)

Dinoflagellates are an important, ubiquitous group of phytoplankton that play a key ecological role in a diverse array of marine ecosystems. Some lineages form symbiotic relationships with corals and coral reef animals, while other are free-living, including those that cause harmful algal blooms, or inhabit brine channels in polar sea ice. Dinoflagellate genomes display idiosyncratic features such as immense genome sizes (i.e. up to 70 times the size of the human genome) and non-canonical splice sites, posing a challenge in de novo genomics. Genome data of some *Symbiodinium* species (the coral reef symbionts) were recently published, but little is known about the molecular mechanisms that underpin the evolution of free-living dinoflagellates. We have sequenced and assembled genomes of two isolates of *Polarella glacialis* (from the Arctic and Antarctica), the free-living sister lineage to *Symbiodinium*. With an estimated genome size of 1.5 Gbp, *Polarella* represents an entry point into the genomes of free-living dinoflagellates. Our preliminary assembly of the Antarctic isolate consists of 94,256 scaffolds, with 61,355 predicted genes. In a comprehensive analysis of 47 publicly available dinoflagellate transcriptomes, we found in *Polarella* a host of significantly enriched domains with potential implications in cold adaptation, including the cold shock and ice-binding domains. Our results provide the first comprehensive overview of key molecular processes and functions in free-living dinoflagellates, specifically relevant to our understanding of cold adaptation and the transition to a symbiotic lifestyle in dinoflagellates.

POA-022

Comparative functional genomics of the obligate endosymbiont Buchnera aphidicola

Rebecca A Chong[1], Nancy A Moran[2]

[1]University of Hawaii at Manoa (United States), [2]University of Texas at Austin (United States)

Symbiotic interaction between hosts and microbes promotes the establishment of new ecological niches and has profound impacts on the evolutionary trajectory of eukaryotic lineages. Many insects harbor maternally transmitted endosymbionts that provision hosts with key nutrients. One key evolutionary consequence of vertically transmitted symbiosis is rapid degradation of symbiont genomes. We use the system of aphids and their obligate endosymbionts, *Buchnera aphidicola*, to explore the evolutionary process of symbiont genome degradation and the functional consequences of this process. We compared complete genome sequences for 40 *Buchnera* strains, including newly sequenced symbiont genomes from 25 different aphid species, to reconstruct the ancestral *Buchnera* genome and characterize patterns of gene loss across lineages. We identified over 700 genes present in the ancestral *Buchnera* genome and describe variable regions of gene loss across the phylogeny. Genes related to cell wall and cell division, DNA repair, and electron transport have been lost in several lineages. Genes involved in transcription and translation are retained throughout the phylogeny along with genes related to amino acid biosynthesis, the primary function of *Buchnera*. We also find repeated gains and losses of plasmids harboring genes involved in leucine and tryptophan biosynthesis. By reconstructing the ancestral genome of *Buchnera* and patterns of gene loss across the phylogeny, we characterize the process and functional consequence of rapid genome degradation in this ancient obligate symbiont.

POA-024

Evolutionary genomics of coral reef symbionts

Raul Augusto Gonzalez-Pech[1], Debashish Bhattacharya[2], Mark A Ragan[2], Cheong Xin Chan[1, 3]

[1]The University of Queensland (Australia), [2]Rutgers University (United States), [3]The University of Queensland (Australia)

Symbiodinium are a specialised group of dinoflagellates, many of which form symbiotic associations with diverse host organisms in coral reefs, including cnidarians, molluscs and foraminiferans. However, "free-living" *Symbiodinium* (*i.e.* isolates not associated with any host) have also been described. Free-living *Symbiodinium* form a distinct subclade within the basal lineage, Clade A. While other dinoflagellates are predominantly free-living, the evolutionary mechanisms that have led to the establishment of symbiosis between *Symbiodinium* and other organisms remain little known. To address these issues, we sequenced and generated draft genomes from eight *Symbiodinium* isolates within Clade A, including symbionts associated with different hosts and for the first time, the free-living *Symbiodinium natans* CCMP2548 (HA3-5). In contrast with the better-understood transition of prokaryotes into endosymbionts or parasites, this study explores the evolution of eukaryotic symbionts. Here, I will present our recent findings based on comparative analyses between *S. natans* and other *Symbiodinium* genomes, and highlight distinct genome signatures that are relevant to the evolutionary transition of *Symbiodinium* from a free-living to a symbiotic lifestyle. I will also highlight the genomic differences between eukaryotes and prokaryotes resulting from their evolutionary transition into endosymbionts. In addition, I will emphasize the remarkable genome sequence divergence among isolates within Clade A, underscoring challenges in *Symbiodinium* genomics. Our data and findings provide a platform for future investigations in *Symbiodinium* biology and evolution.

POA-025

Evolution of exon-intron boundary recognition in coral symbiotic algae

Shinichiro Maruyama[1], Yuu Ishii[1], Konomi Fujimura-Kamada[2], Natsumaro Kutsuna[3, 4], Shunichi Takahashi[2, 5], Takashi Makino[1], Jun Minagawa[2, 5], Masakado Kawata[1]

[1]Tohoku University (Japan), [2]National Institute for Basic Biology (Japan), [3]University of Tokyo (Japan), [4]Inc. (Japan), [5]The Graduate University for Advanced Studies (Japan)

Symbiodinium spp. are dinoflagellate algae known for their abilities to establish symbiotic relationships with many marine invertebrates including reef building corals, as well as their unique genome architectures and gene expression systems. Although recent genomic studies suggested that exon-intron boundary seuqences were non-canonical and that the intronic acceptor site sequence ('AG') and its flanking exonic guanine residue, collectively 'AG"G"', were highly conserved in a coral symbiont *S. minutum*, the molecular mechanisms and evolutionary consequences relevant to the exon-intron recognition remain to be clarified. We identified a mutant strain of *Symbiodinium* sp. showing a single nucleotide substitution, from 'GA-AG' to 'GA-GG', at the exon-intron boundary, resulting in a splicing variation in a pyrimidine biosynthesis gene. Sequence analyses of the cDNAs suggested that the splicing machinery recognized not the downstream 'AG"A"', proximal to the original acceptor site, but rather distant 'AG"G"' as a new acceptor site. Using the genome sequences of three *Symbiodinium* species from different clades, we found that amino acid compositions in the codons located around the exon-intron boundaries were non-homogenous in all the genomes analyzed. Interspecies comparison also showed that base compositions of the exon-intron boundaries were biased in a group of conserved introns, of which the positions were conserved among the orthologous genes in *Symbiodinium* species, relative to all the introns in the genomes. These data suggest that coral symbiont algae possess unique exon-intron boundary recognition mechanisms and be a useful model for studying how introns could affect the evolution of coding sequences in eukaryotic genomes.

POA-026

The evolutionary footprint of lichenization - Towards the characterization of a eukaryotic pioneering holo-organism

Bastian Greshake Tzovaras[1], Arpit Jain[1], Ingo Ebersberger[1, 2, 3]

[1]Inst. of Cell Biology and Neuroscience, Goethe University Frankfurt (Germany), [2]Senckenberg Museum, Frankfurt (Germany), [3]Goethe University, Frankfurt (Germany)

Around 21% of all fungi live in lichen symbioses. These communities are so successful that they often pioneer the colonization of habitats that are otherwise too extreme to support eukaryotic life. The symbiotic interactions of the fungal and the photosynthesizing partners in lichens have been described as ranging from parasitic to mutualistic, depending on the species. However, to what end the symbionts' metabolisms are integrated is largely unknown. We approach this issue by exploring the extent of genomic remodelling as a consequence of the symbiosis in lichenized Lecanoromycetes. We begin with characterizing the holo-genome of the rock-dwelling *Lasallia pustulata*, the first lichen sequenced in a metagenomics approach. Evaluating the quality of gene annotations across five Lecanoromycetes reveals, for all genomes, a considerable fraction of un- or mis-annotated genes, as well as of genes that are overlooked in the downstream ortholog searches. Rigorous quality standards at all steps in the comparative genomics analysis of symbiotic communities are, thus, essential to differentiate between artefacts and genuine lineage-specific losses of evolutionarily old genes and functions. Our analyses reveal a pronounced gene loss in the last common ancestor of the Lecanoromycetes, coinciding with the onset of lichenization. Most prominently, this involves genes of the polysaccharide metabolism indicating that this loss-of-function is complemented by the algal photobiont. So far, we did not find pronounced adaptations in the algal metabolism. Yet, stress response pathways, such as the heat-stress response system, seems to not differ from aquatic green algae, suggesting a stress-protective role of the fungus.

POA-027

Eukaryote genes are more likely than prokaryote genes to be composite

Yaqing Ou[1], James McInerney[1, 2]

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

Species evolution is diverse, not only from the processual process but also from the introgressive process. In the latter, non-homologous recombination can result in new genes being formed by combining parts of existing genes, and likewise, it can result in breaking genes up; these genes are termed composite genes. We set out to examine the extent to which genomes from cells, viruses, and plasmids contain composite genes. We identify composite genes when a given gene shows partial homology to at least two unrelated genes. In order to further analyze composite and component genes, we abstracted our genomic data into graphical form. We constructed sequence similarity networks (SSNs) from 1,190,265 genes comprising the genomes of 36 eukaryotes, 56 archaea, 90 bacteria, 79 viruses and 1,614 plasmids. We then identified non-transitive triplets of nodes in this network and explored the homology relationships in these triplets to see if the middle nodes were indeed composite genes. We identified 221,043 genes (18.57%) as being composites of at least two other genes or partial genes. Composite genes were found to be distributed across all kinds of genomes and across all functional COG categories. Interestingly, the presence of composite genes is statistically significantly more likely in eukaryotes rather than prokaryotes.

POA-029

Identification of the Paramecium bursaria genes involved in endosymbiosis with Chlorella spp.

Jun-Yi Leu[1, 2], Yu-Hsuan Cheng[1, 2], Yen-Hsin Yu[1], Chien-Fu Jeff Liu[1], Trees-Juen Chuang[3, 2], Isheng Jason Tsai[4, 2]

[1]Academia Sinica (Taiwan), [2]National Taiwan University and Academia Sinica (Taiwan), [3]Academia Sinica (Taiwan), [4]Academia Sinica (Taiwan)

Endosymbiosis is one of the major forces driving the evolution of eukaryotic cells, which has occurred multiple times in different lineages during the evolutionary history. However, the initial process of how the endosymbiosis is formed and established in the host cell is still unclear. Here, we use the ciliate Paramecium bursaria and the algae Chlorella spp. to study the early stage of endosymbiosis. The P. bursaria can establish endosymbiosis with various Chlorella species. Previous studies have shown that there are some key cytosolic events that establish endosymbiosis. However, the molecular mechanisms or genes responsible for initiating and establishing the relationship between P. bursaria and Chlorella spp. remain largely unknown. We performed infection experiments of two P. bursaria aposymbiotic strains with different Chlorella species and found that these two strains exhibited different abilities of establishing stable endosymbiosis. The genomes and transcriptomes of these two P. bursaria strains were analyzed in order to identify candidate genes contributing this difference. Our results have shed light on the genetic basis of endosymbiosis establishment. We also investigated the possible benefits of this symbiotic relationship and found that ciliates with symbionts were more resistant to a pathogenic bacterium than the aposymbiotic ones. Currently we are dissecting the underlying mechanisms of this interesting phenomenon.

POA-030

High-resolution metagenomics uncovers microbiota acquisition dynamics in the mussel Bathymodiolus brooksi

Devani Romero-Picazo[1], Tal Dagan[1], Nicole Dubilier[2], Rebecca Ansorge[2], Anne Kupczok[1]

[1]Christian-Albrechts Kiel University (Germany), [2]Max Planck Institute for Marine Microbiology (Germany)

Bathymodiolus mussels inhabit hydrothermal vents and cold seeps worldwide. They harbor thiotrophic (SOX) and methanotrophic (MOX) symbiotic bacteria in their gill tissue that provide them with nutrition. The Bathymodiolus microbiota is thought to be acquired horizontally, however, little is known about its establishment in mussel individuals. Here, we use high-resolution metagenomics of 23 samples including young and old mussels to characterise intra-species diversity changes over the mussels lifetime for both endosybionts. We constructed a non-redundant gene catalog of 4.4 million genes and implemented a gene-based binning pipeline to resolve species bins by means of co-abundance gene segregation. Among three bins with at least 700 genes, two bins have been identified as nearly complete MOX (2,518 genes) and SOX (1,439 genes) core-genomes. Our results reveal different community composition according to age with higher relative abundance of MOX in young mussels. The single-nucleotide polymorphisms (SNPs) density is 1.86 SNPs/kbp and 10.56 SNPs/kbp in the MOX and SOX core-genomes, respectively. A comparison of the genetic diversity among mussels shows that symbiont populations of young mussels have higher intra-host and pairwise inter-host nucleotide diversity (π) than old mussels symbiont populations. Fixation index estimation (F_{ST}) reveals two SOX symbiont subpopulations among old mussels. Taken together, maximum intra-host symbiont diversity is reached early in hosts life. This suggests that the acquisition of symbionts is restricted to an early age of the mussel and that symbiont communities in old mussels result from the reduction of the diversity in young mussels.

POA-032

Rates of Gut Microbiome Divergence in Mammals

Alex Nishida[1], Howard Ochman[1]

[1]University of Texas at Austin (United States)

The variation and taxonomic diversity among mammalian gut microbiomes raises several questions about the factors that contribute to the rates and patterns of change in these microbial communities. By comparing the microbiome compositions of 112 species representing 14 mammalian orders, we assessed how host and ecological factors contribute to microbiome diversification. Except in rare cases, the same bacterial phyla predominate in mammalian gut microbiomes, and there has been some convergence of microbiome compositions according to dietary category across all mammalians lineages except Chiropterans (bats), which possess high proportions of Proteobacteria and tend to be most similar to one another regardless of diet. At lower taxonomic ranks (families, genera, 97% OTUs), bacteria are more likely to be associated with a particular mammalian lineage than with a particular dietary category, resulting in a strong phylogenetic signal in the degree to which microbiomes diverge. Despite different physiologies, the gut microbiomes of several mammalian lineages have diverged at roughly the same rate over the past 75 million years; however, the gut microbiomes of Cetartiodactyla (ruminants, whales, hippopotami) have evolved much faster and those of Chiropterans much slower. Contrary to expectations, the number of dietary transitions within a lineage does not influence rates of microbiome divergence, but instead, some of the most dramatic changes are associated with the loss of bacterial taxa, such as those accompanying the transition from terrestrial to marine lifestyles and the evolution of hominids.

POA-033

Diversity and horizontal gene transfer of nodule bacteria associated with *Lotus japonicus* in natural environments

Masaru BAMBA[1], Seishiro AOKI[2], Tadashi KAJITA[3], Yasuyuki WATANO[4], Hiroaki Setoguchi[5], Syusei SATO[6], Takashi Tsuchimatsu[4]

[1]Chiba University (Japan), [2]The University of Tokyo (Japan), [3]University of the Ryukyus (Japan), [4]Chiba University (Japan), [5]Kyoto University (Japan), [6]Tohoku University (Japan)

Horizontal transfer of genomic islands can provide complex novel traits involved in virulence and mutualism, which enable bacteria to interact with eukaryotic hosts. In the legume-rhizobia symbiosis, the symbiont bacteria have evolved via horizontal transfer of nodulation and nitrogen fixation genes, which are carried by either large plasmids or genomic islands.

Lotus japonicus is an emerging model for investigating the genomics of legume species, and has contributed to the understanding of symbiotic nitrogen fixation. The symbionts of *Lotus* species have been used for the investigation of the horizontal transfer of symbiotic genes in the laboratory and field experiments. However, the level of diversity of associated nodule bacteria and the frequency of horizontal gene transfer are still unclear in natural environments.

Here, we isolated 106 strains that associated with *L. japonicus* collected from 14 sites in Japan, and sequenced three housekeeping (*recA*, *atpD*, and *dnaK*) and four symbiotic (*nodA*, *nodB*, *nodC* and *nifH*) genes. We found that *L. japonicus*-associated strains belonged to diverse lineages of the genus *Mesorhizobium*; however, all the strains possessed strikingly similar symbiotic gene sequences. These results suggested that horizontal transfer of symbiotic genes enabled diverse *Mesorhizobia* to establish symbiotic relationships with *L. japonicus*.

POA-034

Gut metagenomes reveal the evolution of lignocellulolytic abilities across termites

Lucia Zifcakova[1], Thomas Bourguignon[1]

[1]Okinawa Institute of Science and Technology (Japan)

Termites are the dominant decomposers in tropical and subtropical ecosystems. Estimations suggest that termites recycle up to 30% of the total carbon biomass on earth, largely thanks their unique association with complex gut microbial communities that help them to degrade lignocellulose.

In this study, we sequenced gut metagenomes of the cockroach Cryptocercus and seven termite species, representative of the termite phylogenetic diversity, and investigated their abilities to degrade lignocellulose. Using this dataset, we tested the hypothesis that higher termites reached ecological dominance for their enhanced ability to degrade lignocellulose.

Our results show that protozoa only constitute a small fraction of reads in the metagenomes (0.25-0.015%). In contrast, bacteria represent 91-97% of the reads identified by the MG-RAST pipeline. All examined species harbour the same bacterial taxa, but their cellulolytic potential vary substantially between species. The lowest amount of genes involved in polysaccharide degradation was found in the soil-feeding termite Indotermes sp, while the highest amount was found in wood roach Cryptocercus punctulatus and wood feeding higher termite Microcerotermes sp. On the other hand, Mastotermes darwiniensis, which is considered to be the most basal termite group, encoded significantly more genes for endocellulase enzyme involved in cellulose degradation than other termites.

We found no trend of increasing abundance of lignocellulolytic genes in higher termites compared to lower termites. However, lignocellulolytic genes were scarce in the soil-feeding termite Indotermes sp, suggesting that diet, not phylogeny, determines termite gut lignocellulolytic abilities.

POA-035

Large scale comparative genomics reveals the path to genome reduction in the cockroach endosymbiont, Blattabacterium cuenoti.

Yukihiro Kinjo[1], Gaku Tokuda[2], Nathan Lo[3], Thomas Bourguignon[1, 4]

[1]Okinawa Institute of Science and Technology Graduate University (Japan), [2]University of the Ryukyus (Japan), [3]University of Sydney (Australia), [4]Czech University of Life Sciences (Czech Republic)

Most cockroaches harbor an obligate intercellular symbiotic bacteria (endosymbiont), called *Blattabacterium cuenoti* (hereafter *Blattabacterium*), which recycles host nitrogen wastes and provides amino acids and vitamins. This close association was established over 220 million years ago, and has been maintained in all cockroaches but the Euisoptera. The eight *Blattabacterium* genomes sequenced to date show that gene content has been very stable over time, although the strains associated with *Cryptocercus* and *Mastotermes* lost about 10% of their original gene content. One drawback to our current understanding of *Blattabacterium* genome evolution is the low number of available genomes, with limited taxonomic coverage. To address this issue, we sequenced the genomes of 58 new *Blattabacterium* strains and carried out comparative genomic analysis. Sequenced genomes were representative of the cockroach diversity. We found that most strains of *Blattabacterium* have similar genome size (630kbp), although several strains lost up to 10% of their genes. The functional profiles of the lost genes varied among host lineages. While some lineages lost amino acid biosynthetic genes, other predominantly lost cofactor biosynthetic genes. In this talk, we will present the many ways *Blattabacterium* genomes have been eroded over their 220 million years of coevolution with cockroaches.

POA-037

Strong genomewide selection on protein coding sequences of bacterial endosymbionts and obligate pathogens

Saurabh Mahajan[1], Deepa Agashe[1]

[1]National Centre for Biological Sciences (India)

Bacterial endosymbionts and obligate pathogens are thought to have reduced effective population sizes (Ne) due to their obligate host-dependent life cycle with frequent bottlenecks. Reduced Ne manifests in low synonymous site diversity, accelerated evolutionary rates, reduced genome sizes, and reduced GC content. Reduced Ne is also expected to relax selection on protein coding sequences. Indeed, many studies have found increased ratios of non-synonymous substitution rate to synonymous substitution rate (dN/dS) in endosymbionts. However, almost all previous studies estimated dN/dS using methods that are systematically biased by low GC content.. We estimated dN/dS in two lineages of bacterial endosymbionts of insects (Buchnera and Blochmannia) and one obligate animal pathogen (Bartonella) using a recently published method based that accounts for the concomitant reduction in GC content of host-associated lineages. Unexpectedly, and in contrast to earlier studies, we find similar or decreased dN/dS ratio (genome-wide and gene specific) in these organisms compared to related free-living bacteria. Further, we find that the rate of synonymous, but not non-synonymous substitutions is increased in the host-associated lineages. Thus, despite the estimated 100-fold reduction in Ne and efficiency of selection, we observe similar or greater purifying selection on protein coding sequences. Our results imply that the decreased efficiency of selection due to reduced effective population size is nullified by increased selection at non-synonymous sites. We speculate that this results from an increased usage of genes that have survived the genome erosion and are essential for the host-associated lifestyle.

POA-038

Game of Introns: in search of plastid-derived HGTs in plant and algal nuclear genomes

Vera Mukhina[1, 2], Mikhail Gelfand[1, 3, 4]

[1]Institute for Information Transmission Problems (Russian Federation), [2]Vavilov Institute of General Genetics (Russian Federation), [3]Skolkovo Institute of Science and Technology (Russian Federation), [4]National Research University Higher School of Economics (Russian Federation)

Plastids originated from cyanobacteria, engulfed once by the common ancestor of three clades, cyanophytes, red algae, and green algae. Then they spread in other single-cell eukaryote clades by multiple secondary endosymbioses. Here we attempt to trace varying fates of cyanobacterial genes during coevolution: retention in plastid, elimination, or horizontal transfer into the host nuclear genome with subsequent intron invasion. We also investigated how these genes transfer to the secondary symbiotic genomes - do they jump directly from the plastid to the nucleus of the final host or move via the nucleus of the intermediate symbiont. Gene screening across available clades of primary and secondary symbionts allowed us to estimate the gene flows from the plastids to the nuclear genomes and to detect multiple imprints of sequential gene transfer into the nuclear genomes of secondary symbionts. Our data demonstrate that most functional genes have been incorporated into the host nuclear genomes at the early stages of primary symbiosis and carry many introns acquired before the separation of red and green algae.

POA-039

Rapid evolution of distinct Helicobacter pylori subpopulations in the Americas

Koji Yahara[1], Kaisa Thorell[2], Elvire Berthenet[3], Daniel Lawson[4], Jane Mikhail[3], Ikuko Kato[5], Alfonso Mendez[6], Cosmeri Rizzato[7], Maria Bravo[8], Rumiko Suzuki[9], Yoshio Yamaoka[9], Javier Torres[10], Samuel Sheppard[11], Daniel Falush[11]

[1]National Institute of Infectious Diseases (Japan), [2]Karolinska Institutet (Sweden), [3]Swansea University (United Kingdom), [4]University of Bristol (United Kingdom), [5]Wayne State University (United States), [6]ENCB (Mexico), [7]Universita di Pisa (Italy), [8]Instituto Nacional de Cancerologi (Colombia), [9]Oita University (Japan), [10]IMSS (Mexico), [11]University of Bath (United Kingdom)

For the last 500 years, the Americas have been a melting pot both for genetically diverse humans and for the pathogenic and commensal organisms associated with them. One such organism is the stomach dwelling bacterium *Helicobacter pylori*, which is highly prevalent in Latin America where it is a major current public health challenge because of its strong association with gastric cancer. By analyzing the genome sequence of *H. pylori* isolated in North, Central and South America, we found evidence for admixture between *H. pylori* of European and African origin throughout the Americas, without substantial input from pre-Columbian (hspAmerind) bacteria. In the US, strains of African and European origin have remained genetically distinct, while in Colombia and Nicaragua, bottlenecks and rampant genetic exchange amongst isolates have led to the formation of national gene pools. We found three outer membrane proteins with atypical levels of Asian ancestry in American strains, as well as alleles that were strongly fixed specifically in South American isolates, suggesting a role for the ethnic makeup of hosts in the colonization of incoming strains. Our results show that new *H. pylori* subpopulations can rapidly arise, spread and adapt during times of demographic flux, and suggest that differences in transmission ecology between high and low prevalence areas may substantially affect the composition of bacterial populations (PLoS Genetics, 2017).

POA-041

Recombination signal in Mycobacterium tuberculosis stems from reference-guided assemblies and alignment artefacts

Maxime Godfroid[1], Tal Dagan[1], Anne Kupczok[1]

[1]Kiel University (Germany)

DNA acquisition via recombination is considered advantageous as it has the potential to bring together beneficial mutations that emerge independently within a population. Furthermore, recombination is considered to contribute to the maintenance of genome stability by purging slightly deleterious mutations. The prevalence of recombination differs among prokaryotic species and depends on the accessibility of DNA transfer mechanisms. An exceptional example is the human pathogen *Mycobacterium tuberculosis* (MTB) where no clear transfer mechanisms have been so far characterized and the presence of recombination is questioned. Here we analyse completely assembled MTB genomes in search for the footprints of recombination. We find that putative recombination events are enriched in strains reconstructed by reference-guided assembly and in regions with unreliable alignments. In addition, assembly and alignment artefacts introduce phylogenetic signals that are conflicting the established MTB phylogeny. Our results reveal that the so far reported recombination events in MTB are likely to stem from methodological artefacts. We conclude that no reliable signal of recombination is observed in the currently available MTB genomes. Moreover, our study demonstrates the limitations of reference-guided genome assembly for phylogenetic reconstructions. Rigorously de novo assembled genomes of high quality are mandatory in order to distinguish true evolutionary signal from noise, in particular for low diversity species such as MTB.

POA-042

Quantifying population structure of malaria parasites using epidemiological and genomic data

Hsiao-Han Chang[1], Amy Wesolowski[2], Ipsita Sinha[3, 4], Md Amir Hossain[5], M Abul Faiz[6], Olivo Miotto[3], Dominic Kwiatkowski[7], Richard Maude[3, 4], Caroline Buckee[1]

[1]Harvard T.H. Chan School of Public Health (United States), [2]Johns Hopkins Bloomberg School of Public Health (United States), [3]Mahidol University (Thailand), [4]University of Oxford (United Kingdom), [5]Chittagong Medical College Hospital (Bangladesh), [6]Dev Care Foundation (Bangladesh), [7]Wellcome Trust Sanger Institute (United Kingdom)

Malaria is one of the leading causes of disease and death worldwide. Regular circulation of individuals to and from malaria endemic areas undermines local control by reintroducing infections and sustaining local transmission. Quantifying the movement of malaria parasites has become a priority for national control programs, but remains methodologically challenging due to the unique evolutionary and epidemiological features of malaria parasite transmission. In particular, high rates of superinfection with multiple strains and a predominance of asymptomatic infections contribute to difficulty in identifying transmission chains and routes of parasite importation. Here, we assessed the utility of genetic data in combination with epidemiological modeling and detailed travel surveys to measure the spread of malaria parasites in Bangladesh. We collected genetic barcodes of 101 SNPs and epidemiological data from 1,412 patients residing in 143 separate unions in Bangladesh. We found that, at this geographic scale, the proportion of parasites with nearly identical barcodes was highly associated with geographic distance, while standard genetic methods, such as average pairwise difference or F_{ST}, were not. We developed a genetic mixing index that quantifies the likelihood of samples from one location having higher-than-expected relatedness to samples from distant locations. We then integrated genetic and epidemiological data, including case time, location, and travel history, in the same model to infer parasite flows between locations. Our results show distinct regional mixing in the north and south of the malaria-endemic region in Bangladesh, and that, contrary to dogma, most of the parasite mixing did not involve forest areas.

POA-044

Reconstruction of bacterial cell division history to identify new potential antibiotic targets

Pierre Simon Garcia[1, 2], Christophe Grangeasse[2], Celine Brochier-Armanet[1]

[1]Universite Claude Bernard Lyon 1 (France), [2]Institut de Biologie et Chimie des Proteines (France)

The number of death from bacterial infections is estimated to be higher than cancer death cases in 2050. There is thus an urgent need to develop resistance-escaping antibiotics and to identify new therapeutic targets. Cell division and cell cycle proteins are promising candidates although antibiotics targeting these cellular processes are yet to be used in clinic. While they have been deeply studied in a few model bacteria, our current knowledge does not extend to the whole bacteria domain. To fill this gap, we initiated a large scale phylogenomic study to decipher the evolutionary history of the cell division and cell cycle apparatus in Bacteria. Here we present a detailed analysis of the cell division and cell cycle apparatus in *Firmicutes*, one of the main bacterial phyla that encompasses many major pathogens such as *Clostridium difficile*, *Staphylococcus aureus* and *Streptococcus pneumoniae*. Using phylogenetic and reconciliation approaches, genomic contexts and protein domain composition analyses, we reconstructed the evolutive history of genes involved in cell division and cell cycle and highlighted evolutive hotspots associated to massive gene losses and gains at the origin of some clades, e.g. *Bacilli* and *Streptococcaceae*. Our data allow us to disclose functional links among known proteins and identify candidate proteins likely involved in these processes that represent interesting targets to antibacterial drugs.

POA-045

Historical *Y. pestis* genomes provide insights into the initiation and progression of the second plague pandemic

Maria A. Spyrou[1], Marcel Keller[1], Rezeda I. Tukhbatova[1, 2], Elisabeth Nelson[1], Don Walker[3], Sacha Kacki[4], Dominique Castex[5], Sandra Loesch[6], Michaela Harbeck [7], Alexander Herbig[1], Kirsten I Bos[1], Johannes Krause[1]

[1]Max Planck Institute for the Science of Human History (Germany), [2]Kazan Federal University (Russian Federation), [3]Museum of London Archaeology (United Kingdom), [4]Durham University (France), [5]University of Bordeaux (United Kingdom), [6]University of Bern (Switzerland), [7]Ludwig Maximilian University of Munich (Germany)

The second plague pandemic, caused by the bacterium *Yersinia pestis*, is infamous for its initial wave, the Black Death (1347-1352 AD). The pandemic lasted until the 18th century, with more than 6,000 outbreaks documented in Europe during this period. As an important historical example for an emergent disease, the origin of the Black Death and its relationship to the ensuing outbreaks are topics of active research. Although published genomic studies of both ancient and modern diversity point to an East Asian origin, there is considerable lack of historical data to support this scenario. In addition, there is disagreement between the evidence available from genomic and climatic data regarding the source of post-Black Death European outbreaks, with the former suggesting the persistence of a local reservoir in Europe and the latter suggesting recurrent introductions of the bacterium from Central Asia. Here, we nearly triple the number of *Y. pestis* genomes sequenced from the second plague pandemic, by analysing human remains from nine sites located around Europe spanning the 14th-17th centuries. Our data supports an Eastern European entry of the disease during the Black Death, and shows low genetic diversity in the bacterium during the initial wave. In addition, our analysis of post-Black Death outbreaks suggests the local diversification of a single plague lineage that may have given rise to more than one reservoirs of the disease in Europe. Our study provides a comprehensive genetic analysis on the progression of one of the most devastating pandemics in human history.

POA-046

Evolution pathway of the antimicrobial resistance genes

Marcus Shum[1], Tommy Lam[1]

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

Antimicrobial resistance(AMR) is the ability of bacteria to resist antibiotics. It is mainly conferred by the antimicrobial resistance genes (ARGs) present in the genomes or plasmids of the bacteria. The presence of ARG in the environment present the threat for the emergence of AMR carrying bacteria. In fact, origins of many ARGs remain unclear because most studies focused on the clinical isolates where these ARG-acquired resistant strains already emerged, and less were on the environmental samples.

The *mcr-1* plasmid gene was first found in pigs in China in 2016. Since then, it was found carried by animals or human globally. *mcr-2*, *mcr-3* and other *mcr* genes were then discovered. These *mcr* genes confer to the resistance to colistin that is considered as the last-resort antibiotic for treating multidrug-resistant bacteria infections. Spreading of the *mcr* genes poses great threat in human and animal health. We would like to identify the origin of these genes as to control the spreading at source.

We targeted several ARG families including *mcr*, *tet*, *met*, *qnr*, and aimed to study their evolutionary pathways from nature. We searched and identified these ARGs from public genomic and metagenomic databases, and then constructed phylogenetic trees of these sequences. The geographical location and sample collection time where these ARGs were identified were analyzed in the phylogenetic context to track their evolution and dissemination in different sectors. The results of this study shed lights into the origins of these ARGs, providing clues for controlling ARGs at source.

POA-047

Reconstruction of the Killer Whale Oral Microbiome

Courtney A Hofman[1, 2], Rita Austin[1, 2], Michael A. Etnier[4], Krithivasan Sankaranarayanan[1, 3]

[1]University of Oklahoma (United States), [2]University of Oklahoma (United States), [3]University of Oklahoma (United States), [4]Western Washington University (United States)

Biomolecules in human dental calculus have been used to investigate life history, including dietary reconstruction and pathogen evolution, however its broader utility to study other animals has been underexplored. Here we use historic dental calculus samples obtained from marine mammal strandings and research collections to assess the preservation of an oral microbiome signature and investigate its potential for pathogen reconstruction. We conducted metagenomic screening of fourteen killer whales (*Orcinus orca*) from the North Pacific. These samples were collected between 1961 and 2015 and represent two different killer whale ecotypes. DNA extraction was performed following protocols optimized for human dental calculus and yielded between 0.73 and 121.3 ng/mg of DNA. Despite sample ages ranging from 56 years old to 2 years old, the average fragment lengths were short (91+-17 basepairs). Fragment length is correlated (ρ =0.52) with age as expected with degraded DNA. Preliminary analysis shows the presence of several taxa known to inhabit the oral cavity including *Methanobrevibacter, Neisseria, Jeotgalicoccus, Corynebacterium,* and *Mogibacterium*. The offshore killer whale ecotype is routinely associated with worn and abscessed teeth. We recovered *Staphylococcus* reads from one of our offshore specimens and members of this genus are known to cause dental abscesses in humans. These data demonstrate the feasibility of reconstructing the oral microbiome from non-primate dental calculus and open the possibility for investigating oral health in response to environmental change.

POA-049

The Impact of Acquired Copper Resistance Loci on Epidemic Methicillin Resistant Staphylococcus aureus Pathogenesis and Spread

Paul J Planet[1, 2, 3], Ahmed M Moustafa[2], Chanelle Ryan[2], Alejandra Londono[4], Cesar Arias[5], Jeffrey Boyd[7], David Heinrichs[6]

[1]University of Pennsylvania (United States), [2]Children's Hospital of Pennsylvania (United States), [3]American Museum of Natural History (United States), [4]Columbia University (United States), [5]University of Texas, Houston (United States), [6]University of Western Ontario, London (Canada), [7]Rutgers University (United States)

Methicillin Resistant Staphylococcus aureus (MRSA) is a major cause of both community and hospital acquired illness, and the its epidemiological spread has been associated with a small number of clonal lineages in different geographic regions. The mechanisms that promote the widespread geographic dissemination and increased virulence of MRSA are not well understood. Here we examine the possible role of horizontal acquisition copBL locus, which appears to have been independently acquired through horizontal gene transfer in several epidemic lineages immediately prior to their geographical spread. We present an overview of copBL evolution in the S. aureus clade and then focus on the independent acquisition of copB by epidemic strains in South and North America. We show that epidemic MRSA strains present hyper tolerance compared to close relatives that lack the copB locus, nonpolar mutants with deleted copB or copL are more susceptible to copper challenge in both in media and in water. In addition these mutants are attenuated for survival within macrophages, which are known to recruit copper for phagolysosomal killing. Finally, we show that copL and copB mutants are also attenuated in murine models of skin, bone and lung infection. These results suggest that acquisition of the copBL locus could have conferred both an environmental advantage to S. aureus, and a host immune evasion strategy that promoted better spread and more aggressive disease.

POA-052

First *in vivo* evidence of the functionality of a sncRNA of mitochondrial origin (smithRNA) targeting a nuclear transcript and affecting histone methylation

Federico Plazzi[1], Manuel Delpero[1], Andrea Pozzi[1], Marco Passamonti[1]

[1]University of Bologna (Italy)

Small MITochondrial Highly-transcribed RNAs (smithRNAs) were recently described in the species *Ruditapes philippinarum* (Mollusca: Bivalvia). At variance to any other known sncRNA, they are coded by the mitochondrial genome (mtDNA) and they are predicted *in silico* to target nuclear transcripts. This possibility was never suggested before, and provides an unprecedented way for mtDNA to affect nuclear gene expression. Here we report the first *in vivo* evidence of functionality of one of the above predicted smithRNAs, named 'M_smithRNA106t'. The only detected target of this smithRNA species was the homolog of the Histone-lysine N-methyltransferase SETD2, the enzyme connected with the H3K36 tri-methylation in humans. This smithRNA is of particular interest because its final effect would be genome-wide, as it should affect the global histone methylation level. We demonstrated a significant decrease of the H3K36 tri-methylation levels in clams injected with M_smithRNA106t, with respect to the controls. Sex-specific differences in the functionality were also detected. Therefore, this is the first *in vivo* evidence of mtDNA affecting nuclear gene expression through RNA interference. This suggests that the animal mitochondrial genome is much more complex than previously thought, and its functionality should not be only related to coding some OXPHOS subunits.

POA-053

How and why new genes die- de novo microRNAs in the Red Queen race

Guang-An Lu[1]

[1]Sun Yat-sen University, (China)

The prevalence of de novo coding genes is controversial due to the length and coding constraints. Non-coding genes, especially small ones, are freer to evolve de novo by comparison. The best examples are microRNAs (miRNAs), a large class of regulatory molecules ~22 nt in length. Here, we study 6 de novo miRNAs in Drosophila which, like most new genes, are testis-specific. We ask how and why de novo genes die because gene death must be sufficiently frequent to balance the many new births. By knocking out each miRNA gene, we could analyze their contributions to each of the 9 components of male fitness (sperm production, length, competitiveness etc.). To our surprise, the knockout mutants often perform better in some components, and slightly worse in others, than the wildtype. When two of the younger miRNAs are assayed in long-term laboratory populations, their total fitness contributions are found to be essentially zero. These results collectively suggest that adaptive de novo genes die regularly, not due to the loss of functionality, but due to the canceling-out of positive and negative fitness effects, which may be characterized as **quasi-neutrality**. Since de novo genes often emerge adaptively and become lost later, they reveal ongoing period-specific adaptations, reminiscent of the **Red-Queen** metaphor for long term evolution.

POA-054

The impact of translation and RNA-protein interactions in the observed conservation patterns of long non-coding RNAs.

Jorge Ruiz-Orera[1], M.Mar Alba[1, 2]

[1]Hospital del Mar Research Institute, Universitat Pompeu Fabra (Spain), [2]ICREA (Spain)

The advent of high-throughput genomic technologies has revealed that mammalian transcriptomes contain thousands of expressed loci that do not encode conserved long proteins, known as long non-coding RNAs (lncRNAs). Although the majority of lncRNAs have no known functions, a subset of them are known to play roles in gene regulation. The latter often contain short evolutionary conserved sequence segments that are required for their function. Regions that are homologous between human and mouse may correspond to conserved interaction surfaces, but also to unannotated small functional proteins. Here we examined the patterns of ribosome profiling to identify putative translated sequences and other protein-RNA interactions, in conserved and non-conserved lncRNA regions.

We analyzed 9,734 annotated mouse lncRNAs and identified 2,308 regions with significant sequence similarity to human transcripts. A large fraction of them showed promoter overlap, consistent with the conserved expression of the transcript in the two species. Analysis of a high coverage ribosome profiling experiment from mouse hippocampus showed that lncRNA conserved regions contain a higher density of ribosome profiling reads than non-conserved regions. The analysis of read sequence features, such as three-nucleotide periodicity and read length, indicate that conserved regions in lncRNAs are enriched in translated open reading frames (ORFs) as well as in other types of protein-RNA interactions. 8 translated ORFs contain significant selection signatures at the protein level, probably corresponding to functional proteins. Overall, up to 76% of the conserved sequence in lncRNAs can be associated with promoters, antisense gene overlaps and ribosome profiling signatures.

POA-055

Long non-coding RNA ortholog reconstruction from splice sites

Katja Nowick[2, 3, 4], Anne Nitsche[1], Maria Beatriz Walter Costa[1, 3, 4]

[1]University of Leipzig (Germany), [2]Free University Berlin (Germany), [3]University of Leipzig (Germany), [4]University of Leipzig (Germany)

Long non-coding RNAs have become a target of many studies due to their important role in regulation and diseases. LncRNAs are fast evolving and many are primate specific. Although their primary sequence is
poorly conserved, their splice sites are well conserved, enabling better orthology annotation than methods based on sequence alignment. We developed an algorithm, "buildOrthologs", that retrieves the entire ortholog transcripts from other species, using one species as reference. This greedy algorithm recreates the longest valid transcript from a list of splice sites, previously generated by a tool that detects splice site orthology. The validity of the transcript is based on the correct correspondence and order of Donor and Aceptor sites. Global starts and ends are also accounted for. For evolutionary investigation, we applied "buildOrthologs" to the human Gencode lncRNA catalog v26 to retrieve orthologs of four other primates. This allowed us to analyze conservation on the isoform level. To analyze differences in local structure, we developed the "SSS-test" (Selection on the Secondary Structure test, submitted for publication). This statistical test evaluates wheather there is an excess of substitutions leading to structural changes in one lineage. With this test we suggest for future validation a set of human lncRNA candidates with local structures putatively under positive structural selection. These candidates may have important roles in human-specific traits, when compared to closely related species. Our platform consisting of "buildOrthologs" and the "SSS-test" can be applied to create catalogs of lncRNA orthologs for any species and to test for selection.

POA-056

Caste-specific microRNA expression in termites: insights into social evolution

Masatoshi Matsunami[1, 5], Masafumi Nozawa[2], Yudai Masuoka[3, 6], Ryutaro Suzuki[3], Kouhei Toga[3, 7], Katsushi Yamaguchi[4], Kiyoto Maekawa[3], Shuji Shigenobu[4], Toru Muira[1, 8]

[1]Hokkaido University (Japan), [2]Tokyo Metropolitan University (Japan), [3]University of Toyama (Japan), [4]National Institute for Basic Biology (Japan), [5]University of the Ryukyus (Japan), [6]National Agriculture and Food Research Organization (Japan), [7]Nihon University (Japan), [8]University of Tokyo (Japan)

Eusocial insects have polyphenic caste systems in which each caste exhibits characteristic morphology and behavior. In insects, caste systems arose independently in different lineages, such as Isoptera and Hymenoptera. Although a number of caste specific expressed genes have been identified in termites, the regulatory mechanism of these gene expression changes is largely unknown. We hypothesize that the epigenetic regulation may play pivotal roles in these changes among termite castes and microRNAs (miRNAs) may also be involved in polyphenic caste differentiation as an epigenetic regulator. In this study, we therefore performed small RNA sequencing in the subterranean termite (Reticulitermes speratus) and identified the miRNAs that were specifically expressed in the soldier and worker castes. Of the 550 miRNAs annotated in the R. speratus genome, 74 are conserved in insects and 174 are conserved in the three termite species. We found that eight miRNAs (mir-1, mir-125, mir-133, mir-2765, mir-87a, and three termite-specific miRNAs) are differentially expressed (DE) in soldiers and workers. Further, four of the DE miRNAs in soldier and worker termite castes were also differentially expressed in hymenopteran castes, such as nurses and foragers. The finding that Isoptera and Hymenoptera partially shared DE miRNAs among castes suggests that these miRNAs evolved in parallel between phylogenetically distinct species, possibly contributing towards the elucidation of a molecular mechanism of eusociality.

POA-058

Functional sequence constraint on pseudogene 3' UTRs is suggestive of widespread competitive endogenous RNA activity

Cian Glenfield[1], Aoife McLysaght[1]

[1]Trinity College, The University Of Dublin (Ireland)

The competitive endogenous RNA (ceRNA) hypothesis is an attractively simple model to explain the biological role of many miRNAs and miRNA targets. Under this model, there exist transcripts in the cell whose role is to titrate out miRNAs such that the expression level of the genuine target sequence is altered. That it is logistically possible for expression of one miRNA-target(MRE)-containing transcript to affect another is seen in the multiple examples of pathogenic effects of inappropriate expression of MRE-containing RNAs. However, the role, if any, of ceRNAs in normal biological processes and at physiological levels is disputed. Evolutionary analysis, in particular testing for evolutionary constraint, is an unbiased arbiter of biological functionality. If a genomic element displays hallmarks of natural selection then it follows that that element contributes to the fitness of the organism. Importantly, natural selection is sensitive to fitness advantages more subtle than can readily be detected in an experimental setup. Here we find genome-wide evidence for functional sequence constraint acting on the 3' UTRs of pseudogenes, suggestive of a widespread biological ceRNA function, by identifying previously unannotated human pseudogene orthologs across 20 mammalian genomes. We focus also on the oncogene *BRAF* and its pseudogenes (*BRAFP1* in primates and *Braf-rs1* in mouse), as the ceRNA activity of these has been investigated experimentally. We find detailed evidence for functional constraint of the MREs shared between the parent gene and the pseudogene, further supporting the ceRNA hypothesis, at least for some genes in the genome.

POA-060

Exploring the relationship between vaccination and anti-NMDA receptor encephalitis based on microRNA phylogenetic tree

Hsiuying Wang[1]

[1]National Chiao Tung University (Taiwan)

MicroRNA (miRNA) is a short non-coding RNA, which can be used as a biomarker in the early diagnostic of many diseases. Anti-N-methyl-D-aspartate (Anti-NMDA) receptor encephalitis is an acute autoimmune neurological disorder which has been reported to be relate to vaccination. To exploring the relationship between the anti-NMDA receptor encephalitis and Japanese encephalitis vaccination, a protocol based on miRNA phylogenetic tree has been proposed in the literature. In this study, we investigate miRNA biomarker for several other vaccinations and apply this method to explore the association between anti-NMDA receptor encephalitis and these vaccinations. The miRNA phylogenetic trees based on different evolutionary model and linkage function are plotted and compared.

POA-061

Does Lamarckian evolution exist? Direct evidence of small RNA transport from somatic brain tissue to offspring

Elizabeth O'Brien[1], Kathleen Ensbey[1], Paul Baldock[2], Bryan Day[1], Guy Barry[1]

[1]QIMR Berghofer Medical Research Institute (Australia), [2]Garvan Institute of Medical Research (Australia)

Over the past 2 million years, the human brain has nearly tripled in size, coincident with the acquisition of higher order cognitive functions such as creativity, imagination and reasoning. A major contributor to this extraordinarily rapid evolution could be robust transgenerational inheritance, enabling immediate and heritable adaptive responses. However, a mechanism for somatic cell inheritance has to date been elusive and, although small RNAs have been implicated, no direct evidence has yet emerged. Here, we injected adeno-associated virus producing the human-specific miRNA MIR941 into the striatum of adult male mice. We first confirmed after 2, 8 and 14 weeks post-injection that the virus only infected locally at the injection site and did not leak into the bloodstream or lymph. For this we used specific locked nucleic acid (LNA) qPCR primers for rabbit B-globin as rabbit B-globin is introduced into the virus as a marker for viral production. Next, a LNA microRNA qPCR assay for MIR941 showed expression of MIR941 in injection sites at 2, 8 and 14 weeks and in the vas deferens (containing mature sperm) of 8 and 14 week post-injected male mice. The same male mice had been mated at 8 and 14 weeks post-injection, just prior to sacrifice, and remarkably we also found the presence of MIR941 in 8 day old embryos from both these time points. We believe that this is the first direct mechanistic demonstration of somatic cell inheritance and will form the basis of uncovering the full breadth and impact of transgenerational inheritance.

POA-064

M1CR0B1AL1Z3R - A user-friendly software tool for the analysis of microbial genomics data

Oren Avram[1], Tal Pupko[1]

[1]Tel Aviv University (Israel)

The significant technological advances in the last decade brought with them opportunities for large-scale mining and analysis of pathogenic species data in an unprecedented resolution. Such analyses contribute to the comprehensive characterization of complex microbial dynamics within a microbiome and among different strains during a disease outbreak, to name a few. Studying large-scale bacterial evolutionary dynamics posses many challenges. These include data-mining steps, such as gene annotation and orthologs detection, sequence alignment and accurate phylogenetic tree reconstruction. These steps as well as additional analysis-specific computations require the use of multiple bioinformatics tools and software, making the entire process cumbersome and tedious, and prone to errors due to manual handling.

This motivated us to develop an automatic easy-to-use pipeline that integrates basic and advanced analysis components. We are developing a user-friendly software tool (written in Python) called M1CR0B1AL1Z3R. The M1CR0BIALIZ3R tool is a "one-stop shop" for conducting such microbial genomics data analyses via command line or a simple graphical user interface. An example of features which are implemented in M1CR0BIALIZ3R include: (1) Extracting orthologous sets from input genomes; (2) Analyzing presence-absence patterns of genes and rates of gene gain and loss events on each branch; (3) Reconstructing a phylogenetic tree based on the extracted orthologous set; (4) Identifying selective sweeps events(Avram et al., in final preparation); (5) Inferring GC content variation among species lineages. This should allow scientists to analyze hundreds of bacterial genomes, with a click of a button.

POA-067

A journey from an ancient finger print of Rossmann fold enzymes to cofactor engineering

Paola Laurino[1]

[1]Okinawa Institute of Science and Technology (Japan)

Nucleoside based cofactors are presumed to have preceded proteins. The Rossmann fold is one of the most ancient and functionally diverse protein folds. We analyzed an omnipresent Rossmann ribose binding interaction, a carboxylate side chain at the tip of the second beta strand (beta2 Asp or Glu). We identified a canonical motif, defined by the beta2 topology and unique geometry. This motif is uniquely found in Rossmann enzymes that use different cofactors, primarily SAM, NAD and FAD. Ribose carboxylate bidentate interactions in other folds are not only rare but also have a different topology and geometry. Overall, these data indicate the divergence of several major Rossmann fold enzyme classes from a common pre Last Universal Common Ancestor (LUCA) that possessed the beta2 Asp or Glu motif.

While we were studying how Rossmann fold enzyme binding ribose based cofactor evolves, the adenosine mode of binding attracted our attention. Based on this observation we started our cofactor engineering studies to remodel the catalytic site for a new cofactor. Cofactor engineering aims to obtain enzymes with novel cofactor specificities, and orthogonal recognition of a synthetic cofactor instead of the natural one. Our focus is RNA methylases a Rossmann fold protein, that uses S adenosylmethionine as methyl donor to methylate specific RNA target sequences. Although these enzymes are fundamental cellular component, their targets are often unknown and their cellular role remains poorly understood. Our approach provides a powerful tool to study the cellular roles of RNA methylases, including methylation patterns.

POA-068

Relaxed evolutionary constraint of gene expression in the snake venom arsenal leads to diversification and parallelism.

Agneesh Barua[1], Alexander Mikheyev[1]

[1]Okinawa Institute of Science and Technology Graduate University (Japan)

Gene expression changes contribute to complex traits variations in both individuals and populations. However, it is not entirely clear as to how patterns of inheritance and adaptive landscape influence gene expression of multiple traits over deep evolutionary time. Being comprised of proteinaceous cocktails, snake venoms are unique in that the expression of each toxin can be quantified and mapped to a distinct genomic locus and traced for up to tens of millions of years. Using a generalized linear mixed model we analysed expression data of toxin genes from 45 snake species, and estimated the effect of phylogeny on trait relations and combinations. We find that combinations of toxins do not phylogenetically covary, meaning all combinations are in principle possible. We observed niche conservatism, where snakes from the three venomous families evolved envenomation strategies focused on four major toxins: metalloprotease, three-finger toxins, serine protease, and phospholipase A2. However, we also observe the frequent convergence of envenomation strategies by distantly related snake. We believe that the lack of association between toxins allowed snakes to explore various phenotypes for the venom, and through this exploration, the four major toxins emerged as consistently providing the most adaptive advantage. This caused snakes to diversify via lineage-specific evolution of these toxins. While the results showcase the importance of evolutionary constraints in shaping the venom phenotype of snakes, they also provide the first quantitative and comparative approach towards elucidating the long-term role of gene expression for evolution of the snake venom phenotype.

POA-069

Evolutionary evidence for independent origins of genes essential for the proper establishment of left-right asymmetry in amphibians and mammals

Juan Cristobal Opazo[1]

[1]Universidad Austral de Chile (Chile)

Growth differentiation factors 1 (GDF1) and 3 (GDF3) are members of the transforming growth factor superfamily (TGF-beta) that is involved in fundamental early-developmental processes that are conserved across all vertebrates. The evolutionary history of these genes is still under debate due to ambiguous definitions of homologous relationships among vertebrates. Thus, the goal of this study was to unravel the evolution of the GDF1 and GDF3 genes of vertebrates, emphasizing the understanding of homologous relationships and their evolutionary origin. Surprisingly, our results revealed that the GDF1 and GDF3 genes found in amphibians and mammals are the products of independent duplication events of an ancestral gene in the ancestor of each of these lineages. The main implication of this result is that the GDF1 and GDF3 genes of amphibians and mammals are not 1:1 orthologs. In other words, genes that participate in fundamental processes during early development, such as establishing the left-right identity of the body plan, have been reinvented two independent times during the evolutionary history of tetrapods.

POA-070

Genomic signatures of a Mannheimia haemolytica lineage associated with bovine respiratory disease

Michael Clawson[1], Gennie Schuller[1], Aaron Dickey[1], Robert Murray[2], Michael Sweeney[3], Michael Apley[4], Keith DeDonder[5], Sarah Capik[6, 7], Robert Larson[4], Brian Lubbers[4], Brad White[4], Jochen Blom[8], Dayna Brichta-Harhay[1], Timothy Smith[1]

[1]United States Department of Agiculture, Agricultural Research Service (United States), [2]Zoetis (United States), [3]Zoetis (United States), [4]Kansas State University (United States), [5]Veterinary and Biomedical Research Center (United States), [6]Texas A&M (United States), [7]Texas A&M (United States), [8]Justus-Liebig-University Giessen (Germany)

A major challenge in the modern age of microbiology is to develop narrow-spectrum interventions that target pathogens, but not the remaining microbiome of a host, including less- or non-virulent strains of the same pathogenic species. *Mannheimia haemolytica* is a model study organism in that regard. *M. haemolytica* is both a normal resident of the upper respiratory tract of cattle, and a major cause of bovine respiratory disease. Sequencing of over 1,100 isolate genomes has shown that there are two major lineages, or genotypes of *M. haemolytica* (1 and 2) in North American cattle. While both genotypes are found in the upper and lower respiratory tract of cattle with or without signs of disease, genotype 2 predominantly associates with the diseased lungs of cattle over genotype 1. Over 13,000 genome-wide polymorphism alleles separate the two lineages. Additionally, an integrative conjugative element with variable numbers of antimicrobial resistance genes has been found in all genotype 2 isolates examined to date, and rarely in genotype 1 isolates. The two genotypes share a core genome of 1,880 proteins, and differ by 112 and 179 proteins with specificity for genotype 1 and 2, respectively. Analyses of the proteins specific to each genotype, and those shared in the core, particularly those with beta-barrel topology which is indicative of outer membrane localization, point towards the design of next-generation vaccines that could target only genotype 2 *M. haemolytica* to minimize disruption to the bovine microbiome, or both genotypes if the need arises.

POA-071

Age-dependent patterns of adaptation to diet in Drosophila melanogaster

Grant Allen Rutledge[1], Kevin H Phung[1], Laurence D Mueller[1], Michael R Rose[1]

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

A variety of anthropologists and physicians claim that the health of present-day humans would be enhanced by reversion to "Paleo" diets. Against them, a few assert that long-agricultural populations are well-adapted to agricultural diets, due to the speed with which natural selection can fashion effective adaptations to novel diets. But theoretical analysis based on Hamilton's forces of natural selection suggests that both might be wrong: populations might adapt to a novel environment quickly at early ages, but only slowly and incompletely at later adult ages. Numerical calculations support this general Hamiltonian hypothesis of early-weighted adaptation to novel environments and diets. Experimental tests for age-dependent adaptation to a novel diet were performed on populations of *Drosophila melanogaster*. The results support the Hamiltonian hypothesis of age-dependent adaptation, with populations performing better on their ancestral or long-standing diet, compared to an evolutionarily recent diet, only at later ages. Additionally, populations performed poorly on an entirely novel diet compared to an evolutionarily recent diet that has been sustained for hundreds of generations, particularly at earlier ages. These findings suggest that humans could revert to an ancestral diet at later ages to alleviate some chronic disorders. However, at earlier ages, long-agricultural human populations might be best able to achieve reasonable health on an agricultural diet.

POA-072

Evolutionary biology meets synthetic biology: designing a translational machinery that enhances incorporation of non-proteinogenic amino acids into proteins by evolutionary analysis

Mariko F. Matsuura[1], Sarah Lucas[2], Vanessa E. DeLey Cox[3], Eric A. Gaucher[1, 3, 4]

[1]Georgia State University (United States), [2]Georgia Institute of Technology (United States), [3]Georgia Institute of Technology (United States), [4]Georgia Institute of Technology (United States)

Non-proteinogenic amino acids (NPAAs) are amino acids (AA) that are not genetically encoded in organisms. However, NPAAs play important roles in the evolution of life: (1) they are found in meteorites and are also formed in the Miller-Urey experiment thus highlighting their potential importance as prebiotic organic molecules (e.g. D-, beta-AA); (2) they are intermediates of biosynthetic pathways; (3) and, lastly, they regulate the functions of many present-day enzymes and proteins (e.g. phospho-, acetyl-AA).

To study the roles of NPAAs in proteins, it is necessary to incorporate NPAAs in all positions of a protein. However, current ribosomal NPAA incorporation methods have limitations: (1) multiple and consecutive NPAA incorporations are challenging; (2) *in vitro* NPAA-tRNAs synthesis is hard to scale up; and (3) misincorporation of proteinogenic AA (PAA) occurs.

The goal of this study is to create proteins containing multiple and consecutive NPAAs by engineering Elongation Factor-Tu (EF-Tu). EF-Tu is one of essential components of the translational machinery, which accurately delivers PAA-tRNAs to the ribosome by finely tuned interactions between EF-Tu and PAA-tRNAs. In our study, EF-Tu variants were designed to accommodate NPAA-tRNAs by exploiting the reconstructing evolutionary adaptive paths (REAP) method so to improve *in vivo* NPAA incorporation efficiency. Then, an *in vivo* assay was developed to confirm and screen the variants activity. Results obtained by the assay and potential applications using the variants will be discussed. This is an example of research that applies concepts of evolutionary biology to the field of synthetic biology.

POA-074

A single pheromone receptor gene shared among most bony vertebrates

Masato Nikaido[1], Hikoyu Suzuki[1], Takehiko Ito[1], Junji Hirota[1]

[1]Tokyo Institute of Technology (Japan)

Pheromones are crucial for eliciting social and sexual behaviors in diverse animal species. The vomeronasal receptor type-1 (V1R) genes, encoding members of a pheromone receptor family are highly variable in number and repertoire among mammals due to extensive gene gain and loss. Here, we report a novel pheromone receptor gene belonging to the V1R family, named ancient V1R (ancV1R), which was found to be shared among most bony vertebrates from basal lineage of ray-finned fish to mammals. Phylogenetic and syntenic analyses suggest that ancV1R represents an orthologous gene retained for >400 million years of vertebrate evolution. Interestingly, ancV1R pseudogenizations observed in some species are coincident with degeneration of vomeronasal organ in higher primates, cetaceans, and birds, etc., suggesting crucial roles of ancV1R in the vomeronasal function. Furthermore, ancV1R exhibits unique molecular features with respect to its unexpected expression patterns, being expressed in all mature vomeronasal sensory neurons and co-expressed with both V1Rs and V2Rs, that challenges the "one neuron-one receptor" rule. Our findings of the evolutionarily conserved pheromone receptor across most bony vertebrates may shed insight into the mechanism for general pheromone detection and signal transduction in vertebrate vomeronasal sensory neurons.

POA-075

Ecological influence of sediment bypass tunnels on macroinvertebrates in dam-fragmented rivers using DNA metabarcoding

Joeselle Serrana[1], Sakiko Yaegashi[1, 2], Shunsuke Kondoh[1], Bin Li[1], Christopher Robinson[3], Kozo Watanabe[1]

[1]Ehime University (Japan), [2]Yamanashi University (Japan), [3]EAWAG (Switzerland)

Sediment bypass tunnels (SBTs) are guiding structures in reservoirs used to reestablish sediment regimes downstream. Previous studies monitoring the ecological effects of SBT operation on downstream reaches suggest a positive influence of SBTs on the recovery of riverbed condition and macroinvertebrate community through traditional morphology-based surveys. However, morphology-based macroinvertebrate assessments are not just costly and time-consuming but the large number of morphologically cryptic, small-sized and undescribed species results to a coarse taxonomic level of identification. In this study, we utilized metabarcoding analysis to assess the influence of SBT operation on macroinvertebrate communities by estimating species diversity and pairwise community dissimilarity in dam-fragmented rivers with operational SBTs in comparison to dam-fragmented (i.e., without SBT) and free-flowing rivers (i.e., without dam). Community dissimilarities between the up- and downstream sites assessed for the Reuss River (Pfaffensprung dam) and Rabiusa River (Egschi dam) were relatively low, similar to the free-flowing rivers, while observed values for Albula River (Solis dam) were relatively high, similar or higher compared to the dam-fragmented rivers. Total sample abundance of the morphologically-identified specimen was significantly positively correlated to read abundance which validates and reinforce the use of quantitative estimates for the diversity analysis of our metabarcoding data. We report that macroinvertebrate community dissimilarity decreases with increasing operation time and frequency of SBT. These could be major factors influencing the recovery of riverbed features that would subsequently support the recovery of the downstream macroinvertebrate community similar or close to the upstream community.

POA-076

Evolution of rapid life cycle through deletion of a genetic hotspot after recent gene duplication in Boechera stricta

Cheng-Ruei Lee[1], Eric Schranz[2], Thomas Mitchell-Olds[3]

[1]National Taiwan University (Taiwan), [2]Wageningen University & Research Center (Netherlands), [3]Duke University (United States)

Differences in the timing of vegetative-to-reproductive phase transition have been independently and repeatedly evolved in different plant species. Due to their specific biological functions and positions in pathways, some genes are "genetic hotspots" of repeated evolution - independent mutations on these genes caused the evolution of similar phenotypes in distantly related organisms. While many studies have investigated these genetic hotspots, it remains unclear how gene duplications, which create functional redundancy, influence the repeated phenotypic evolution and patterns of pleiotropy. Here we characterized the genetic architecture underlying a novel rapid-flowering phenotype in *Boechera stricta* and investigated the candidate genes *BsFLC1* and *BsFLC2*. The deletion of *BsFLC1* conferred rapid flowering and loss of vernalization requirement, and the expression patterns of functional *BsFLC1* suggested function consistent with the *Arabidopsis homolog*. In contrast, *BsFLC2* did not appear to suppress flowering and had accumulated multiple amino acid substitutions in the relatively short evolutionary timeframe after gene duplication. These non-synonymous substitutions greatly changed the physicochemical properties of the original amino acids, concentrated non-randomly near a protein-interacting domain, and had greater substitution rate than synonymous changes. Here, we showed that, after recent gene duplication of the genetic hotspot *FLC*, the evolution of rapid phenology was achieved through the neo- or sub-functionalization of *BsFLC2* followed by the deletion of *BsFLC1*.

POA-077

Whole genome sequencing of a Japanese endemic pit viper, habu, *Protobothrops flavoviridis* reveals accelerated evolution of venom protein genes enriched in microchromosomal regions.

Hiroki Shibata[1], Takahito Chijiwa[2], Naoko Oda-Ueda[3], Kazuaki Yamaguchi[2], Shosaku Hattori[4], Kazumi Matsubara[5, 6], Yoichi Matsuda[6], Ryo Koyanagi[7], Kanako Hisata[8], Yasuyuki Fukumaki[1], Motonori Ohno[2], Eiichi Shoguchi[8], Noriyuki Satoh[8], Tomohisa Ogawa[9]

[1]Kyushu University (Japan), [2]Sojo University (Japan), [3]Sojo University (Japan), [4]University of Tokyo (Japan), [5]Nagoya City University (Japan), [6]NagoyaUniversity (Japan), [7]Okinawa Institute of Science and Technology Graduate University (Japan), [8]Okinawa Institute of Science and Technology Graduate University (Japan), [9]Tohoku University (Japan)

Evolution of novel traits is a challenging subject in biological research. Several snake lineages developed elaborate venom systems to deliver complex protein mixtures for prey capture. To understand mechanisms involved in snake venom evolution, we decoded here the ~1.4-Gb genome of a Japanese endemic pit viper, habu, *Protobothrops flavoviridis*. We identified 73 snake venom protein genes (SV) and 251 non-venom paralogs (NV), belonging to 24 gene families. Molecular phylogeny revealed an early divergence of SV and NV gene copies, suggesting that one of the four copies generated through two rounds of whole-genome duplication was modified for use as a toxin in the venom. Among them, both SV and NV gene families of the four major venomic components, metalloproteinase, serine protease, C-type lectin-like protein and phospholipase A2 were extensively duplicated after their diversification. An accelerated evolution was evident in their SV genes but not in NV counterparts. On the other hand, genes for the other 20 families those were not extensively duplicated showed no evidence of accelerated evolution. We also observed that venom-related SV and NV gene copies are significantly enriched in microchromosomes than in macrochromosomes, suggesting the implementation of the genomic architecture in the multiplication and the accelerated evolution in the venom-related genes.

POA-078

The effective population size is correlated to census population size in mammals.

Jennifer James[1], Adam Eyre-Walker[1]

[1]University of Sussex (United Kingdom)

The factors that determine the level of genetic variation of a species remain one of molecular evolution's enduring mysteries. The level of neutral genetic diversity depends upon the mutation rate and effective population size, and the latter might reasonably be expected to be correlated to census population size. However, genetic diversity is generally uncorrelated to any measure of census population size. To help understand what factors determine the level of genetic diversity, we compiled mitochondrial DNA sequence diversity data from >600 mammalian species. Controlling for phylogeny by using paired comparisons we find that synonymous nucleotide diversity is simultaneously positively correlated to range size, and negatively correlated to body size and latitude, with the strongest effect being that of range size. Failing to control for phylogeny yields very different results. To investigate whether the correlation between diversity and range size is due to variation in effective population size, we also analysed a measure of the efficiency of selection - synonymous nucleotide diversity divided by the sum of the synonymous and non-synonymous nucleotide diversities. We find that this measure is negatively correlated to range size. It therefore seems that the effective population size does increase with census population size. However, the relationship is weak both in terms of its slope and the variance explained.

POA-079

Closing the Lipid Divide: Phylogenetic analysis of phospholipid biosynthetic pathways in Archaea and Bacteria

Gareth A. Coleman[1], Richard D. Pancost[2], Tom A. Williams[1]

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

Archaea and Bacteria represent the two fundamental domains of life. One of the differences between the two domains lies in their membrane phospholipids, which are synthesised by two distinct biosynthetic pathways with non-homologous enzymes. This apparent 'lipid divide' has led to many evolutionary questions, including whether the last universal common ancestor (LUCA) had a membrane, and if so, what the nature of this membrane was. However, much evidence suggests a less clear cut divide than was previously thought. Many lipids found in the environment exhibit a mixture of bacterial and archaeal traits, with their provenance and biosynthetic pathways remaining unclear. Taken together with evidence for extensive horizontal gene transfer between Archaea and Bacteria, this prompted us to investigate the distribution of phospholipid biosynthesis enzymes across the two domains and look for evidence of transfer of these genes. We downloaded the sequences for the archaeal enzymes from 43 archaeal genomes, and the bacterial enzymes form 64 bacterial genomes. We performed BLAST searches for the archaeal enzymes in bacterial genomes and vice versa, and made Maximum Likelihood and Bayesian trees for each enzyme. These trees were rooted using minimum ancestor deviation (MAD) and the molecular clock. Our results show that inter-domain transfer is extensive, and that the distribution of phospholipid biosynthesis genes across the tree of life is patchy, with a diversity of different lipids being found in various lineages. Our rooting results further suggest that archaeal genes can be mapped back to LUCA, suggesting that LUCA had an archaeal-like membrane.

POA-080

Differences between de novo genes and their non-functional precursors can result from neutral constraints on their birth process, not necessarily from natural selection alone

Lou Nielly-Thibault[1, 2, 3], Christian R Landry[1, 2, 3]

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

The *de novo* origination of protein-coding genes is receiving increasing attention for its contribution to the evolution of proteomes and phenotypes. The role of natural selection in shaping the properties of these genes is of particular importance. Studies in the field typically use non-coding DNA and GC-content-based random-sequence models to generate random expectations for the properties of novel polypeptides. Deviations from these expectations have been interpreted as the result of natural selection. However, interpreting such deviations requires a yet-unattained understanding of the raw material of *de novo* gene birth and its relation to novel functional polypeptides. We combine modelling with analysis of published translation data to show how the importance of the "junk" polypeptides that make up this raw material goes beyond their average properties and their filtering by natural selection. We find that the variance of the properties of junk polypeptides and their correlation with the rate of evolutionary turnover affect the properties of novel functional polypeptides. We also bring empirical support to a GC-content-based random-sequence model of the raw material of *de novo* gene birth in the yeast *Saccharomyces cerevisiae*. Under this model, GC content can have different and even opposite effects on the means of some polypeptide properties between the raw material and the products of *de novo* gene birth. Our results provide a theoretical framework that can serve as a guide for the design and interpretation of future empirical studies of novel genes and as a basis for further theoretical developments in this field.

POA-081

An initiative for genetic data collection from underrepresented countries and populations

Kimberly F McManus[1], Meghan Moreno[1], Joanne Kim[1], Kasia Bryc[1], Joanna Mountain[1]

[1]23andMe (United States)

It is well-known that currently available human genetic data do not adequately represent the breadth and depth of the world's diversity. The lack of genetic research on people of diverse ancestries greatly limits a comprehensive understanding of genetic variation in *Homo sapiens*, as well as hinders understanding of genetic disease risk in populations underrepresented in genetic research.

23andMe's Global Genetics Project, launched in 2018, seeks to genotype individuals with four grandparents from large populations throughout the world that are underrepresented in genetic research. This initiative follows our 2016-17 African Genetics Project, which enrolled over 1,100 people with four grandparents from certain African countries, including hundreds of samples from each of Ethiopia, Somalia and Sudan.

Over the next few years, we aim to enroll 6,000 participants representing 40 countries in regions including central, western and southeastern Asia; northern, western and southern Africa; Oceania; and Central and South America. Consenting participants, who are recruited from the US, are asked to provide a saliva sample and complete a survey about their family's birthplaces, cultural affiliations and languages.

These data will be used for a variety of ancestry and health research projects. For ancestry research, they will be used to improve understanding of population structure, diversity and migrations throughout the world. They will be used to further health research in underrepresented populations by exploring models of disease risk in these populations.

POA-082

Population structure in pre-contact North America: a whole-genome ancient DNA study

Christiana Scheib[1, 2], Hongjie Li[3], Vivian Link[4], Christopher Kendell[6], Genevieve Dewar[6], Peter William Griffith[1], Alexander Moerseburg[1], John R. Johnson[7], Amiee Potter[8, 9], Susan L. Kerr[10], Phillip Endicott[11], John Lindo[12], Marc Haber[5], Yali Xue[5], Chris Tyler Smith[5], Manj Sandhu[5], Richard Durbin[5], Joseph G. Lorenz[13], Tori D. Randall[14], Zuzana Faltyskova[1]

[1]University of Cambridge (United Kingdom), [2]University of Tartu (Estonia), [3]University of Illinois Urbana-Champaign (United States), [4]University of Fribourg (Switzerland), [5]University of Toronto (Canada)

On behalf of all authors. See emailed abstract

POA-083

Genetic features of the Korean short-necked clam, Ruditapes philippinarum, via next-generation sequencing and comparative genomic analyses and their gene flow among Asian-Pacific Countries

Hye Suck An[1], Seyoung Mun[2, 3], Jiyoung Woo[1], Young Se Hyun[1], Ha Yeun Song[1], Jongsu Yoo[1], Kyudong Han[2, 3]

[1]National Marine Biodiversity Institute of Korea (Republic of Korea), [2]Dankook University (Republic of Korea), [3]DKU-Theragen institute for NGS analysis (DTiNa) (Republic of Korea)

The short-necked clam, *Ruditapes philippinarum*, is an important bivalve species in worldwide aquaculture including Korea. In spite of its importance in marine resource, the comprehensive genetic study of short-necked clam is rare. We`ve reported the whole-genome sequencing with *de novo* assembly from the short-necked clam and whole-transcriptome analysis with total RNA sequencing across its three different tissues (foot, gill, and adductor muscle). Our study conducted identification of the repetitive elements including simple sequence repeats (SSRs) and non-coding RNAs (ncRNAs), and taxonomy profiling to provide more inclusive understanding of the short-necked clam genome. Interestingly, we found that the gene family related with the innate immune response has been remarkably expanded in *R. philippinarum*. Based on the whole-transcriptome data, we also identified differential expressed genes (DEG) across three tissues and validated tissue-specific expressed genes using real-time PCR. Furthermore, its gene flow was examined across its northwestern Pacific range by screening variation of eight microsatellite loci in the present study. All 10 populations in Korea, Japan and China exhibited moderate genetic diversity. Pairwise fixation index (*F*st) suggested a low to moderate level of genetic differentiation among populations. Although significant relationship was not observed between genetic and geographic distances among the sampled populations, four populations of the Yellow Sea were differentiated from the other six populations based on the results of pairwise *F*st, three-dimensional factorial correspondence analysis and STRUCTURE, which implied gene flow within each group. We will more discuss several conservation and management strategies based on the findings.

POA-084

Comparative genome analysis of *Ralstonia solanacearum*causing potato bacterial wilt in Korea

Heejung Cho[1], Young Kee Lee[1], Seungdon Lee[1], Dong Suk Park[1], Jeong-Gu Kim[1]

[1]National Institute of Agricultural Sciences, Rural Development Administration (Republic of Korea)

Ralstonia solanacearum, causal agent of bacterial wilt, is one of the most destructive phytopathogen in the world. *R. solanacearum* has unusual broad host range over 450 plant species of 50 botanical families. This bacterium distributed worldwide encompassing tropical, subtropical, and temperate region. With these features, this species are very diverse and complex and call as pathogenic *Ralstonia solanacearum*species complex (RSSC). Here, we sequenced the genomes of twenty-five strains possessing distinctive host range (potato, tomato, eggplant, and pepper). Phylogenetic relationship analysis of these newly sequenced genomes was performed with previously published genome data of nine *R. solanacearum* species including phylotype I, IIA, IIB, III, and IV. With this result, we could suppose the relation of RSSC evolution and host range. Subsequently, functional genome comparisons were analyzed to investigate candidate genes related to the host adaptation. The strains of same pathotype group exhibited considerable repertoires for infection of tomato, eggplant, or pepper. By analyzing the type III secretion system effectors (T3Es), it was found three host-specific effectors; RipS3 (Skwp3) and RipH3 from the tomato-pathogenic strains and RipAC (PopC) from the eggplant-pathogenic strains. This study suggests that host range of *R. solanacearum* species is conferred by combination of host-specific effectors and other supportable virulence factors involving on regulatory mechanisms, secretion systems, and hydrolytic enzymes, etc.

POA-085

Phylogenetic incongruence of microdiversity in a marine bacterial population

Xiaojun Wang[1], Haiwei Luo[1]

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

How microdiversity evolves in natural bacterial populations remains unknown. We sampled 28 Sulfitobacter strains associated with sponges in the Red Sea and diatoms in Japanese Seto Inland Sea. These isolates are clustered into three clades differing by up to 0.6% in the 16S rRNA gene. Interestingly, all three possible clade relationships receive nearly equal support from gene trees. We tested alternate mechanisms that likely gave rise to this considerable phylogenetic incongruence, including rampant recombination between clades, incomplete lineage sorting (ILS) because of the short interval between speciation events, and fragmented speciation (FS) due to stepwise acquisitions of adaptive alleles. Only three gene trees with intermingled clade membership exclude the recombination hypothesis. The ancestral populations consistently show a greater genetic diversity than the extant populations in nearly all gene trees, which contradicts with the principle that only the gene trees supporting the speciation order have a greater ancestral diversity if ILS acted. FS therefore remains as the most likely mechanism. To further support FS as the underlying speciation model, we evaluated alternate bacterial speciation models. As the isolates do not cluster by geography, the allopatric speciation is rejected. In addition, substantial incongruence among gene trees rejects the genome-wide sweep model. Furthermore, the absence of homoplasious SNPs among clades goes against the gene-specific sweep model. Taking together, FS via stepwise acquisitions of moderately adaptive alleles followed by limited allele spread due to moderate recombination rate maintains genetic diversity of this population.

POA-087

The crossover landscape is more conserved than the double-strand-break landscape in yeast evolution

Haoxuan Liu[1], Jianzhi Zhang[1]

[1]University of Michigan (United States)

The rate of meiotic recombination varies across the genome with hot- and cold-spots, which has important implications for genome evolution. The hotspot paradox hypothesis predicts that recombination hotspots are evolutionarily unstable. However, evolutionarily stable hotspots of meiotic double-strand breaks (DSBs) were previously reported in divergent yeast species. Meiotic DSBs represent recombination initiations, only a subset of which are resolved as crossovers during meiosis. Hence, it remains possible that the crossover landscape is much less conserved than the DSB landscape. Here we investigate this possibility by generating a high-resolution map of recombination events in Saccharomyces paradoxus through whole-genome sequencing of fifty meiotic tetrads. We then compare this map with the corresponding map of its sister species S. cerevisiae. We find that the rate of crossover is lower in S. paradoxus than in S. cerevisiae. Nevertheless, the crossover landscape is not only conserved between the two yeasts that differ by ~15% in genome sequence but also more conserved than the DSB landscape. We provide evidence that this elevated conservation is explainable by the near-subtelomere preference of crossover in both species. This preference is caused not by crossover interference but likely by subtelomere positioning during chromosomal pairing. We conclude that the yeast crossover landscape is conserved, contrasting the hotspot paradox hypothesis.

POA-088

Population Genetic Models for Complex Disease Evolution

Jeremy J Berg[1], Guy Sella[1]

[1]Columbia University (United States)

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

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

POA-090

RNA-seq of single spermatogenic cysts shows gradual loss of dosage compensation but little evidence for meiotic X chromosome inactivation in Drosophila

Yumei Huang[1], Aimei Dai[1], Yixin Zhao[1], Xu Shen[1], Tian Tang[1]

[1]Sun Yat-sen University (China)

Two kinds of sex chromosome-specific regulation, namely dosage compensation and meiotic sex chromosome inactivation (MSCI), have evolved repeatedly in species with heteromorphic sex chromosomes. It is unclear how the X chromosome is regulated in the Drosophila testes. In particular, the existence of MSCI in Drosophila remains controversial. In this study, we performed single-cyst RNA-seq of stage-specific cell types from the Drosophila melanogaster male germline. First, we show dosage compensation is incomplete in Drosophila testes. Second, expression reduction of the X chromosome versus the autosomes accelerates at the pre-meiotic stage of the Drosophila male germline; the expression ratio of the X chromosome to the autosomes (X: A ratio) reaches a minimum slightly greater than one half in secondary spermatocytes and then persists in the post-meiotic cells. In contrast, X inactivation in human and mouse starts at the meiotic stage and results in drastic reduction of the X: A ratio in the post-meiotic cells. Third, expression of the dosage compensation complex (DCC) progressively decreases during Drosophila spermatogenesis. The distance of X-linked genes to high-affinity sites (HASs) of DCC is highly associated with their downregulation in spermatogenesis. These results suggest that gradual loss of dosage compensation would be the most parsimonious explanation for the X chromosome-specific expression reduction in the Drosophila male germline. Further analyses of male-biased genes and retrogenes show little evidence for MSCI. Our findings shed new lights on the different patterns of X chromosome regulation between Drosophila and mammals.

POA-091

What drives the chromosomal clustering of functionally related genes?

Haiqing Xu[1], Jianzhi Zhang[1]

[1]University of Michigan (United States)

There is mounting evidence that functionally related genes tend to be chromosomally clustered in eukaryotic genomes even after the exclusion of genes formed by tandem duplication, but the driving force of this phenomenon remains elusive. We propose that, because neighboring genes are likely to be controlled by the same chromatin domain, the stochastic expression variations of neighboring genes tend to be coordinated such that the expression ratio between them is more stable than that for unlinked genes. Consequently, gene clustering could be advantageous when the expression ratio of the clustered genes needs to be tightly regulated, for example, due to the accumulation of toxic compound when the expression ratio is misregulated. To test this hypothesis, we focus on the yeast (Saccharomyces cerevisiae) GAL gene cluster, which emerged through the relocations of originally unlinked genes in evolution. Specifically, the chromosomally adjacent GAL1, GAL7, and GAL10 encode enzymes catalyzing consecutive reactions in galactose catabolism, with a cytotoxic intermediate metabolite. By measuring the among-cell fluctuation of the protein expression ratio between linked and unlinked alleles in diploid cells, we confirm that linkage enhances the coordination of expression fluctuations of GAL genes. We then use CRISPR/Cas9-based genome editing techniques to perturb the physical linkage between the GAL genes followed by fitness essays. Indeed, clustering of GAL genes confers a significant fitness benefit in galactose but not glucose media. We conclude that minimizing the variation of expression ratio in the face of expression noise could drive the emergence of functionally related gene clusters.

POA-092

Evolution of S100A3 and PADI3 genes during the mammalian lineage

Takashi Kitano[1], Tadashi Minato[1]

[1]Ibaraki University (Japan)

S100A3 gene encodes a member of S100 family of proteins containing 2 EF-hand calcium-binding motifs which is highly expressed in the human hair cuticle. The 51st arginine residue of S100A3 protein is citrullinated by peptidyl arginine deiminase 3 (PADI3) specifically, and the citrullinated S100A3 protein is related to maturation of cuticle. Both members of S100 family and PADI family are tandemly duplicated, and mammals and birds have S100A3 and PADI3 genes, while amphibians and fishes do not have them. In the phylogenetic tree of the S100A gene family, the common ancestral branch of mammals of S100A3 genes showed relatively higher dN/dS (the number of nonsynonymous substitutions per nonsynonymous site/the number of synonymous substitutions per synonymous site) value (1.09), in contrast with lower dN/dS values (average 0.18) in mammalian branches. In the phylogenetic tree of the PADI gene family, the common ancestral branch of placental mammals and marsupials of PADI3 genes also showed relatively higher dN/dS value (0.98), in contrast with lower dN/dS values (average 0.10) in mammalian branches. The results suggest that the both S100A3 and PADI3 genes experienced relaxations of functional constraints rather than positive selection in the common ancestral branch of mammals after the divergence from birds and reptiles, and functional constraints of the both S100A3 and PADI3 genes were resumed after the acquisition of hair cuticle specific function in mammals.

POA-094

Short tandem repeats in the human, cow, mouse, chicken, and lizard genomes are concentrated in the terminal regions of chromosomes

Kazuharu Misaawa[1]

[1]Tohoku University (Japan)

Repetitive sequences in a genome cause mapping errors, especially in the case of short reads, because of the presence of similar or identical sequences. Distribution of repetitive sequences in a genome must be studied to distinguish between mappable and unmappable regions. Previous studies showed that short tandem repeats (STRs) are clustered in the terminal regions of chromosomes in the human genome. It is an open questionwhether formation of STRs in the terminal regions of chromosomes occurs only in humans. The present study investigated the distribution of STRs in the human, cow,mouse, chicken, and lizard genomes. In this study, itwas shown that STRs were concentrated in the terminal regions of chromosomes not only in the human genome, but also in the mouse, cow, chicken, and lizard genomes. The results suggested the mechanism through which STRs are shared by amniotes in which mammals, birds and lizards are included. Thus, we must be careful with the genomic sequences at chromosome ends of amniotes by using the next generation sequencers.

POA-095

Plastid genome mutational hotspots across gymnosperms with application for phylogenetic and barcoding studies

Edi Sudianto[1, 2, 3], Chung-Shien Wu[2], Shu-Miaw Chaw[1, 2]

[1]Taiwan International Graduate Program (Taiwan), [2]Academia Sinica (Taiwan), [3]National Taiwan Normal University (Taiwan)

Plastid sequences have been widely used in molecular biology studies for the past two decades. A number of plastid loci, including *rbcL, matK, ycf1*, the *trnL-UAA* intron, and the *trnH-psbA* intergenic spacers, were previously proposed to be good markers in phylogenetic and barcoding fields. However, their universality and discriminating power were largely based on studies of flowering plants. In contrast, exploration of useful genetic markers has not been conducted across the plastid genomes (plastomes) of the four multigeneric clades of gymnosperms: cycads (10 genera, 300 spp.), gnetophytes (3 genera, 70 spp.), Pinaceae (conifers I; 11 genera, 200 spp.), and cupressophytes (conifers II; 55 genera, 400 spp.). Here we report and compare plastomic mutational hotspots across 54 genera from the four gymnosperm groups (excluding ginkgo). Our results indicate that the degree of sequence divergence is generally higher in non-genic (including introns and intergenic spacers) than genic loci (Mann-Whitney test, $P < 0.001$ for all pairs). The top ten most divergent loci are distinctively different for each group. As a result, we propose that plastomic mutational hotspots evolved independently among the four gymnosperm groups. Moreover, we found that most of the hotspots are clustered, suggesting concerted evolution in plastid genes. These clustered loci constitute specific mutational islands that will be useful for future phylogenetic studies and species discrimination.

POA-096

Evolutionary changes in the thermosensory system contributed to the acquisition of heat tolerance in Buergeria japonica tadpoles inhabiting hot springs

Shigeru Saito[1, 2], Claire T. Saito[1], Takeshi Igawa[3], Shohei Komaki[4], Makoto Tominaga[1, 2]

[1]Okazaki Institute for Integrative Bioscience (National Institute for Physiological Sciences) (Japan), [2]SOKENDAI (The Graduate University for Advanced Studies) (Japan), [3]Hiroshima University (Japan), [4]Iwate Tohoku Medical Megabank Organization (Japan)

Temperature is a critical environmental factor for organisms, and extreme heat or cold causes detrimental effects. However, some of the species acquired resistance to harsh thermal environments and thereby occupying the niches where most other species are unable to utilize. Tadpoles of *Buergeria japonica* possess extreme heat resistance and even inhabit geothermal hot springs. This species offers the opportunity to understand the evolutionary mechanism of thermal adaptation processes to such extreme environments. In the present study, we focused on the role of thermosensory system in the acquisition of heat resistance in *B. japonica* by examining their behavioral responses to heat. *B. japonica* tadpoles tolerated high temperature up to about 41 degrees C and, above that temperature, they showed an abnormal swimming behavior. However, the tadpoles did not avoid 41 degrees C in the thermal selection assay using two-chamber system. We then cloned TRPV1 which serves as a heat sensor and characterized its functional property by electrophysiological approaches. In contrast to TRPV1 from clawed frogs which was activated around 40 degrees C, TRPV1 from *B. japonica* did not respond to heat stimulation up to 45 degrees C, although its channel function was retained. All these observations suggested that *B. japonica* tadpoles reduced thermal sensitivity to noxious heat, which enables them to reside in extreme thermal environments such as hot springs. Therefore, the evolutionary change in thermosensory system partly contributed to the acquisition of heat tolerance in *B. japonica*.

POA-097

Rapid Evolution of Vision in Sea Snakes

Bruno F Simoes[1, 2], Filipa L Sampaio[3], Julian C Partridge[4], David M Hunt[4], Nathan S Hart[5], Davide Pisani[1], Belinda SW Chang[6], David J Gower[3], Kate L Sanders[2]

[1]University of Bristol (United Kingdom), [2]The University of Adelaide (Australia), [3]The Natural History Museum, London (United Kingdom), [4]The University of Western Australia (Australia), [5]Macquarie University (Australia), [6]University of Toronto (Canada)

Sea snakes are among the most ecologically specialised of all squamate reptiles. This group diverged from terrestrial Australian elapid snakes approximately 6-16 MY and radiated into at least 60 species found from open ocean waters to coastal and mangrove habitats. Sea snakes are highly adapted to their marine lifestyle, having paddle-shaped tails, sealed nostrils, and respiratory traits that allow them to remain active underwater for several hours. However, how their visual system adapted to this unique environment is still poorly known. Previous studies suggested a shift from UV sensitive vision to blue sensitivity and photoreceptor transmutation (cones expressing rod-related genes) in two sea snake species. Our results from sequencing visual pigment genes for 151 sea snakes of 49 species suggest that the visual system of sea snakes is highly dynamic. At least 7 transitions are inferred between UV sensitivity and blue sensitivity and these may be linked to diel activity patterns. More interestingly, some species display short-wavelength (UV-Blue) trichromacy, a visual genotype that is rarely reported in vertebrates and may represent a trans-species polymorphism maintained by balancing selection. The fast diversification of the sea snake visual system and novel short-wavelength trichromacy makes this snake lineage exceptional to study the evolution of vision among vertebrates.

POA-100

The recombination landscape in species and subspecies of wild murid rodents

Ben Jackson[1], Tom Booker[1], Peter Keightley[1]

[1]University of Edinburgh (United Kingdom)

Recombination affects evolution directly through, for example, GC-biased gene conversion and recombination-associated mutation, and indirectly, by breaking down linkage disequilibria, which allows alleles at different sites in the genome to follow different evolutionary trajectories. This has consequences for the extent of Hill-Robertson interference and the dynamics of neutrally evolving loci linked to sites under selection. Understanding how recombination rate varies across genomes is therefore an important step towards answering many fundamental questions in evolutionary genetics. The ability to assay recombination rates using linkage disequilibrium information means that studies incorporating detailed recombination information are no longer confined to species with genetic maps, which are costly and time-consuming to produce. This is advantageous, because it allows researchers to quantify recombination rates in any population for which they have genetic polymorphism data and further, linkage-based maps can provide higher resolution than genetic maps. In this study we generate linkage-based maps for the brown rat, *Rattus norvegicus*, *Mus spretus* and three subspecies of the house mouse, *Mus musculus*. We use these data to understand how recombination rates have changed over evolutionary time in these taxa at broad and fine scales. We also investigate the effect of recombination rate variation on other evolutionary processes and patterns, including genetic differentiation and GC-biased gene conversion.

POA-101

Comparative study of lactate-mediated neural plasticity genes and its implication to long-term memory formation

Amal Abdulrhman Bajaffer[1, 2]

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

Lactate is known to work as an energy source in muscles, neurons and other tissues. Recently it was suggested that lactate works as a signaling molecule in neuronal plasticity system in long-term memory (LTM). Therefore, it is of particular interest to know how lactate gets involved in the neuronal plasticity function during the evolution. For this purpose, in this study, we examined the evolutionary origin and process of lactate-mediated neuronal plasticity (LMNP) system. We examined a phylogeny of six LMNP genes (Arc, c-Fos, Egr1, BDNF, C/EBP, and NMDAR) whose expressions were induced by lactate in neurons. As a result, we found that the emergence times of all the 5 LMNP genes, except Arc, were much before the emergence of vertebrates and very different, suggesting that the LMNP system evolved gradually by adding necessary genes to the system until it has become the present form by the emergence time of the common ancestor of vertebrates. When we extended this approach to 243 LTM-related genes of mice, we found that their ancestral genes emerged at different points of time before the emergence of vertebrates. We, therefore, conclude that the LTM system including the LMNP system has been formed by gradual participation of necessary genes in this system during evolution.

POA-103

Helicobacter pylori suggests early human migration in Asia

Rumiko Suzuki[1], Osamu Matsunari[1], Naruya Saitou[2, 3], Yoshio Yamaoka[1]

[1]Oita University Faculty of Medicine (Japan), [2]National Institute of Genetics (Japan), [3]The Graduate University for Advanced Studies (Japan)

Helicobacter pylori (*H. pylori*) is a gram-negative bacterium that colonizes human stomachs. About a half of the world population is estimated to be infected by this bacterium. *H. pylori* transmits by intimate contact to an infected person during early childhood when the immune system is not fully developed. Basically infection occurs within a family and transmits vertically. Because of such an infectious tendency, genealogy of *H. pylori* correspond well with that of human population. We collected wide variety of *H. pylori* strains from South to East Asian countries and found there are tow unique *H. pylori* groups in Ryukyu Islands, the southeast end of the Japanese archipelago. One of the *H. pylori* groups in Ryukyu formed a sub-branch near to East Asian strains distinct from the main island of Japan. The other group formed a sub-branch near to South Asian strains showing a divergence time earlier than any East Asian strains. Such unique *H. pylori* groups suggest that multiple human populations had reached to Ryukyu area before subsequent migrants from China or Korea came with farming to form the major population in the main island of Japan. Furthermore, population structure analysis suggested relationship between one of the Ryukyu groups and Oceanic strains. This result implies this Ryukyu group might have been brought via southern sea.

POA-104

A developmental switch generating phenotypic plasticity is part of a conserved supergene in Pristionchus nematodes

Bogdan Sieriebriennikov[1], Neel Prabh[1], Mohannad Dardiry[1], Hanh Witte[1], Waltraud Roeseler[1], Manuela R Kieninger[1, 2], Christian Roedelsperger[1], Ralf J Sommer[1]

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

Supergenes are polymorphic multi-gene complexes that regulate the formation of alternative phenotypes, such as in the iconic examples of primrose heterostyly or butterfly mimicry. Despite recent progress, in the majority of examples to date the identities of the causal genes are unknown or their roles remain untested. Also, alternative morphs are traditionally linked to different haplotypes, but the supergene concept may also extend to phenotypic plasticity, i.e. to environmentally induced phenotypes formed in isogenic background. Here, we show the first example of a supergene regulating plasticity. The locus controlling developmental switching between predatory and microbivorous morphs in the nematode *Pristionchus pacificus* contains two sulfatases and two α-N-acetylglucosaminidases (*nag*) in an inverted tandem configuration. We provide functional characterization of all supergene constituents using CRISPR/Cas9-based reverse genetics and show that the nag genes and the previously identified *eud-1*/sulfatase have opposite phenotypic effects. These genes show non-overlapping neuronal expression and epistatic relationships. In contrast to other animal models and similar to the duplication that generated the supergene in *Primula* primroses, the synteny within the *Pristionchus* supergene is conserved in the entire genus, demonstrating that supergene organization can be retained over tens of millions years of evolution and multiple speciation events. However, the locus architecture is different in other dimorphic genera of the same family. Interestingly, divergence between paralogs in the supergene is counteracted by gene conversion as inferred from phylogenies and observed experimentally in the genotypes of CRISPR/Cas9-induced mutants. Thus, supergenes can control alternative phenotypes generated by plasticity or polymorphisms.

POA-105

Universally High Transcript Error Rates in Bacteria

Weiyi Li[1], Michael Lynch[2]

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

Errors can occur at any level during the replication and transcription of genetic information. Genetic mutations, which are derived mainly from replication errors, have been extensively studied in evolutionary research. However, many fundamental details of transcript errors, such as their rate, molecular spectrum and selective constraints, remain largely unknown. To globally identify transcript errors, we applied an adapted rolling-circle sequencing (CirSeq) approach in *Escherichia coli*, *Bacillus subtilis*, *Agrobacterium tumefaciens*, *Mesoplasma florum*, and consistently revealed high transcript error rates, 3 to 4 orders of magnitude higher than the corresponding genetic mutation rates. With a total of 13,013 transcript errors identified from all four species, the molecular spectrum was uncovered in great detail. In addition to a widely known C-to-U substitution bias, a G-to-A bias was also observed in *M. florum*, which has the lowest known effective population size in bacteria. We also characterized the potential functional effects of observed transcript errors. To our surprise, an enrichment of nonsense errors was observed at the 3' end of mRNA transcripts, suggesting a Nonsense-Mediated Decay (NMD)-like quality-control mechanism in prokaryotes. We also evaluated the variation of error rates within and across species in coding and non-coding RNA (ncRNA) regions, and revealed a stronger selective constraint on errors in the ncRNA transcripts. Most intriguingly, the total error rate of ncRNA regions negatively correlates with the effective population size, which supports a drift-barrier hypothesis for the transcript error rate evolution.

POA-106

Evolution of isoprenoid biosynthesis from bacteria to eukaryotes

Yosuke Hoshino[1], Eric Gaucher[1, 2]

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

Isoprenoids (or terpenoids) represent one of the largest groups of organic compounds in nature and are distributed universally in the three domains of life. Understanding the evolutionary history of isoprenoid biosynthesis in each domain of life is critical since isoprenoids are deeply interwoven in the architecture of life and thus would have had indispensable roles in the early evolution of life. Among isoprenoids, sterols are universally found in eukaryotes and are integral components of eukaryotic membranes used to regulate cell rigidity and fluidity. In nature, several other isoprenoids such as hopanoids and tetrahymanol have also been discovered in many bacteria and some eukaryotes, respectively, and are considered to have analogous functions to sterols in their host organisms. Even though a common origin of all isoprenoids has been inferred, their evolutionary history is still enigmatic. Our study provides a detailed phylogenetic analysis of enzymes involved in the isoprenoid biosynthesis. Our analysis suggests that the transition from bacterial hopanoids to eukaryotic sterols entailed a complex trajectory including both vertical and horizontal transfers of isoprenoid biosynthesis genes among bacteria and eukaryotes. In particular, aerobic members of δ-proteobacteria (myxobacteria) are inferred to have had an integral role in the evolution and dissemination of the ability to synthesize isoprenoids in various organisms including both bacteria and eukaryotes.

POA-107

Native Metals And CO2 Reduction In Early Biochemical Evolution

Martina Preiner[1], Mingquan Yu[2], Sreejith J Varma[3], Kamila B Muchowska[3], Filipa L Sousa[4], Joana C Xavier[1], Harun Tueysuez[2], Joseph Moran[3], William F Martin[1, 5]

[1]Heinrich-Heine-University (Germany), [2]Max-Planck-Institut fuer Kohlenforschung (Germany), [3]Universite de Strasbourg, Institut de Science et d'Ingenierie Supramoleculaires (France), [4]University of Vienna, Department of Ecogenomics and Systems Biology (Austria), [5]Universidade Nova de Lisboa (Portugal)

When it comes to life's origin, there is only one thing we can say for sure: Without energy release, no chemical reactions can take place that could ultimately lead to complex chemicals and metabolism. Chemical energy at hydrothermal vents (mainly the H_2/CO_2 redox couple) is especially interesting here. Hydrogen is an ancient source of electrons while CO_2 is an ancient source of carbon. Prebiotic CO_2 reduction (carbon fixation) has been an issue for this early life hypothesis because the midpoint potential of H_2 is not conductive to direct reduction to complex carbon compounds. Modern anaerobes employ a mechanism called flavin based electron bifurcation involving iron-sulfur proteins to reduce CO_2 with H_2. But abiotically, FeS alone is not a likely reductant for CO_2 either, because it cannot readily engage in CO_2 reduction, which in known biological systems is always a two-electron reaction. Some acetogens and methanogens can actually grow on native iron (Fe^0) as a reductant, so might native metals solve the question about early CO_2 reduction? It appears that they do. We are investigating the intermetallic compound awaruite which occurs naturally in hydrothermal vents and contains native nickel (Ni^0) and Fe^0. To this day, several enzymes involved in carbon fixation employ those metals in their active centers. Using HPLC and NMR analytics we can show that in combination with awaruite, CO_2 and water can react to more complex carbon compounds like acetate, pyruvate and other molecules that could have played an important role in early metabolism.

POA-108

Origins of novel protein sequences de novo and by sequence divergence

Nikolaos Vakirlis[1], Anne-Ruxandra Carvunis[2], Aoife McLysaght[1]

[1]Trinity College Dublin (Ireland), [2]University of Pittsburgh (United States)

The genetic basis of evolutionary novelty is a central topic in evolution and biology. Novel protein sequences and protein functions play a crucial role in the emergence of novel phenotypes and eventually new species. The study of mechanisms of new protein coding gene origination has revealed multiple routes by which these can arise, including divergence of preexisting genes (preceded by gene duplication or not) as well as de novo emergence from previously noncoding genomic sequences. Although these 2 processes are fundamentally different, distinguishing them is a major challenge because they both give rise to "orphan" genes, sequences that have no detectable homologs in the genomes of other organisms. As a result, their relative contributions and evolutionary importance remain a matter of debate. Studying them separately however is an essential step in understanding the origin of the observed patterns of taxonomic distribution of gene families across the tree of life. Difficult questions such as what constitutes a novel gene and how can we estimate the timing of its origination lie at the core of the problem. I will present methodological advances that can allow us to reconstruct the history of innovation at the sequence level, quantify the contributions of the different mechanisms and gain insight into the properties and functions of the resulting novel genes.

POA-110

Evolution with recombination using state-of-the-art computational methods

Felipe Medina Aguayo[1], Richard Everitt[1]

[1]University of Reading (United Kingdom)

Abstract The ClonalOrigin model ([Didelot et al. 2010]) can be a good approximation to the process describing bacterial recombination, which is typically modelled using full ancestral recombination graphs. Inference in the ClonalOrigin model is performed via a reversible-jump MCMC (rjMCMC) algorithm, which attempts to jointly explore: the recombination rate, the number of recombination events, the departure and arrival points on the clonal genealogy for each recombination event, and the sites delimiting the start and end of each recombination event on the genome. However, as known by computational statisticians, the rjMCMC algorithm usually performs poorly due to the complexity of the target distribution since it needs to explore spaces of different dimensions. Recent developments in Bayesian computation methodology have provided ways to improve existing methods and code, but are not well-known outside the statistics community. We show how exploiting new computational methods can lead to faster inference when using the ClonalOrigin model.

POA-111

Systematic testing of a unified protocol for extracting DNA and proteins from ancient dental calculus

Zandra Fagernas[1], Maite Iris Garcia-Collado[2], Jessica Hendy[1], Courtney Hofman[3], Camilla Speller[4], Christina Warinner[1, 3]

[1]Max Planck Institute for the Science of Human History (Germany), [2]University of the Basque Country (Spain), [3]University of Oklahoma (United States), [4]University of York (United Kingdom)

Archaeological dental calculus is a rich source of ancient DNA, proteins and microremains, providing valuable information about human history and evolution of the oral microbiome. However, calculus deposits are limited in size and often rare in early hominin specimens. Performing multiple extractions (e.g. DNA, proteins, microremains) is often unfeasible for small samples, and current protocols are chemically incompatible. A protocol enabling simultaneous extraction of multiple lines of evidence from a single sample is therefore desirable, for maximising information yield from this finite resource.

Here we present a new unified protocol for DNA and protein extraction from dental calculus. The protocol was systematically tested on samples representing different time periods and anticipated states of preservation, at low (2 mg) and high (10 mg) starting amounts. Overall, the unified protocol recovered 45-71% of the DNA obtained using a standard protocol. Well preserved calculus yielded a lower percentage of DNA through the combined protocol than poorly preserved calculus, possibly due to inefficient cell lysis in the unified protocol, which lacks proteinase K during digestion. Protein yields through the unified protocol were approximately 40% higher than through a protein-only protocol, possibly due to a freezing step in the unified protocol. These results suggest that the unified protocol has sufficiently high biomolecule yields to be preferred over separate extractions, especially for samples with poorer preservation. However, method-associated biases in downstream results must be considered. We conclude with preliminary results on the downstream effect of extraction method on reconstructed microbial communities and recovered proteins.

POA-112

The evolution of the ribosome and its impact on translation dynamics

Khanh Dao Duc[1], Sanjit Batra[1], Nicholas Bhattacharya[2], Yun S Song[1, 3]

[1]UC Berkeley (United States), [2]UC Berkeley (United States), [3]UC Berkeley (United States)

Recent advances in sequencing and structural biology have demonstrated the significant impact of ribosomes on translation dynamics, via complex interactions with the translated sequence. To understand the details of these interactions, a more quantitative understanding of the ribosome structure across different species is needed. In our work, we compile and compare more than 40 recently obtained ribosome structures from cryo-EM data, coming from bacteria, archaea, and eukarya. In particular, we focus on the ribosome exit tunnel, a long narrow structure containing the nascent polypeptide chain. Upon extracting the structure and introducing new metrics for comparison, we show that the evolutionary tree reconstructed from the tunnel geometry is in remarkable agreement with known phylogenetic information based on DNA sequence evolution. Furthermore, we identify domains of conservation, and show that the separation between bacteria and eukarya is mainly due to the insertion of a specific protein during ribosome assembly. Interestingly, by analyzing the charge conservation of aligned sequences of homologous ribosome proteins, we show that it is correlated with the geometric conservation of the tunnel.These results suggest a mechanism of facilitating the movement of charged nascent polypeptide chains through the tunnel during the early stages of translation. Finally, our results are in agreement with 1) the variation of ribosome elongation rates along the mRNA transcript (inferred from ribosome profiling data) and 2) the variation of amino-acid residue charges along the translated protein sequence, suggesting co-evolution of the proteome and the ribosome exit tunnel.

POA-115

Evolution and genetic control of gene expression variability and noise in humans

James J Cai[1, 2]

[1]Texas A&M University (United States), [2]Texas A&M University (United States)

One of the fascinating observations in transcriptomics is that the expression level of the same gene can be highly variable. The variability can be quantified as the magnitude of variance in gene expression among unrelated individuals, or as the gene expression noise at the single-cell level. Our study thus focuses on the two distinct measures of variability in gene expression. First, population-level variability refers to the variance in mRNA abundance of a gene across samples derived from genetically unrelated individuals. Second, single-cell variability---i.e., gene expression noise---refers to the cell-to-cell variability in mRNA abundance measured across single cells of the same sample. We have conducted QTL mapping for population-level variability and identified abundant genetic loci associated with gene expression variance across individuals. We also showed that there is a significant positive correlation between the two measures of variability across genes, suggesting that there might be a common genetic basis responsible for the population-level gene expression variability and gene expression noise. In addition to genetic factors, epigenetic and environmental factors modulating gene expression variability and noise will be discussed.

POA-117

Chromosome-wide stochastic co-fluctuations of gene expression in mammalian cells

mengyi Sun[1], jianzhi zhang[1]

[1]University Of Michigan (United States)

Gene expression is subject to stochastic variation, but to what extent and by which means such variations are coordinated among different genes are unclear. We hypothesize that neighboring genes on the same chromosome co-fluctuate in expression because of their common chromatin dynamics, and indeed detect this signal from allele-specific single-cell RNA-sequencing data of mammalian cells. Unexpectedly, however, the co-fluctuation extends to genes that are over 60 million bases apart. We provide evidence that the long-range effect arises from shared chromatin accessibilities of linked loci attributable to 3D chromatin proximities, which are much closer intra-chromosomally than inter-chromosomally. We further show that genes encoding components of the same protein complexes tend to be chromosomally linked, likely resulting from natural selection for intracellular among-component dosage balances. These findings have implications for both the evolution of genome organization and optimal design of synthetic genomes in the face of gene expression noise.

POA-118

Differential movement of song, morphology, and genes across the black-capped/Carolina chickadee hybrid zone in Missouri

Alana Alexander[1, 2], Mark Robbins[1], Jesse Holmes[1, 3], Robert Moyle[1, 3], A. Townsend Peterson[1, 3]

[1]University of Kansas (United States), [2]University of Otago (New Zealand), [3]University of Kansas (United States)

The black-capped (*Poecile atricapillus*) and Carolina chickadee (*P. carolinensis*) contact zone occurs across a narrow latitudinal band in the USA, ranging from Kansas in the west to New York in the east. Genetic and morphological studies in Pennsylvania (PA) and Ohio (OH) demonstrate northward movement through time of the hybrid zone located within the contact zone, likely in response to climate change. In contrast, analysis of song data in Illinois suggested little movement of the hybrid zone in this area. Here, we focussed on characterising a temporal shift in the understudied western portion of the hybrid zone in Missouri (MO), using 67 samples from 1980 and 92 samples from 2016, including 3 black-capped and 2 Carolina chickadee reference samples obtained from outside the hybrid zone. We compared and contrasted patterns in the morphological, song and genetic (using 11,834 SNPs derived from ddRADseq) datasets for these birds. While the hybrid zone did appear to move northwest across our temporal sample based on genetic markers, it did not move at the rapid rate seen in other areas in the United States (e.g. PA, OH). We suggest this slower movement is due to differences in warming between areas: the temperature change between 1976-2016 in PA (where the contact zone has moved rapidly) is as much as 50% greater than in MO.

POA-119

Structural phylogenetics with confidence

Ashar Malik[1], Anthony Poole[2], Jane Allison[1]

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

Protein structure exhibits greater evolutionary preservation than sequence, due to its closer relationship to function and therefore to phenotype. Comparison of protein structures therefore represents an opportunity to uncover evolutionary signal that escapes conventional sequence-based methods. While structural phylogenetics has been attempted in the past, there is no method to assess the robustness of the inferred relationships. We have overcome this problem by developing a structural analogue to the bootstrap method used with sequence data. We use molecular dynamics simulations to sample alternative conformations for each protein, allowing these to be utilised to assess support for a structure-based phylogeny. We illustrate our method on the Ferritin-like superfamily, showing that structural phylogenetics can successfully extract evolutionary signals from protein structures. Ultimately, phylogenetic analysis of structure in cases where sequence similarity is too low to establish homology will allow us to refine and reconstruct very deep evolutionary relationships by comparing protein structures.

POA-120

High-throughput sequencing using combinatorial profiling

Luisa Teasdale[1], Andreas Zwick[1]

[1]CSIRO (Australia)

Many questions in evolutionary biology only require a small number of genetic loci but sequenced from thousands of samples. This goal is currently prohibited by cost and is technically not practical given current sequencing technologies require individual labelling of samples (i.e. 'indexing'). We have developed a computational approach that enables the simultaneous sequencing of genes for thousands of samples in a single high throughput sequencing run. Instead of individual sequence labels, we use combinatorics to track and match sequences to samples. This approach (termed 'combinatorial profiling') has previously been used for projects such as the detection of rare variants in medical research but has not before been used to retrieve full sequences for every sample encoded. We present a complete computational pipeline to encode and decode the pooled sequences. We also compare different approaches of combinatorial profiling and validate these approaches through computational simulation and sequencing of combinatorially pooled samples. We show that this approach vastly reduces costs when sequencing hundreds or thousands of samples, and is robust to sample drop out. This technique, however, does require a minimum level of sequence divergence. Combinatorial profiling has a vast array of applications and is particularly suited to projects where several loci are needed for thousands of specimens, such as producing reference genetic databases of museum collections.

POA-121

Bacterial K-strategies on evolution experiments under various resource limitation

Takahiro Komori[1], Saburo Tsuru[2]

[1]Osaka University (Japan), [2]The University Of Tokyo (Japan)

All organisms live under resource-limited environments. In an environmental condition with resource shortage, the K-strategy, by which organisms increase their carrying capacity through evolution, can be a more important adaptation strategy than r-strategy, by which organisms increase their growth rate. A previous study experimentally demonstrated K-strategy evolution in a laboratory and proposed that organisms can increase carrying capacity by switching energy generating pathway under carbon source limitation. But it has not been clarified whether there are any other mechanism to realize K-strategy evolution or how often these evolutionary events occur. To answer these questions, we reports consequences of K-strategy evolution under two distinct nutrient limitations, one is single amino acid shortage and the other is nitrite source shortage. We used 96-well microplates to propagate *Escherichia coli* repeatedly. Cultivation medium included limited amount of histidine (amino acid) or annmonia (nitrogen source) and histidine autotroph was cultivated in the former condition. The grown culture with the highest optical density (OD) among the wells were selected and transferred to the next rounds, with dilution to 1 cells per well and we explored its consequences. When annmonium was limited, there was no significant change of OD. On the other hand, when histidine was limited, OD at the saturation point increased. These result indicated frequency of beneficial mutations for carrying capacity depends on nutrient conditions. Whole genome sequencing revealed functional variety of the mutated genes during the fitness-increasing period under histidine limitation, suggesting loose genetic constraints to improve carrying capacity in histidine utilization.

POA-122

Evolution of male courtship songs in the Drosophila nasuta species cluster

Matthew James Nalley[1], Wynn Meyer[1], Doris Bachtrog[1]

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

The courtship song of Drosophila is frequently involved in species recognition and sexual selection. Species from the *Drosophila nasuta* species cluster are widely distributed throughout South-East Asia, and show varying degrees of pre- and post-zygotic isolation. Here, we describe the courtship songs of multiple species in the nasuta clade. We developed a novel method to analyze this group's courtship songs and characterize commonly studied acoustic components such as burst duration (BD), pulse length (PL), pulse number (PN), inter-burst interval (IBI), and inter-pulse interval (IPI). We identify three main song types in this group (sine song, pulse song, and rasps), and we found that the species differ with respect to their usage of these principle song types. We will present a phylogenetic analysis of courtship song in the nasuta clade and preliminary mapping results to determine the genetic basis of evolutionary changes in song among species in this group.

POA-123

Next-Generation Transcriptome Assembly: Strategies and Performance Analysis

Adam Voshall[1, 2], Sairam Behera[3], Xiangjun Li[4, 2], Edgar B Cahoon[2, 4], Etsuko N Moriyama[1, 2]

[1]University of Nebraska-Lincoln (United States), [2]University of Nebraska-Lincoln (United States), [3]University of Nebraska-Lincoln (United States), [4]University of Nebraska-Lincoln (United States)

Accurate and comprehensive transcriptome assemblies lay the foundation for a range of molecular biological analyses. With the arrival of next-generation sequencing technologies it has become possible to acquire the whole transcriptome data rapidly even from non-model organisms. However, the problem of accurately assembling the transcriptome for any given sample remains extremely challenging, especially in species with a high prevalence of recent gene or genome duplications, those with alternative splicing of transcripts, or those whose genomes are not well studied. In order to analyze the performance of transcriptome assemblers, we developed simulation protocols to computationally generate RNAseq data that present biologically realistic problems such as gene expression bias and alternative splicing. Using these simulated RNAseq data, we compared the accuracy, strengths, and weaknesses of nine representative transcriptome assemblers including three genome-guided, four *de novo*, and two ensemble methods using a range of parameters (*e.g.*, k-mers). Finally we developed a new ensemble trasncritome assembly approach that performs superior (with higher accuracy and F_1-score) to any of individual assemblers we compared.

POA-124

On the multiple ways of calculating F_{ST} from DNA sequence polymorphism

Songeun Lee[1], Yuseob Kim[1, 2]

[1]Ewha Womans University (Republic of Korea), [2]Ewha Womans University (Republic of Korea)

Wright's F_{ST} is a simple and widely used measure of genetic differentiation between populations, particularly for detecting loci perturbed by recent selection. However, since it was originally defined for polymorphism at a single locus, applying it to DNA sequence data with multiple polymorphic sites may not be straightforward. In this study, we explore possible ways of defining and calculating F_{ST} for multiple linked SNPs, focusing on whether they ensure consistency and statistical power to detecting outlier while preserving the original aim of measuring population differentiation. Factors influencing the outcome of calculation includes sample size differences between demes, the concept and procedure of "weighting" between demes, minor allele frequencies at SNPs, missing values in NGS data, and whether sampled sequences are phased or not. It is also important whether F_{ST} should be calculated for individual SNPs first and then averaged or calculated after within- and between-deme sequence diversities are estimated from multiple SNPs. We find that the latter method with exclusion of SNPs with low minor allele frequency achieves best performance in detecting outliers caused by positive selection in a two-deme model. Our exploration will contribute to establishing consistency across studies using F_{ST} as a fundamental statistic.

POA-125

Cyclostomes' Hox genes provide insights into the evolutionary origin of temporal colinearity in vertebrates

Juan Pascual-Anaya[1], Iori Sato[1], Fumiaki Sugahara[1, 2], Shinnosuke Higuchi[1], Jordi Paps[3], Ren Yandong[4], Wataru Takagi[1], Adrian Ruiz-Villalba[5], Kinya G. Ota[6], Wen Wang[4], Shigeru Kuratani[1]

[1]RIKEN (Japan), [2]Hyogo College of Medicine (Japan), [3]University of Essex (United Kingdom), [4]Kunming Institute of Zoology, Chinese Academy of Sciences (China), [5]Foundation of Applied Medical Research (FIMA), University of Navarra (Spain), [6]Marine Research Station, Institute of Cellular and Organismic Biology, Academia Sinica (Taiwan)

Hox genes are key developmental genes, which are important for the specification of the body structures along the anterior-posterior axis of animal embryos. Hox genes are generally linked in the same genomic locus, forming cluster, and in most bilaterians are expressed in order along the anterior-posterior main axis of the embryo. In jawed vertebrates (gnathostomes), Hox genes are also expressed in a temporal order, with genes in the 3 prime part expressed earlier than in the 5 prime. This has recently been called whole-cluster temporal collinearity (WTC). However, in some invertebrates, Hox gene are expressed temporally in small subgroups of genes, a phenomenon called sub-cluster level temporal collinearity (STC). However, which condition is ancestral, or whether the Hox genes last common ancestor of vertebrates were expressed according to the WTC remain a mystery. In this study, we have first screened the genome and embryonic transcriptome of the hagfish, a jawless vertebrate, for Hox genes. Hagfish Hox genes expression patterns in the hindbrain are very well conserved with those of jawed vertebrates, with some differences that can account for species-specific differences. Next, we have performed a comprehensive analysis of Hox genes temporal expression in both the lamprey and the hagfish, representing the major lineages of agnathans, and found that these are expressed according to WTC. Our results indicate that WTC is a conserved mechanism of Hox gene expression that have been conserved during the last 500 million year despite drastically different genome evolution and morphological outputs between jawless and jawed vertebrates

POA-126

The Genetic mechanism underlying the ERV abundance in vertebrates

Wanjing Zheng[1], Yoko Satta[1]

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

Endogenous retroviruses (ERVs) are a class of endogenous viral elements in the genome that are highly similar to and can be derived from retroviruses. If a retrovirus invades and integrates into the genome of germ line and subsequently be transmitted vertically, it becomes an ERV. Many LTR retro-transposons are the ERVs that lost extracellular mobility through inactivation or deletion of the gene env that encodes the proteins forming the viral envelope. The expansion of ERVs in the genome of a host species can be promoted by its historical exposures to retroviral infections, and can be suppressed by its resistant mechanism, including the removal. We are interested in revealing the genes that are functionally associated with the ERV abundance in vertebrates. In our study on the functional evolution of the cytoplasmic innate sensors of non-self RNA, RIG-I-like receptors in birds, we found that the evolutionary rate of one of them, RIG-I, was negatively correlated with ERV abundance in birds. To further understand what role the ERVs or retroviruses play in the functional evolution of vertebrate genome, we initiated a genome-wide screen for the genes that might have evolved under association with ERV abundance (mainly in birds).

POA-127

Divergent evolution of olfactory receptor repertoire in New and Old World primates revealed by target capture and massive-parallel sequencing

Ryuichi Ashino[1], Yoshihito Niimura[2, 3], Kazushige Touhara[2, 3], Amanda D. Melin[4, 5], Shoji Kawamura[1]

[1]The University of Tokyo (Japan), [2]The University of Tokyo (Japan), [3]JST (Japan), [4]University of Calgary (Canada), [5]University of Calgary (Canada)

Primates are generally regarded as vision-oriented animals, for which other senses, especially olfaction, are less important. However, this notion is questioned by recent studies and requires further investigation. Here, by employing targeted capture and massively parallel (Next Generation) sequencing methods, we studied the entire gene repertoire of olfactory receptors (ORs) from diverse species of New and Old World primates. New World monkeys (NWMs) are particularly suitable for understanding the interplay of different senses in evolutionary and ecological contexts because of their diversity in color vision and diets. Contrary to a sensory "trade-off" prediction, the proportion of OR pseudogenes was not highest in the howler monkey among NWMs, the sole routine trichromatic NWM genus, compared to other species with dichromacy/ trichromacy polymorphism. The number of intact and defective OR genes appeared to differ only slightly among taxa, whereas gene composition differed considerably among species by repeated gain and loss of OR genes throughout phylogeny. These results depict a feature of active turnaround of OR gene contents in evolution of Anthropoidea primates.

POA-128

Nucleotide divergence between L and M opsin genes in New and Old World primates

Yuka Matsushita[1], Naoko Takezaki[2], Amanda D. Melin[3, 4], Shoji Kawamura[1]

[1]The University of Tokyo (Japan), [2]Kagawa University (Japan), [3]University of Calgary (Canada), [4]University of Calgary (Canada)

In New World monkeys, howler monkeys (*Alouatta*) are the only genus having L and M opsin genes juxtaposed on the X chromosome. We previously reported that they also have a high frequency of L/M hybrid opsin genes, which diverge from normal L and M opsins in absorption spectra. Our previous study on gibbons, an Old World primate taxon, showed that the spectral difference between L and M opsins is maintained by purifying selection against homogenization due to gene conversion in central exons. However, it remains to be elucidated whether such homogenization and purifying selection are also apparent in *Alouatta* and also in other Old World primate taxa. Here, we employ target capture and massively parallel (Next Generation) sequencing methods to examine nucleotide divergence between the L and M opsin genes including all introns in various Old World primates and in *Alouatta*. Contrasting to Old World primates, *Alouatta* showed high nucleotide divergence between the L and M opsin genes throughout the gene region including introns, resembling the pattern of inter-allele difference of the L/M opsin gene in other New World monkeys with polymorphic color vision. Thus, homogenization due to gene conversion is not obvious in *Alouatta*. Nevertheless, *Alouatta* species have L/M hybrid opsin genes at high frequency. This suggests a distinct mode of selection in *Alouatta* from Old World primates allowing color vision variation as in other New World monkeys.

POA-129

Unequal allele frequencies of the L/M opsin gene in New World monkeys

Shoji Kawamura[1], Yuka Matsushita[1], Anthony Di Fiore[2], Filippo Aureli[3, 4], Amanda D. Melin[5, 6]

[1]The University of Tokyo (Japan), [2]University of Texas at Austin (United States), [3]Universidad Veracruzana (Mexico), [4]Liverpool John Moores University (United Kingdom), [5]University of Calgary (Canada), [6]University of Calgary (Canada)

Polymorphic color vision is observed in most New World primates, where females are either trichromatic or dichromatic and males are dichromatic. This is due to allelic variation of the L/M opsin gene on the X chromosome. Our previous study showed that the L/M opsin alleles are maintained by balancing selection in wild populations of capuchin monkeys and spider monkeys. However, knowledge is still wanted on the mode of natural selection acting on these various forms of color vision. We have continued collecting L/M opsin genotype data from wild populations of capuchin monkeys, spider monkeys and woolly monkeys. While equality of frequency is expected among L/M opsin alleles to maximize the number of trichromatic individuals if trichromacy is simply more advantageous than dichromacy, the longest-wave allele was most prevalent with 60-70% frequency irrespective of species or locality. The majority status of the longest-wave allele results in increase of dichromats whose spectral difference between the L/M class and S class opsins are maximum. This could benefit dichromats in chromatic discriminability. On the other hand, this results in decrease of trichromats whose spectral difference between two L/M class opsins is maximum, thus seems disadvantageous for trichromats. However, this trend does not seem to hold for small-bodied species, such as callitrichines, showing different patterns of skewness. The observed unequality of L/M opsin alleles supports the notion that trichromat benefit does not always surpass opposing dichromat benefit and that different alleles could be maintained by different demands among vision types.

POA-131

Speciation Results from Gene Regulatory Evolution

Chia-Hung Yang[1], Samuel V. Scarpino[1, 2, 3]

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

Gene Regulatory Networks (GRNs) describe the inter-dependencies of gene expression and encode information for individual development on the molecular level. These GRNs bridge the gap between inheritance factors and physiological traits, whose evolution therefore becomes a potential candidate for understanding the process of speciation. Here, we study how GRN evolution can rapidly generate reproductive isolation even between lineages from the same ancestors and under identical selection pressure.

We model how GRNs evolve by genetic shuffling during sexual reproduction, drift resulting from finite population sizes, and natural selection on GRN function. Numerical simulations suggest that a population initially comprised of a unique GRN for each individual rapidly achieves 100% survival and then fixes a single GRN. Furthermore, given a constant environment, isolated populations--starting from a common ancestral pool--fix alternative GRNs due to genetic drift. We observe inviable hybrids with varying orders of incompatibilities by interbreeding 100% survival lineages. Even for simple GRNs experiencing generic evolutionary pressures, genetic incompatibilities become enriched for higher-order interactions.

Our work reveals an unconventional mechanism of speciation, which differs from the classical Dobzhansky-Muller model. Intriguingly, we discover that a genotypically rich population experiences a relatively fast, yet stochastic, loss of genotypes during evolution. Such genotypical loss eliminates potential incompatibilities in the offspring genetic pool, resulting in populations no longer affected by natural selection. Consequently, interbreeding these lineages resurrects previously removed incompatibilities and leads to inviable hybrids with a non-negligible chance. Speciation is therefore interpreted as a consequence of drift and GRNs' high dimensionality.

POA-134

Speciation genetics of Pristionchus nematodes

Kohta Yoshida[1], Ralf J. Sommer[1]

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

Although the identification of causal genes for hybrid incompatibility is a fundamental topic in speciation research, there are at least two major limitations. First, in many systems it is difficult to culture species in the laboratory and to perform genetic analyses. Second, closely related species suitable for the analysis of speciation are often difficult to find in nature. While nematodes, in principle, have the potential to overcome these limitations, sufficient number of taxonomic samplings to find closely related species pairs have often not been performed. We focus on *Pristionchus*nematodes, which have a necromenic association with beetles. These nematodes are easily cultured and amenable to many genetic tools, such as genome sequences and CRISPR/Cas9 engineering. Our sampling of beetles around the world resulted in more than 30 new species of *Pristionchus* in the last decade. With these materials, we are running two large projects. First, we use two closely related gonochoristic species and study F2 hybrid progeny that allows screening for genes involved in nascent hybrid incompatibility during on-going speciation. Second, we use the model hermaphrodite *Pristionchus pacificus* and its closely related gonochoristic species. We found recovery of hermaphroditic fertility in backcrosses of hybrids to *P. pacificus* and established multiple lines with introgression of the gonochoristic species into the *P. pacificus*genetic background. We are currently sequencing these lines to screen for genes involved in hybrid sterility in the hermaphrodites.

POA-136

Mitochondrial-Y chromosome interactions and local adaptation to the Mother's Curse

J. Arvid Agren[1], Manisha Munasinghe[2], Andrew G. Clark[1, 2]

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

The maternal inheritance of mitochondrial genes (mtDNA) means that mothers may pass on mutations that are beneficial in her daughters, but deleterious in her sons, a phenomenon called the Mother's Curse. Accumulation of such male-biased mtDNA mutations should lead to selection in males for nuclear compensatory modifiers that alleviate the effect. Both theory and empirical work suggest that this antagonistic co-evolution can result in reproductive isolation through the accumulation of Bateson-Dobzhansky-Muller incompatibilities. Here, we use analytical models, computer simulations, and experiments in *Drosophila melanogaster* to show that mtDNA-sex chromosome interactions are potential hotspots for such incompatibilities. In particular, we demonstrate that the Y chromosome, being strictly paternally is an especially good candidate location for suppressor mutations that that alleviate the deleterious effect of the Mother's Curse. We then use 36 otherwise isogenic *Drosophila melanogaster*strains from five worldwide locations differing only in the geographical origin of their mitochondrial genome and the Y chromosome to experimentally examine this effect. If Y-linked suppressors are important for reproductive isolation, co-evolved locally adapted mtDNA-Y combinations from the same population should outperform novel combinations. Through both phenotypic assays and RNA sequencing we show that mtDNA-Y interactions can on occasion compensate for the Mother's Curse, but depend what aspect of male fitness considered. Overall, this study provides a new framework to understand how genetic transmission asymmetries lead to genetic conflict-driven reproductive isolation.

POA-137

EVOLUTION IN CERIANTHARIA (CNIDARIA), A HOLISTIC VIEW: ASPECTS ON MITOCHONDRIAL DNA, LIFE CYCLE, SYMBIOSIS AND TOXINS

Sergio N Stampar[1], Maximiliano M Maronna[2], Celine Lopes[1], Hellen Ceriello[1], James D Reimer[3], Adam Reitzel[4], Jason Macrander[4], Marymegan Daly[5], Michael Broe[5], Mei Lin Neo[6], Nicholas Wei Liang Yap[6], Marcelo V Kitahara[7, 8], Alvaro E Migotto[8], Andre C Morandini[2, 8]

[1]Universidade Estadual Paulista (UNESP), FCL/Assis (Brazil), [2]Instituto de Biociencias, Universidade de Sao Paulo (Brazil), [3]University of the Ryukyus, Faculty of Science (Japan), [4]University of North Carolina at Charlotte (United States), [5]The Ohio State University (United States), [6]National University of Singapore (Singapore), [7]Universidade Federal de Sao Paulo (Brazil), [8]Universidade de Sao Paulo (Brazil)

More than 200 years of infrequent studies on Ceriantharia morphology created a classification system within the group, and for the last century it was classified as an order of anthozoans, divided into three families. This group of tube anemones are one of the least understood groups of cnidarians where its phylogenetic affinities remains controversial, even considering current methods and molecular markers used to infer phylogeny. In fact, recent phylogenetic studies have established Ceriantharia as a subclass sister to all other Anthozoa. This study integrates and review several issues, relating to ceranthid life cycles, symbioses, mitochondrial genomes and toxin profiles, to assess the traditional classification system within Ceriantharia. Our results indicate that previously established subgroups are not coherent and artificial groupings are likely derived from misinterpretations of homology among morphological characters. In turn, the symbiosis data indicate that despite the tubes variety there are no divergence among groups associated with Ceriantharia, so the tube structure can be a factor to contrast subgroups in Ceriantharia which support data concerning life cycles. Furthermore, our results suggest differences among the structures of the mitochondrial genome for groups with distinct life cycles (cerinula or planula). Finally, data on toxins also appear to be consistent with this divergence between two large subgroups of Ceriantharia. Altogether, reconstructions based on morphology and/or molecular markers (non-genomic approaches) may highlight inconsistent patterns and even non-homologous characters Our data demonstrate that Ceriantharia contains much deeper diversity than has been assumed, and from this phylogenetic framework a new classification is proposed.

POA-140

Transcriptomic analyses of sex chromosome meiotic drive in Drosophila

Sung-Ya Lin[1], Shu Fang[2], Chau-Ti Ting[1, 3], Catherine Montchamp-Moreau[4]

[1]Academia Sinica & National Taiwan University (Taiwan), [2]Academia Sinica (Taiwan), [3]National Taiwan University (Taiwan), [4]CNRS, IRD, Paris-Sud University and Paris-Saclay University (France)

Meiotic drive is the non-Mendelian transmission of alleles or chromosomes during gametogenesis. It can trigger intragenomic conflicts with impact on genome evolution and speciation. *Sex-ratio* (*SR*) meiotic drives, favoring the transmission of X relative to Y chromosome, lead to strong female-biased progeny of affected males. *SR* meiotic drives have been reported in several independent lineages, but the molecular mechanism remains largely unclear. In *Drosophila simulans*, it is known that many genes are involved in the Paris *SR* system, but only *HP1D2*, a member of the Heterochromatin Protein 1 (HP1) gene family, was identified. To systematically identify other *SR*-related genes, the transcriptomic differences of testicular expression between the wild-type and *SR* strains were compared. Here we show a list of candidate genes which showed differential expression between wild-type and *SR* strains. Among these genes, there are more up-regulated genes than down-regulated genes. Also, these genes are highly enriched in genes associated with chromatin assembly, RCAF complex, protein heterodimerization, and nucleosomal DNA binding. The study will shed insights into the possible players involved in the sex chromosome meiotic drive and improve our understanding on the molecular and cellular bases of meiotic drive.

POA-141

Entensive and fine-grained introgressions in hybrid zones suggest genome-wide adaptive differences between mangrove species

Xinfeng Wang[1], Ming Yang[1], Zixiao Guo[1], Ziwen He[1], Suhua Shi[1]

[1]Sun Yat-sen University (China)

The extent and significance of interspecific gene flow to species evolution has long been debated. Patterns of gene flow in contact zones vary across the genome and studies of such patterns allow us to quantify the genetic difference and understand the tempo and mode of speciation. Here, we addressed this question using a population genomics approach with 43 resequenced genomes from 10 populations of two Mangrove Rhizophora sister species, R. mucronata and R. stylosa. The two species diverged approximately 2.94 million years ago and are morphologically distinct. To investigate the genomic architectures underlying their divergence, we sampled these two species both in their allopatric ranges and sympatric hybrid zones. Combing phylogenetic methods and the D statistics, we find pervasive interspecific gene flow between sympatric populations in hybrid zone Daintree River, Australia, with introgressed regions fine-grained and scattered throughout the genome. Highly divergent regions resistant to gene flow identified with fixed difference measure harbor genes involved in stress response and flower development, pointing to species-specific adaptation. Given species distinction in morphology regardless of geographical context and the comparison of genomic divergence in sympatry vs allopatry populations, our results indicate that species integrity can be maintained in face of substantial gene flow via divergent selection on numerous localized genomic regions. Thus, the formation of nascent species may involve a rather diffuse genomic architecture with genome-wide adaptive differences underlying species differentiation.

POA-142

Oviposition decision-making behavior in the specialist Drosophila sechellia

RumiKondo[1], Ami Hagihara[1], Yoshiko Iwamoto[1], Yukiko Oishi[1]

[1]Ochanomizu Univsersity (Japan)

Selecting a suitable oviposition site is an important reproductive need for females. Host shift could be an important factor of early speciation, however, genetic and neural mechanism underlying oviposition decision making behavior is not known. We examined the oviposition behavior of *D. sechellia* (*D. sec*), a specialist that exclusively oviposit on the toxic noni fruit. We found that females check the oviposition substrate before each oviposition for decision-making. Their egg is retained either until oviposition conditions are met or until the embryo reaches hatching time. Front tarsi are important to sense oviposition substances and adaptation occur after loss of tarsi. Interspecific hybrid with *D. simulans* showed similar oviposition choice as *D. sec*indicating that *D. sec* 's noni choosing behavior is a dominant trait.

POA-143

Sequencing a Large Repetitive Reproductive Gene in Abalone

Alberto Marcos Rivera[1], Willie J Swanson[1]

[1]University of Washington (United States)

The molecular mechanisms mediating sperm-egg interactions remain poorly understood. However, a common feature of genes involved in reproduction is their rapid evolution. This rapid evolution may contribute to reproductive isolation between populations. The seven species of California abalone have overlapping habitats and mating seasons, but produce few wild hybrids due to species-specific fertilization. Two major abalone interacting reproductive proteins are lysin and VERL (Vitelline Envelope Receptor for Lysin). Lysin is secreted by sperm, and VERL is a gigantic ~1 MDa molecule that is part of a glycoprotein egg coat which surrounds the egg. While lysin is well characterized on a genetic and biochemical level, only one VERL sequence from one abalone species has been determined. VERL has large array (~10kb) of repeated 500 bp ZP-N modules. This large repetitive region has hindered many sequencing efforts (both Sanger and Illumina) due to difficulties with repeat assembly. I am characterizing VERL sequence and repeat number variation, which may have functional consequences for species-specific fertilization since lysin is experimentally known to bind the repeat array. I am exploring multiple strategies for amplifying and isolating VERL to sequence the gene with PacBio technology, which produces long reads capable of spanning the entire repeat array. VERL will be sequenced from multiple individuals in the seven California abalone, which will reveal intraspecific and interspecific variation in repeat number and sequence content. This analysis could illuminate the relationship between VERL evolution, species-specific fertilization, and speciation and provide molecular insights into the rapid co-evolution of sperm-egg interactions.

POA-144

When sex-specific hybrid incompatibility does not make sense: Parent-of-origin-specific transcriptome analysis of a pseudohaplodiploid male hybrid

Andres De la Filia[1], Dominik Laetsch[1], Laura Ross[1]

[1]University Of Edinburgh (United Kingdom)

Reduced hybrid viability frequently arises due to incompatibilities between the genomes of parental species and contributes to reproductive isolation. Certain reproductive systems, however, are expected to protect against such deleterious interactions in males. Among these is paternal genome elimination (PGE), a pseudohaplodiploid system which has independently evolved several times in arthropods. PGE males are diploid but transmit maternally-inherited chromosomes only, while paternal homologues are excluded from sperm. Genetic conflict between parental alleles has been invoked to explain PGE evolution, predicting an arms race between paternal and maternal genomes over transmission to following generations. Consequently, in some PGE species such as the mealybug *Planococcoccus citri*, paternal chromosomes are heterochromatinized during development and are thought to remain silenced so that possible paternal genome responses to resist elimination are prevented. This implies that males should only express maternal alleles and therefore escape hybrid incompatibility, but extreme male-biased mortality is recurrently observed among offspring of crosses between *P. citri* and the closely related *P. ficus*. In this study, we present a parent-of-origin allele-specific transcriptome analysis of hybrid mealybugs to provide a first genome-wide estimation of how completely paternal genomes are silenced under PGE. We show that expression is globally biased towards the maternal genome but detect activity of paternal chromosomes in somatic and reproductive tissues. Our results provide a first direct insight into gene expression patterns in PGE males and offer a solid ground to further explore the role of disruption of paternal genome silencing as a postzygotic barrier between PGE species.

POA-146

The population history of the Green Monkeys inferred from the X to autosome diversity ratio

Moises Coll Macia[1], Mikkel Heide Schierup[1]

[1]Aarhus University (Denmark)

The X chromosome evolves differently from the autosomes because multiple factors affect them asymmetrically. The diversity in the X chromosome is influenced by the fact that its effective population size (Ne) is 3/4 of an autosome. However, other factors, such as selection in genes or sex-specific generation times, makes the levels of diversity depart from the 3/4 expectation. In this study, we compare the diversity levels of the X chromosome and autosomes (ratio X/A) of 163 vervet monkeys (genus Chlorocebus) to understand their population history. These species are highly diverse and widely spread through Africa, divided in 6 species (C. sabaeus, C. tantalus, C. aethiops, C. cynosures, C. hilgerti and C. pygerythrus), although we can find some populations in certain Caribbean islands, where they were introduced in the colonization period. As it has been shown in humans, we observe a strong correlation of the ratio X/A with the amount of diversity in the different populations; Caribbean populations, which went through multiple bottlenecks and have smaller Ne, have lower ratios than the African populations with bigger Ne. Furthermore, X/A ratio increases with distance from the nearest gene, due to the strong selective pressure in the X chromosome. These high (> 3/4) values of the X/A rate found far from genes can be explained with demographic models that accounts for sex-specific demographic events. Finally, lower levels of shared polymorphisms in the X compared to the autosomes are found between species that had ancient hybridization due to hybrid incompatibility.

POA-147

High-throughput investigation of the species-specific binding of coevolving abalone reproductive proteins

Jolie A Carlisle[1], Willie J Swanson[1]

[1]University of Washington (United States)

Reproductive proteins mediating sperm-egg interactions are characterized by accelerated evolution. Strong selective pressure to maintain successful fertilization paired with differences in optimal male/female reproductive strategies can promote arms race dynamics (sexual conflict) that drive the rapid coevolution of protein-protein interactions (PPIs). This rapid coevolution can create molecular boundaries to hybridization and contribute to speciation. The marine gastropod abalone is a classic model for studying sperm-egg interactions. Despite having overlapping habitats and breeding seasons, natural hybridization events between the seven species of sympatric Californian abalone are rare. During abalone fertilization, sperm lysin dissolves the vitelline envelope of the egg by binding repetitive domains of the vitelline envelope receptor for lysin (VERL). Vitelline envelope dissolution by lysin is species specific, presumably due to the rapid coevolution of lysin and VERL creating differences in conspecific and heterospecific binding affinities (K_D). However, the functional consequences of lysin and VERL diversity on binding affinity has not been determined despite its potential importance as a molecular block to hybridization. Using Yeast Synthetic Agglutination (YSA), a novel, high-throughput approach for quantitatively measuring PPI binding affinities, I am measuring the K_D between lysin and its multiple VERL binding sites for all conspecific and heterospecific pairs. This unprecedentedly thorough functional characterization of a PPI mediating gamete recognition is a unique opportunity to provide experimental support for the hypothesis that the rapid coevolution of reproductive proteins contributes to the maintenance of species boundaries and, perhaps, species formation.

POA-148

Laboratory evolution reveals distinct genetic and phenotypic adaptation in two parallel lineages originating from nearly identical enzyme progenitors.

Charlotte Miton[1], Eleanor Campbell[2], Colin Jackson[2], Nobuhiko Tokuriki[1]

[1]University of British Columbia (Canada), [2]Australian National University (Australia)

A long-standing question in evolution asks whether adaptation relies on stochastic events, or rather follows a deterministic path, repeatedly imposed by specific molecular or environmental constraints. Haldane postulated in 1932 that similar selection pressures exerted on close relatives would repeatedly drive adaptation to similar genotypic and phenotypic solutions. Numerous studies since reported striking examples of evolutionary convergence or parallelism at the organismal, molecular or genetic levels, where similarities emerged multiple times independently. Thus, the systematic comparison of evolutionary trajectories, whether natural or in the laboratory, should reveal the common underlying molecular mechanisms shaping these trajectories. How closely related, genetically speaking, should orthologous proteins be to trigger a repeated evolutionary outcome? To explore these questions, we 'replayed' the evolution of an enzyme, PTE phosphotriesterase, which was previously evolved to hydrolyze an arylester substrate. Starting from a phenotypically similar immediate neighbor that differs at the genotypic level by a unique amino acid mutation (S254 instead of R254), and is, we repeated the laboratory evolution under identical experimental conditions. While a detailed kinetic, mutational and structural characterization of the evolutionary intermediates revealed genotypic similarities between the trajectories, we observed marked phenotypic differences in enzyme fitness and structure. An analysis of mutational tolerance in the mutants demonstrated that extensive epistasis prevents the fixation of the S254 trajectory mutations on the original R254 variants and reciprocally, disclosing the emergence of genetic incompatibility and contingency during adaptation. Our findings emphasize the difficulty of predicting evolutionary outcomes, even when evolution originates from virtually identical starting points.

POA-149

Do selfish plasmids drive evolution in budding yeast?

Michelle Hays[1, 2], Janet Young M[1], Harmit Singh Malik[1]

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

Selfish genetic elements exploit host cell machinery for their own reproduction, thereby reducing the fitness of their hosts. How do host cells evolve to defend themselves against these genetic parasites? We hypothesize that the 2-micron plasmid, naturally found in many budding yeasts, is one example of a selfish genetic element. The plasmid must hijack host cellular machinery to replicate and segregate its genome, and confers a 1-3% fitness cost to the host in the process. There is no known benefit to the host for harboring the 2-micron, and indeed the only genes found on the plasmid itself promote plasmid stability and high copy number maintenance.

We have identified several natural *Saccharomyces cerevisiae* isolates that naturally do not harbor the 2-micron plasmid. When we reintroduce the 2-micron plasmid, these strains rapidly and reproducibly lose the plasmid once again, suggesting this plasmid loss is a heritable trait. Furthermore, this plasmid loss phenotype is a genetically dominant trait. Taken together, we hypothesize these strains may have evolved a restriction factor targeting the 2-micron plasmid. We are currently using a quantitative trait locus mapping strategy to determine the host factors underlying this restriction phenotype. This genetically tractable, yet naturally occurring, host-parasite model will provide insight into how host cells can evolve to fight genetic parasites, and may reveal adaptation in replication or segregation systems: two processes that are not normally expected to evolve under antagonistic genetic pressures.

POA-150

The role of rapidly evolving fertilization genes in threespine stickleback reproductive isolation

Emily E Killingbeck[1], Damien B Wilburn[1], Gennifer E Merrihew[1], Michael J MacCoss[1], Catherine L Peichel[2], Willie Swanson J[1]

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

Sperm-egg compatibility is essential to the evolutionary success of any sexually reproducing organism, yet the proteins that mediate gamete interactions often evolve at extraordinary rates. In threespine stickleback fish (*Gasterosteus aculeatus*), reproductive isolation is common in many recently derived populations throughout the Northern Hemisphere, but the precise biochemical mechanisms driving this isolation are unknown. Stickleback are classic models of molecular adaptation and speciation, and while rapidly evolving reproductive proteins are probable candidates underlying this reproductive isolation, they remain unexplored in this model evolutionary system. Tandem mass spectrometry was used to characterize the secreted proteomes of stickleback eggs from Lake Union, Washington. High-resolution mass spectra were acquired, with homologs of common vertebrate egg proteins identified. Evolutionary rate analysis (d_N/d_s) of these homologs across fish from superorders within the Teleosts indicates positive selection. In contrast to mammals, the genes encoding the major egg proteins are tandemly duplicated in the stickleback genome. Such duplications provide a substrate for diversification that can drive rapid evolution, and suggest a potential mechanism underlying sexual conflict within stickleback populations and ultimately speciation.

POA-155

Identifying Loci Under Selection Against Gene Flow in IM Models using linked loci

Sadoune Ait Kaci Azzou[1], Daniel Wegmann[1], Anja Westram[2]

[1]University of Fribourg (Switzerland), [2]University of Sheffield (United Kingdom)

Migration is a major evolutionary force homogenizing evolutionary trajectories of populations by promoting the exchange of genetic material. The influx of new genetic material may be facilitated or hampered by selection. While the rate of migration is the same for all loci, at least on autosomes, selection is likely to vary across loci. As a consequence, and due to genetic drift, the effective gene flow is expected to vary along the genome in organisms undergoing recombination. Identifying these loci is key to understand the role of selection in shaping phenotypic differences between populations. Here we introduce a new coalescent based method to infer locus-specific migration rates jointly with demographic parameters affecting all loci under realistic isolation with migration models. Our method is computationally efficient because the same sampled genealogies can be used for all loci and because most calculations can be done analytically in our setting. Importantly, our approach accounts for linkage between sites through autocorrelation in locus-specific migration rates. This results in an increased power to detect deviations from back-ground effective migration rates, which makes our method particularly suited to identify selection on complex traits with often rather weak selection on individual loci. We illustrate the power of our method by inferring locus-specific effective migration rates between ecotypes of the coastal snail Littorina saxatilis that maintain strong phenotypic differences despite intense gene flow.

POA-156

The genetic affinities of southern Africa hunter-gatherers prior to the arrival of farming and pastoralist populations.

Mario Vicente[1], Peter Ebbesen[2], Carina Schlebusch[1]

[1]Uppsala University (Sweden), [2]University of Aalborg (Denmark)

Southern African Khoe-San groups collectively refer to forager (San) and herder (Khoekhoe) communities who all speak Khoisan languages. The Khoisan linguistic phylum includes various 'click' language families that are linguistically unrelated to each other. Throughout history, Khoe-San communities have been largely isolated until the arrival of pastoralists and farmers in the region starting approximately 2,000 years ago. Assessing Khoe-San regional genetic prehistory has been challenging due to genetic contribution from immigrant farmers and pastoralists into their original gene pool, obscuring much of the past population affinities and gene-flow among these autochthonous communities.

In this study we re-evaluate a combined genome-wide dataset of previously published southern Africa Khoe-San populations in conjunction with novel data. After excluding exogenous genome segments originating from interaction with immigrant farmers, herders and colonists, the genetic diversity of 20 Khoisan-speaking groups fitted an isolation-by-distance model well, reflecting the southern African landscape. Even though isolation-by-distance explains most affinities between autochthonous groups, signals of contact and admixture between different Khoe-San groups could be detected. We found evidence of admixture between geographically distant Khoe-San groups and some of these connections are also reflected in linguistic relatedness.

POA-157

Fitting an isolation-migration model to MSMC estimates to infer population sizes and migration rates over time

Ke Wang[1], Stephan Schiffels[1]

[1]Max Planck Institute for the Science of Human History (Germany)

Given the rising number of complete human genome sequences worldwide, many novel methods have been developed for studying population history. The Multiple Sequentially Markovian Coalescent (MSMC, Schiffels and Durbin 2014) is a popular tool for analysing the ancestral relationship between populations from genome sequence data. Estimates of coalescence rates within and across populations as measured by MSMC can be used to infer population sizes and divergence times back in time. Particularly, the midpoint of the relative cross coalescence rate (CCR), calculated as the time point when the relative CCR hits 0.5, is often used as a heuristic estimate for the divergence time. Here we investigate the robustness of this approach with simulations of various demographic and migration scenarios. We find that the CCR-midpoint-approach can lead to wrong estimates of the true divergence time if the two populations have different demographic history since the split time, or if migration and admixture occured. We propose an alternative approach to interpret MSMC results, by fitting an explicitly structured Isolation-migration model to the inferred coalescence rates from MSMC. We show that this approach improves interpretability and robustness of MSMC results. We show an application of our approach by re-analysing dozens of world-wide populations from the SGDP (Mallick et al. 2016) project.

POA-158

Gene flow in space and time: differential patterns among syntopically occurring species of the water flea genus Daphnia

Anne Thielsch[1], Klaus Schwenk[1]

[1]University of Koblenz-Landau (Germany)

Variation in the mode of reproduction (e.g., alternation of asexual and sexual reproduction) might determine population genetic structure and influence patterns of gene flow among populations. Cyclic parthenogenesis, common among zooplankton organisms, is largely regulated by environmental factors: during favourable conditions, amictic individuals reproduce asexually, while deteriorating environmental cues induce sexual reproduction, which results in dormant stages that enable dispersal in space (mainly via waterfowl) and time (biological archives in lake sediments). Although zooplankton organisms are regarded as effective dispersers, genetic studies often show that the influence of gene flow among spatially distributed populations is negligible. In addition, information on the genetic influence of individuals hatched from dormant stages on active populations is rare. Our aim was a combined analysis of the population genetic structure of two species from the genus *Daphnia* to evaluate the influence induced by the two dispersal modes available in cyclic parthenogens. Therefore, we assessed spatial (37 European localities) and temporal changes (monitoring of one lake) in genetic diversity using microsatellite and mitochondrial DNA markers. Interestingly, *D. galeata* and *D. longispina* differ strongly in their spatial population genetic structures across central Europe. While *D. galeata* populations are genetically rather homogenous, suggesting a certain connectivity of populations, *D. longispina*populations are highly differentiated across localities. Our first results on the temporal patterns also indicate differences among species, corroborating the results on the spatial scale, as the differentiation of *D. longispina* populations is increasing more rapidly over time than the differentiation of *D. galeata* populations.

POA-159

Analysis of Pattersons D and fD for Detecting Regions of Introgression

Lesly Lopez[1], Emily Jane McTavish[1], Emilia Huerta-Sanchez[1]

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

Neanderthals interbred with early modern humans, resulting in Neanderthal alleles in the modern-day human genome. Evidence of introgression into non-African populations has been found. One method for detecting introgression genome wide is the Pattersons D statistic, or ABBA-BABA statistic. However, it is found to produce false positive when applied to small genomic windows or to regions of low genetic diversity. The f_D statistic was derived from the Pattersons D to address these issues. We find that there are problems with using Pattersons D and f_D in areas with a low number of informative sites. The f_D statistic and f_D statistic only use sites where the donor population has the derived allele. Tracts of introgressed DNA which contain derived alleles from the donor population also include ancient alleles that are not being accounted for in these tests. These include ancient alleles that the recipient population shares with the donor population, but that are rare in the population where introgression is unlikely to have occurred. We predict that integrating these sites into the statistics will provide a stronger test to detect signs of introgression. By investigating the methods that are being used to identify introgressed genomic regions, we can recognize their limitations and reduce the instances when false positives are given.

POA-160

Dissecting the Pre-Columbian legacy along the Andes-Amazonia divide

Guido Alberto Gnecchi Ruscone[1], Stefania Sarno[1], Sara De Fanti[1], Cristina Giuliani[1], Tullia Di Corcia[2], Chiara Barbieri[3], Patrizia Di Cosimo[5], Ricardo Fujita [4], Antonio Gonzalez Martin [6], Zelda Alice Franceschi [7], Olga Rickards[2], Donata Luiselli [5], Marco Sazzini[1], Davide Pettener[1]

[1]Universita di Bologna (Italy), [2]Universita di Roma Tor Vergata (Italy), [3]Max Planck Institute for the Science of Human History (Germany), [4]Universidad de San Martin de Porres (Peru), [5]Universita di Bologna (Italy), [6]Universidad Complutense de Madrid (Spain), [7]Universita di Bologna (Italy)

Most of the recent population genomics studies on present-day and ancient Amerindians have been mainly focused on disentangling the first migration routes that brought modern humans into North America. Conversely, many aspects related to subsequent diffusions throughout the continent, especially as concerns the dynamics of the peopling of South America, still need to be addressed from a genome-wide perspective.

In particular, due to both environmental and historical conditions the Andes and Amazonia are supposed to have shaped different South American gene pools since the first peopling processes. Nevertheless, several aspects regarding the origins and interactions between Andean and Amazonian populations are is still unresolved. One of the reasons for this lack of information is that the study of Pre-Columbian history based on modern populations' DNA legacy is extremely puzzling due to the confounding effects of recent and extensive European and African admixture in present-day South Americans.

To explore the genetic sub-structure of Andean and Amazonian populations, as well as to infer their demographic history and different ancestral contributions, we generated high-resolution genome-wide data for 200 newly sampled individuals belonging to 10 ethnic groups from Peru, Bolivia and Argentina. The limited genomic traces of recent European and African admixture in the considered populations allowed us to perform fine haplotype-based analyses that revealed new insights into the local patterns of differentiation and admixture that characterized the history of South American populations.

POA-162

Estimated Effective Migration Surfaces with Rare Variation and application to Anopheles gambiae

Daniel N. Harris[1, 2, 3], Timothy D. O'Connor[1, 2, 3]

[1]University of Maryland School of Medicine (United States), [2]University of Maryland School of Medicine (United States), [3]University of Maryland School of Medicine (United States)

Reconstructing migration histories across geographic space is essential to understand regions' demographic history. Estimated Effective Migration Surfaces (EEMS), (Petkova et al. 2016, Nature Genetics), has primarily been used with common variation which are biased towards migration events in the distant past. Therefore, we aimed to improve EEMS's sensitivity to recent demographic events by optimizing EEMS to function with rare genetic variation. We compared two additional distance metrics to the pair-wise difference (PWD) used by EEMS: 1) cosine similarity and 2) weighted PWD which assigns a higher priority to rare variants. We evaluated each of these distances within the EEMS framework with simulations introducing a migration barrier 0 to 1000 generations ago. EEMS computed with all or common variants (MAF greater than or equal to 1%) performed better at distinguishing the migration barrier from 1000 to 700 generations ago with all distance metrics. However, from 600 until 10 generations ago, EEMS run on only rare variants (MAF < 0.5%) was better able to identify the migration barrier with the weighted PWD approach performing the best. We then applied EEMS to 444 western African mosquitoes (*Anopheles gambiae*) from the Ag1000G Consortium Phase 1. We find differences in migration patterns when we model rare and common variants in EEMS, with an additional putative migration barrier, which have implications for understanding the timing and spread of this malaria vector in Africa.

POA-163

Maximum Likelihood Implementation of an Isolation-with-Migration Model with Three Species for Testing Speciation with Gene Flow

Tianqi Zhu[1], Ziheng Yang[2], Daniel Dalquen[2]

[1]Academy of Mathematics and Systems Science, Chinese Academy of Sciences (China), [2]University College London (United Kingdom)

We develop a maximum likelihood (ML) method for estimating migration rates between species using genomic sequence data. A species tree is used to accommodate the phylogenetic relationships among three species, allowing for migration between the two sister species, while the third species is used as an out-group. A Markov chain characterization of the genealogical process of coalescence and migration is used to integrate out the migration histories at each locus analytically. Our implementation can accommodate tens of thousands of loci, making it feasible to analyze genome-scale data sets to test for gene flow. We calculate the posterior probabilities of gene trees at individual loci to identify genomic regions that are likely to have been transferred between species due to gene flow. We conduct a simulation study to examine the statistical properties of the likelihood ratio test for gene flow between the two in-group species and of the ML estimates of model parameters such as the migration rate. Inclusion of data from a third out-group species is found to increase dramatically the power of the test and the precision of parameter estimation. We compiled and analyzed several genomic data sets from the Drosophila fruit flies. Our analyses suggest no migration from D. melanogaster to D. simulans, and a significant amount of gene flow from D. simulans to D. melanogaster, at the rate of about 0.02 migrant individuals per generation. We discuss the utility of the multispecies coalescent model for species tree estimation, accounting for incomplete lineage sorting and migration.

POA-168

Transgenic rhesus monkeys carrying the human MCPH1 gene copies show human-like neoteny of brain development

Bing Su[1], Lei Shi[1], Xin Luo[1], Yongchang Chen[2], Cirong Liu[3], Min Li[1], Hong Wang[2], Yanjiao Li[2], Yuyu Niu[2], Yundi Shi[4], Martin Styner[4, 5], Qiang Lin[1], Jin Jiang[1], Weizhi Ji[2]

[1]Kunming Institute of Zoology, Chinese Academy of Sciences (China), [2]Institute of Primate Translation Medicine, Kunming University of Science and Technology (China), [3]National Institute of Neurological Disorders and Stroke, National Institutes of Health (United States), [4]University of North Carolina (United States), [5]University of North Carolina (United States)

Brain size and cognitive skills are the most dramatically changed traits in humans during evolution, and yet the genetic mechanisms underlying these human-specific changes remain elusive. Here, we successfully generated transgenic rhesus monkeys carrying human copies of MCPH1, a key gene for brain development with fixed human-specific mutations during evolution. Brain image and tissue section analyses indicated an altered pattern of neural cell differentiation, resulting in a delayed neuronal maturation and neural fiber myelination in the transgenic monkeys, similar to the known brain developmental delay (neoteny) in humans. Further brain transcriptome analysis showed a marked expression suppression of neuron-differentiation-related genes, providing a molecular explanation to the observed brain developmental delay of the transgenic monkeys. The presented data represents the first attempt to experimentally interrogate the genetic basis of human brain origin using a transgenic monkey model, and it values the use of nonhuman primates in understanding human unique traits.

POA-170

No evidence for recent selection at *FOXP2* among diverse human populations

Elizabeth G Atkinson[1, 2, 3], Amanda J Audesse[4], Julia A Palacios[5, 6], Dean Bobo M[1], Ashley E Webb[4], Sohini Ramachandran[5], Brenna M Henn[1, 7]

[1]Stony Brook University (United States), [2]Massachusetts General Hospital (United States), [3]Broad Institute of Harvard and MIT (United States), [4]Brown University (United States), [5]Brown University (United States), [6]Stanford University (United States), [7]University of California, Davis (United States)

FOXP2, a textbook gene initially identified for its role in human speech, contains two nonsynonymous substitutions in exon 7 that are uniquely derived in the human lineage. This led to speculation that *FOXP2* played a key role in the development of modern language, as neighboring intronic variation appeared compatible with a recent selective sweep in the ancestors to all humans. Evidence for a recent selective sweep in *Homo sapiens*, however, is now at odds with the presence of these substitutions in archaic hominins. Here, we comprehensively reanalyze *FOXP2* in hundreds of genomes from globally distributed modern human populations to test a hypothesis of recent selection. Neither positive nor balancing selection statistics support a recent selective event in *FOXP2*. The original signal appears to have instead been a result of sample ancestry composition. Our tests do identify an intronic region containing a transcription factor binding site that affects FOXP2 expression and that is unusually enriched for highly-conserved polymorphisms. Strong evolutionary constraint among taxa but variability within Homo sapiens is compatible with a modified functional role for this intronic region, such as a recent loss of function in humans. We conducted further follow-up on this intron via Sanger sequencing, RNAseq on human prefrontal cortex, and RT-PCR in immortalized human brain cells and find that it is expressed in relevant tissues. Our findings represent a major revision to the understanding of the selective history of *FOXP2*, a gene regarded as vital to the evolution of the human species.

POA-171

The Role of CNTNAP2 in the Evolution of the Human Synapse

Frances St George-Hyslop[1, 2], Frederick J. Livesey[1], Toomas Kivisild[2]

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

There are significant morphological and developmental differences between the synapses of humans and other primates. Human cortical neurons have longer dendrites and more elaborate dendritic branching than chimpanzees or macaques. These differences may have contributed to the higher cognitive abilities of *Homo sapiens*. Contactin-Associated Protein-like 2 (*CNTNAP2*) was selected as a candidate gene underlying species differences in synapse development and function, based on the presence of six Human Accelerated Regions (HARs) within the gene, and previous work in mice showing knockdown of *CNTNAP2* reduces dendritic branching and dendritic spine density *in vitro* and *in vivo*. Using human stem cell-derived forebrain neurons, the effects of *CNTNAP2* over-expression and loss-of-function on dendrite and synapse development are being investigated. Parameters being measured include dendrite length, dendritic branching, and dendritic spine density. In addition to this, we are also studying the function of the six HARs within the introns of *CNTNAP2*. These sequences are hypothesized to be enhancers that promote the expression of the gene, and thus the development of dendrites and dendritic spines in the human cortex. We are using a luciferase enhancer assay to study the gene regulatory potential of these six HARs (and their orthologs from other primates) in human and mouse experimental systems. This work may shed light on the longstanding question of why humans are unique as a species. Crucially, it will also inform on the causes - and potential treatments - of neurological diseases associated with mutations in *CNTNAP2*, including Autism Spectrum Disorder and Intellectual Disability.

POA-173

Molecular signatures of autism in prefrontal cortex

Ilia Kurochkin[1], Ekaterina Khrameeva[1, 2], Anna Tkachev[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]Max Planck Institute for Evolutionary Anthropology (Germany), [4]CAS-MPG Partner Institute for Computational Biology (China)

Autism Spectrum Disorder (ASD) is a group of complex diseases of brain development connected with abnormalities in brain functions, including socialization and communication: key behavioral capacities that separate humans from other species. This suggests that ASD may involve alterations in evolutionarily novel, human-specific developmental processes. To test this, we sequenced genomes of 24 autism patients, and found an enrichment of SNPs with differences in allele frequency between autism samples and unaffected controls in metabolic pathways. To further study the effect of these SNPs on metabolic pathways, we measured concentrations of polar compounds in the prefrontal cortex of 73 autism patients and control individuals, as well as 80 chimpanzees and macaques, from natal to adult age. We show that 202 metabolites with concentration differences between autism cases and unaffected controls are strongly overrepresented in glutathione, purine and pyrimidine metabolism pathways that were previously reported to be associated with ASD. Also, we find that metabolites with human-specific concentration changes are enriched in purine and pyrimidine metabolism pathways associated with ASD, suggesting that a recently evolved developmental pattern of metabolite concentrations unique to humans might be altered in autism.

POA-175

Towards the most comprehensive lncRNA catalogue and in-depth evolutionary analysis of human long non-coding RNAs

Amin Saffari[1], Peter Stadler[2, 3, 4], Katja Nowick[1]

[1]Freie Universitat Berlin (Germany), [2]Leipzig Universitat (Germany), [3]Institute (United States), [4]Max Planck Institute (Germany)

A majority of the human lncRNAs are expressed in the brain, and multiple lines of evidence have linked them to pivotal brain functions. Although several databases exist for lncRNAs, there is still a huge lack in understanding their role in the nervous system.

Here, we use in-house and publicly available RNA-seq data from different regions of the human brain to create a complete catalogue of lncRNAs, including sequence annotation, structural information and evolutionary changes. Our gene models have been predicted using a novel transcript assembler which preserves the read pairing and orientation utilising maximum flow algorithm in its splice graph. We implemented a dynamic programming method which uses different characteristics such as the signature of ORF conservation (corresponding to indel patterns) and different biochemical substitutions to calculate the coding capacity of different isoforms. Since lncRNAs are evolving rapidly, we inferred the power of natural selection and its influence on their structures.
To detect recent selection in modern humans, we measured differences in selection for splice sites that are shared with archaics and compared them to those that are modern human-specific. Interestingly we found relaxation of natural selection on some splice sites in modern compared to archaic humans.

Our results indicate that most of the annotated lncRNAs in current gene catalogues are incomplete. We expect that completion of our catalogue and studies of selective constraints on brain lncRNAs will aid in understanding some phenotypic changes and disease vulnerability in specific populations.

POA-176

Investigation of single cell genome instability in normal and neurodegeneration brains

Raheleh Rahbari[1], Sarah Geurs[2], Marie-Christine Galas[4], Bart Dermaut[3], Thierry Voet[1, 2]

[1]Sanger Institute (United Kingdom), [2]KU Leuven (Belgium), [3]Ghent University Hospital (Belgium), [4]Inserm (France)

Tau-mediated neurodegeneration, may be determined during neurodevelopment when Tau induces mitotic problems leading to somatic chromosomal instability and potentially profound genetic heterogeneity in the brain. In this study, we integrated the novel single-nucleus Genome &Transcriptome sequencing (snG&T-seq) to investigate genomic instability, and diversification in normal and tau-mediated brains. Here, we present our preliminary findings from whole genome and transcriptome sequencing of ~600 single neurons from frontal cortex across five normal brain donors with mean age of 58 years (ranges from 41 to 72yrs). From transcriptome data, we can detect on average 3,000 genes that are expressed per single cell. We identified 545 differentially expressed genes with adjusted p-value < 0.01 (non-parametric Kruskal-Wallis test). Principal component analysis revealed three main clusters. Top marker genes per each cluster (SC3 framework1), suggest these three clusters belong to two major neuronal cell types: Excitatory and Inhibitory neurons, also reported in the recent studies of human brain cell types2. Moreover, our data from parallel genome analysis show megabse-scale copy number variations, that are also traceable in the transcriptome data. In following we are proceeding with an integrative DNA-RNA analysis to sift true from false CNVs by co-detection of corroborating gene dosage events. We will use this informatics pipeline to investigate genome instability in tau-mediated and Alzheimer brains using snG&T-seq.

POA-179

Weak purifying selection in genomic regions with high diversity in great apes

David Castellano[1], Adam Eyre-Walker[2], Kasper Munch[1]

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

Natural selection is expected to be more efficient in genomic regions that are more diverse because both the efficiency of natural selection and genetic diversity are expected to depend upon the local effective population size (N_e). Here, we estimated nucleotide diversity at putatively neutral non-coding sites, θ_{NC}, and nonsynonymous sites, θ_N, and a measure of the efficacy of selection, the ratio θ_N/θ_{NC}, in great apes using whole genome data from the Great Ape Genomes Project. When genes are ranked and bin according to θ_{NC}, we observe a quadratic relationship between $\theta_N/\theta_{NC}$and θ_{NC} in humans, gorillas and orangutans, even after accounting for the statistical dependence between θ_N/θ_{NC} and θ_{NC}. When we control for the variation in the mutation rate and genes are ranked and bin according to the ratio between θ_{NC} and divergence at putatively neutral non-coding sites (D_{NC}), the relationship becomes negative. We also control for the effect of biased gene conversion on neutral substitution rate by analyzing only A<->T and C<->G substitutions. We find that variation in mutation rate is a major determinant of both variation in diversity and the efficiency of natural selection within the great apes genome, while recombination rate is not. We conclude that purifying selection is weak in high diversity regions of humans, gorillas and orangutans because high diversity regions have a high mutation rate and this increases Hill-Robertson interference.

POA-180

Evidence that the motility organelle of Mycoplasma pneumoniae is under translational selection and insights on its pathogenic lifestyle

Hassan Sibroe Abdulla Daanaa[1], Ali Mostafa Anwar[1]

[1]Faculty of Agriculture, Cairo University, 12613, Giza (Egypt)

Mycoplasma pneumoniae(Mpn) is an obligate pathogen whose reduced genome appears to have emerged through purifying selection. Adaptation and survival of this microbe depends on attachment of a motility organelle to epithelial cells of its host, *Homo sapiens*. As these bacteria lack genes for nucleic and amino acid biosynthesis, they are strongly reliant on host resources. The preferential use of synonymous codons, codon usage bias(CUB), is an important adaptation for many microbes: highly expressed genes undergo the strongest translational selection. We employed a systems approach using available data to identify structures/processes under translational selection in the *Mpn* genome. We find that overall CUB in *Mpn* is normally distributed, and that ribosomal protein-coding genes(RPCGs) and those of the motility organelle(MPCGs) are significantly more biased than the genome, and these biases are consistent with adaptation to the tRNA pool of *H. sapiens*. These results concurred with mRNA expression data that revealed MPCGs followed by RPCGs as the most abundant transcripts. Proteomics data did not completely parallel transcript levels. MPCGs proved the most abundant proteins however, some less biased genes(i.e. metabolic enzymes) showed higher protein concentrations than RPCGs. The dissimilarities between mRNA and proteomic data suggest strong translational regulation in *Mpn*. Overall, the results indicate:1)an adaptive use of CUB by *Mpn* to rapidly express MPCGs, vital in high quantities for survival and propagation. 2)a CUB-transcription association. 3)a co-evolutionary history through parasite CUB dependency on anticodons of its host, suggesting an evolutionary force that would favor changes in host CUB over time.

POA-181

Population genomics of three North American conifer species: sugar pine, loblolly pine, and douglas-fir

Lida Anita To[1], Kristian A. Stevens[1], Marc W. Crepeau[1], David B. Neale[2], Charles H. Langley[1]

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

Population geneticists have long been fascinated with the ecological forces and genetic mechanisms that space, shape, and maintain genetic polymorphism across the genome. Whole-genome sequencing of the huge genomes of widely distributed, long-lived conifer species may offer insights into the applicability of fundamental theories, especially when juxtaposed to the extensive information from forestry and paleobiology. To address these issues, we leverage whole-genome sequence data of three gymnosperms from the division Pinophyta. Conifers make interesting population genetic study candidates due to their high rates of gene flow, as well as strong signals of local adaptation. Following the whole-genome resequencing of a sample of three North American conifer species' haploid genomes across their natural ranges -- sugar pine (n = 12; 31 Gbp), loblolly pine (n = 12; 22 Gbp), and douglas-fir (n = 13; 18 Gbp) -- as well as a single "outgroup" genome for each species: Western white pine, slash pine, and big cone douglas-fir -- we characterize and analyze the level of polymorphism and divergence across the genomes of these three conifer species to detect evidence of direct and linked selection, as well as infer their demographic histories to uncover signals of large scale changes in population sizes, such as signals of glacial refugia. Initial analyses test for parallel bottlenecks during the Last Glacial Maximum (~33,000 years ago). The impacts of linked selection, background selection and hitchhiking in genomes as large as these, as well as the histories of transposable element dynamics in Pinophyta will be discussed.

POA-182

GC-biased gene conversion and structural variation in highly recombining social insects

Takeshi Kawakami[1], Andreas Wallberg[2], Matthew T Webster[2]

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

Recombination rate is highly variable between and within species. Social hymenopterans, such as honeybees (*Apis melifera*), are known to have extremely high recombination rates, which are >20 times higher than humans. However we still lack a clear picture of the evolutionary significance of such extreme rate and its impact on genome evolution. Here we directly inferred meiotic crossover and gene conversion events by sequencing whole-genomes of 347 haploid males collected from 36 colonies of European honeybees (*A. m. melifera*), African honeybees (*A. m. scutellata*), Cape bees (*A. m. capensis*), and bumblebees (*Bombus terrestris*). Crossover rate was significantly higher in honeybees than bumblebees (66 and 18 crossovers per genome, respectively). We also found significant heterogeneity in crossover rate between colonies of honeybees with the lowest crossover rates (47 crossovers per genome) exhibited by the Cape bee, a subspecies that is able to reproduce asexually. Importantly, out of the total of 5,500 gene conversion events identified in the four taxa (mean number of gene conversions = 12-65 per genome), nucleotide transmission was significantly biased toward G/C bases over A/T bases in all four taxa (transmission bias c = 0.13-0.16). Moreover, we identified 48,901 deletions, 13,166 insertions, 144 inversions, and 1,004 duplications segregating in the honeybee population, and their distribution was positively associated with recombination rate along the genome. We discuss distribution of crossover breakpoints and gene conversions along the genome and potential impacts of recombination on the evolution of genome structure and base composition.

POA-183

Cytosine Methylation affects the Mutational Spectrum Beyond the Base Itself

Vassili Feodorovich Kusmartsev[1, 2], Tobias Warnecke[1, 2]

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

DNA modifications can affect the mutation rate of the bases which they modify. In particular, methylated cytosines at CpG sites have long been known to spontaneously deaminate at much higher rates than if they were unmethylated. Intriguingly, data from NMR experiments, molecular dynamic simulations, and other studies suggest that CpG methylation has an impact on the mechanical properties of the DNA helix that goes beyond single base. It is therefore conceivable that cytosine methylation impacts lesion formation or repair dynamics on a local scale, beyond the methylated base. Here, we investigate this possibility using the incidence rate of rare SNPs as a proxy for mutation rate. By comparing methylated and unmethylated sites - matched for sequence context, nucleosome association and chromatin state - we find that, in humans, CpG methylation significantly protects adjacent bases against transversion mutations (plus-minus 3 around the focal CpG). In particular, G-to-C and A-to-T transversion mutation rates are reduced by up to 19% at the base directly adjacent to the methylated CpG. Using additional population-level variation data, we go on to characterise similar local differences in methylated-dependent mutation spectra in mouse, Arabidopsis, bee, and rice. Although the overall impact of methylation on neighbouring mutational dynamics is comparatively weak, it nonetheless has a significant cumulative impact on genomes over evolutionary time scales. Our research suggests that methylation affects genome evolution beyond the modified base itself, and that the cumulative impact of epigenetic modifications on evolution remains to be fully characterised.

POA-184

Resolving the evolutionary impact of polymorphic gene duplications in humans at the haplotype level

Marie Saitou[1], Omer Gokcumen[1]

[1]University at Buffalo (United States)

Gene duplications are principal drivers of evolution. Several variable gene duplications in humans have been shown to contribute to phenotypic diversity. However, the evolutionary forces that maintain variable gene duplications across the human genome are largely unexplored.

To bridge this gap, we conducted genome-wide analyses of all 11,048 polymorphic duplications reported in 1,000 genomes phase 3 dataset. Specifically, we conducted enrichment analyses and confirmed previous observations that polymorphic gene duplications are evolving with significantly less negative selection constraint as compared to polymorphic deletions. In parallel, we are developing novel model-based tests to determine the adaptive forces, if any, which shape the global distribution of duplication variants in humans.

In parallel, to further understand the haplotype architecture of the derived duplications, we developed a linkage-disequilibrium based method. Using this method, we resolved the insertion sites and haplotypes that harbor 14 common polymorphic duplications. For example, we found that a partial duplication of a well-characterized pigmentation-related HERC2 gene, showed unusually high allele frequency differences between human populations (3% in CEU, 70% in CHB and 30% in YRI). This observation suggests the presence of a selective sweep in European populations and balancing selection in East Asian populations. By studying the haplotype carrying the derived duplicate, we were able to provide additional, haplotype-level evidence of such local selections. Our study provides a first look at the evolutionary impact of much understudied polymorphic gene duplications in human populations and presents methodological insights for future studies.

POA-185

Fine-scale characterization of genomic structural variation hotspots in the human genome reveals adaptive and biomedical roles

Yen-Lung Lin[1], Omer Gokcumen[1]

[1]University at Buffalo (United States)

Genomic structural variants (SVs) distributed nonrandomly across the human genome and cluster into "hotspots." Such hotspots have been implicated in critical evolutionary innovations as well as serious medical conditions. However, the evolutionary and biomedical features of these hotspots remain incompletely understood. Here we analyzed data from 2504 genomes from 1000 Genome Project Consortium and constructed a refined map of 1,148 SV hotspots in human genomes. By studying the genomic architecture of these hotspots, we verified the nonallelic homologous recombination (NAHR) mediated by segmental duplications as the primary mechanistic driver of SV hotspots. Still, we found that a majority of SV hotspots are not dependent on segmental duplication content, potentially implicating mechanisms other than NAHR. Evolutionarily, we found that the majority of SV hotspots are within gene-poor regions and evolve under relaxed purifying selection or neutrality. However, we found a small subset of SV hotspots in gene-rich regions that are previously associated with anthropologically crucial traits, including blood oxygen transport, olfaction, synapse assembly, and antigen binding. We provide evidence that balancing selection may have maintained these SV hotspots. Biomedically, we found that the SV hotspots coincide with breakpoints of clinically relevant large, *de novo* SV significantly more often than genome-wide expectations. As such, the mutational instability in SV hotspots likely enables chromosomal breaks that lead to pathogenic structural variation formations. Our study contributes to a better understanding of the mutational landscape of the genome and implicates both mechanistic and adaptive forces in formation and maintenance of SV hotspots.

POA-189

The impact of gremlin mutations on cancer risk

Kazuki K Takahashi[1], Hideki Innan[1]

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

Cancer is considered as a "disease of the genome" because cancer initiation and progression are generally caused by a number of somatic mutations, mainly in cancer-causing genes or cancer driver genes.

Since the development of the next-generation sequencing technology, a large number of cancer cells together with normal cells from the same patients were sequenced and analyzed to identify cancer drivers. However, most of previous studies focused only on somatic mutations, and gremlin mutations have been nearly ignored.

We here analyzed over 6000 patients sequence data to explore the impact of germline mutations in cancer genes to cancer risk. We found very few gremline mutations that are significantly enriched in cancer patients. Nevertheless, we found that the accumulation of many rare nonsynonymous and synonymous mutations in tumor suppressor gene would affect cancer risk. The number of rare gremlin mutations per patient is negatively correlated with the diagnose age in many cancer types, indicating that rare gremlin mutations could potentially increase cancer risk. It is also found that the effect of gremlin mutations seems to be larger for rarer mutations. These results suggest that germline mutations should have non-negligible effect on determining cancer risk especially when many mutations are accumulated, although the impact of each may be very small.

POA-190

Evolution of codon bias and gene expression in the highly AT-biased genome of Dictyostelium discoideum

Janaina Lima De Oliveira[1], Atahualpa Castillo Morales[1], Jason Wolf[1], Chris Thompson[2]

[1]University of Bath (United Kingdom), [2]University College London (United Kingdom)

Alternative synonymous codons are not used with equal frequencies both among species and among genes from the same genome. Comparative studies have found that, among species, codon usage bias (CUB) can be partially predicted by differences in genomic GC content. Similarly, on an intragenomic scale, genes lying within GC-richer regions use GC3-richer synonymous codons, accounting for differences in CUB within mammalian genomes. This suggests that GC content plays an important role in shaping patterns of codon usage. However, base composition cannot explain CUB among genes when such effects are evenly distributed across the genome. Instead, there is increasing evidence of the relevance of selection on synonymous codons to optimise gene expression. Since most studies on eukaryotes have focused on species with unbiased GC-content genomes, it is often difficult to disentangle the relative effects of base composition and selection on CUB. To address this, we analysed patterns of codon usage in the highly AT-biased genome of the social amoeba Dictyostelium discoideum and show how base composition and selection interplay on shaping variation at synonymous sites in this species.

POA-192

Evolutionary origins of taxonomically restricted genes in Drosophila genus

Karina Zile[1, 2], Christophe Dessimoz[1, 2, 3]

[1]University College London (United Kingdom), [2]Swiss Institute of Bioinformatics (Switzerland), [3]University of Lausanne (Switzerland)

The majority of the protein coding genes in extant genomes evolved by divergence from ancestral genes. Taxonomically restricted genes (TRGs), in contrast, evolve from non-coding DNA regions, in an alternative frames of coding regions and from DNA regions created by structural rearrangement. TRGs have been shown to contribute greatly to phenotypic novelty in several clades. Here we examine the evolutionary origins of TRGs that came into existence after the speciation of *melanogaster* subgroup in *Drosophila* genus. High quality genome assemblies are available for five species in this clade. Being separated by only about 12.8 million years of divergence, these genomes provide sufficient amount of homology signal to enable us to answer our questions. To identify TRGs we first used OMA orthology prediction algorithm to locate gene families specific to this clade, and then validated these candidates based on sequence similarity to homologous DNA region(s) in sister taxa. Unlike the methods based on the lack of detectable homology to known proteins, our approach seeks to discriminate between newly evolved genes and highly divergent copies of established genes and thus allows us to study the evolutionary origins and dynamics of TRGs. We report the relative rates of TRG emergence from previously coding (in an alternative frame) and non-coding regions, and regions created by structural rearrangement. By analysing GC-content of coding and non-coding genome regions we quantify the relationship between nucleotide landscapes and lineage-specific evolution of TRGs.

POA-193

The Red-Queen evolutionary dynamics of recombination hotspots

Thibault Latrille[1], Laurent Duret[1], Nicolas Lartillot[1]

[1]Universite de Lyon 1 (France)

In humans and many other species, recombination events cluster in narrow hotspots distributed across the genome, whose locations are determined by the protein PRDM9. Surprisingly, hotspots are not shared between human and chimpanzee, suggesting that hotspots are short-lived. To explain this fast evolutionary dynamics of recombination landscapes, an intra-genomic Red-Queen model, based on the interplay between two antagonistic forces, has been proposed. On one hand, biased gene conversion, results in a rapid extinction of hotspots in the population. On the other hand, the resulting genome-wide depletion of recombination induces strong positive selection favoring new Prdm9 alleles recruiting new hotspots across the genome, thereby restoring normal levels of recombination. However, this Red-Queen scenario has not been formalized as a quantitative model.

We propose a stochastic population-genetic model of the Red-Queen dynamic of recombination. Although we model the system at the level of population, assumptions of our model are based on mechanisms occurring at the molecular level during recombination. Analytical and numerical results suggest that the genetic diversity at the PRDM9 locus is seamingly neutral, although each allele is under positive selection. Moreover, our model suggest that the extinction of hotspots leads to genome-wide depletion of recombination hotspots that is independent of population size, and that the high mutation-rate of PRDM9 is accountable for a restoration of this depletion of hotspots. Together, our model provides analytical predictions at the population level based on molecular mechanisms.

POA-194

Compensatory back mutation in mitochondrial genome of primates

Kazuhiro Satomura[1], Naoki Osada[1], Toshinori Endo[1]

[1]Hokkaido University (Japan)

Most of mutations arisen in molecular evolution are selectively neutral or deleterious to the organisms, according to the neutral theory. In general, deleterious mutations are removed by purifying selection, but fortuitously weakly deleterious mutations are fixed as selectively neutral mutations, e.g. in the case of small population size. To repair the fixed deleterious mutation, the shortest evolutionary route is that the mutation is reversed to ancestral state, i.e. back mutation. The back mutation occurred on deleterious mutation is relatively adaptive to repair the fixed deleterious mutation, the shortest evolutionary route is that the mutation is reversed to ancestral state, i.e. back mutation. Since the back mutation occurred on deleterious mutation is relatively adaptive, the number of back mutations should be larger than the expected value. In order to verify the number of back mutations and to investigate the influence of purifying selection on back mutation, we analyzed 13 mitochondrial genes of 79 primates genomes in this study. As the result, the number of back mutation was larger than the number under selectively neutral. We discussed about the influence of natural selection on back mutations on amino acid sequences in primates mitochondrial genes.

POA-195

The effect of genetic connectivity on the strength of natural selection in Chlamydomonas reinhardtii

Sara El-Shawa[1], Rob Ness[1, 2]

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

Genetic networks represent the complex biological interactions amongst genes, proteins and metabolites. Numerous advances have demonstrated the power of genetic networks to model and predict the effects of gene knock-outs and other mutations. While such experimental studies are important, evolutionary genetics is only beginning to utilize networks to understand the consequences of mutations on fitness and how networks have evolved. For instance, highly connected genes are predicted to experience increased purifying selection when compared to genes with low connectivity. Our study seeks to quantify the position of genes in the network predicts the direction and strength of selection. We are using whole genome population genomic data, with a draft genome of a closely related species and the *Chlamydomonas reinhardtii* genetic network (iRC1080) to analyze patterns of diversity and divergence across all genes in the network. Our analysis will estimate the distribution of fitness effects of new mutations, the proportion of fixed mutations that are adaptive and how these measures vary with network characteristics such as connectivity and centrality.

POA-196

Translational Efficiency and The Evolution of Position-Dependent Codon Bias

Nelson Morrow[1], Ashley Teufel[2], alon Diament[3], Claus Wilke[2]

[1]University of Texas at Austin (United States), [2]University of Texas at Austin (United States), [3]Tel Aviv University (Israel)

The translation of mRNA by ribosomes is the fundamental process underlying protein synthesis. It has been observed that mRNA sequences with codons corresponding to abundant tRNAs are more highly expressed, implying a relationship between expression level and codon usage. This observation has implications for position-dependent codon usage. In fact, the first 30-50 codons have been shown to translate with relatively low efficiency, on average. This initial group of low efficiency codons is often referred to as a ramp. It is hypothesized that these ramps serve to restrict the number of ribosomes allocated to any one sequence at a time to prevent ribosomal traffic jams. Here we simulate the codon evolution of a network of yeast genes evolving under selection for translation efficiency and minimizing translation error with the use of a ribosome flow model (RFM), to examine if these constraints result in formation of ramps. This ribosome flow model considers the dynamic nature of the translation process, modeling translation rates, protein abundance levels, and ribosomal densities interconnected through a pool of free ribosomes. We expect that under some parameterizations of our RFM selection for translation efficiency and the minimization of translation error will result in ramps at the beginning of some sequences in order to discourage ribosomal traffic jams and increase overall network efficiency.

POA-199

Realistic evolutionary simulations for population genetic inferences with SLiM 3

Benjamin C. Haller[1], Philipp W. Messer[1]

[1]Cornell University (United States)

Modern inference frameworks in population genetics rely crucially on simulation tools that allow us to test evolutionary hypotheses and study patterns of genetic variation. Coalescent and Wright-Fisher models have long provided the foundation for such tools, but the applicability of these models is increasingly being questioned due to concerns about their strong underlying assumptions. With the advent of ever more powerful inference methods, there is also a growing realization that inferences from these models could be profoundly biased because they omit important population details such as fine-scale structure, realistic migration patterns, and mate choice preferences. Here we present SLiM 3, a software package for individual-based, genetically-explicit evolutionary simulations that breaks free of the limitations of the Wright-Fisher model. With SLiM 3, it is now easy to model individual variation in survival, mate choice, fecundity, migration, dispersal, and other behavior. Population structure can be accurately simulated, whether as discrete subpopulations connected by migration, or as the actual movements of individuals over a continuous landscape. Non-overlapping generations are now supported, including arbitrary age structure and individual effects of age upon all behaviors (mating, dispersal, etc.). SLiM 3 still provides the same highly-flexible scriptability, interactive graphical modeling, and cross-platform compatibility at the command line as previous versions, all as open-source software freely available on GitHub. As our understanding of fine-scale population structure and migration patterns improves at an unprecedented pace, SLiM 3 will allow us to exploit and harness such information for the next generation of population genetic inference frameworks.

POA-200

Relative Levels of Nonneutral vs. Neutral Polymorphism under Balancing Selection through Storage Effect

Yeongseon Park [1], Yuseob Kim[1]

[1]Ewha Womans University (Republic of Korea)

Since nonneutral polymorphism is the genetic basis of phenotypic variation, how it is generated and maintained is of great interest in population genetics. A number of mechanisms are known to maintain polymorphism in nonneutral loci, including overdominance, selection with spatial or temporal heterogeneity, and storage effect. The latter was suggested based on the idea that different species/alleles can coexist when they react differently to the fluctuating environment and loss under unfavorable environment can be buffered. This study further explores the storage effect and investigates the relative abundance of polymorphism in nonneutral versus neutral loci when mutations at multiple loci have random additive effects on a phenotype. We used forward-in-time, individual-based computer simulations to examine multi-locus genetic variation under seasonally fluctuating selection with storage effect, either due to a refuge from selection or a seed bank. In our model, the fitness of an individual is given by the deviation of the phenotype from the optimum that fluctuates seasonally in either complete or incomplete symmetry. Under various parameter combinations, we observed that the ratio of nonneutral to neutral polymorphism is larger than one. It suggests that, for protein-coding sequences, more nonsynonymous than synonymous polymorphism can exist under storage effect. We further investigate if this mechanism can explain the excess of nonsynonymous polymorphism observed in a population of malaria parasite *Plasmodium falciparum*.

POA-201

Hitchhiking in space, with varying selection intensity and migration rates among demes

Yichen Zheng[1], Thomas Wiehe[1]

[1]University of Cologne (Germany)

Genetic hitchhiking has classically been considered in the framework of panmixis. However, substructure with variable migration rates among demes, and with differing ecological conditions, is the standard in real-life populations. Building on, and extending, earlier models of hitchhiking in space, we study the dynamics of beneficial alleles in a substructured population when exported from a deme of origin to other demes by migration and contrast this to a scenario of convergent evolution, i.e. when beneficial alleles at the same locus arise independently in different demes. Furthermore, we consider the ecologically important cases when the fitness advantage and migration rates vary among demes and ask under which conditions adaptive alleles may spread in the entire population. In particular, intermittent (punctuated) migration, i.e. when periods of isolation are intermitted by bursts of migration, is an important scenario to better understand the conditions of survival of species living in extreme environments. We present results of extensive computer simulations and of a preliminary analysis of beetle and gregarine populations from the Atacama Desert. We also present results from machine learning algorithms performed on multi-dimension statistic datasets derived from simulated and biological samples, which are designed to classify selection and migration patterns to determine the source deme of adaptation and selection heterogeneity in space.

POA-202

Properties of haplotype-based Fst computed as a function of haplotype length

Rohan S Mehta[1], Alison F Feder[1], Noah A Rosenberg[1]

[1]Stanford University (United States)

When the population differentiation statistic Fst is computed from the haplotypes of increasingly long genomic regions, multiple distinct patterns might be expected. First, as a genomic region increases in length from one SNP to several, haplotypes potentially become more distinctive to specific populations, resulting in a high haplotype-based Fst. However, as the region lengthens further, individual haplotypes become increasingly rare, so that the high haplotype diversity might generate low Fst. We study the properties of haplotype-based Fst in a model in which SNPs are sequentially added to a genomic region one at a time. When adding SNPs that are independent of existing haplotypes, we find that Fst usually decreases with increasing haplotype length, increasing only when the minor allele frequency of a new SNP is large and the SNP has a high value of Fst itself. Using data from the Human Genome Diversity Panel, we find that Fst trajectories for increasingly long haplotypes tend to either decrease monotonically or to contain a single peak at small haplotype lengths, depending on the population structure present in the first several SNPs. Owing to the relationship between Fst and heterozygosity, high values of Fst occur only for small haplotype lengths, with almost all differentiation eventually becoming diluted by increasing haplotype diversity. Our results contribute to interpretations of the relationships among population differentiation statistics computed at multiple genomic scales.

POA-203

Efficient and exact computational solutions to Wright-Fisher Markov models with varying population sizes

Ivan Krukov[1], Bianca de Sanctis[1], A. P. Jason de Koning[1]

[1]University of Calgary (Canada)

The Wright-Fisher Markov model (WF) underlies a wide variety of results in population genetics and molecular evolution. WF describes the dynamics of allele frequency changes per generation for a single panmictic population of constant size. However, real populations fluctuate in size over time, and many studies have suggested that these fluctuations are important for both the distribution of weakly deleterious variants in extant populations and perhaps to the rate of adaptation as well. Typically, to account for variable population size, diffusion theory and other approximations are employed to obtain tractable solutions. However, WF Markov models that move between a discrete set of population sizes can also be constructed and analyzed using the theory of absorbing Markov chains. Importantly, such approaches require none of the classical approximations of theoretical population genetics (weak selection, weak mutation, large population size, continuous-time).

We previously developed an exact, sparse, parallel linear algebra solution to the direct computational analysis of the WF model, WFES (Wright Fisher Exact Solver). Here, we present an extension of that work, which accounts for changes in population size - the Markov-Modulated Wright-Fisher model. The approach remains computationally exact, is efficient and fast, and is completely general with respect to the underlying assumptions of the model of population genetics. Using this model, we show how standard approximations to effective population size behave under different switching dynamics, selection, and dominance regimes. This model is implemented in our new program, WFES2, and will be made available on our Github site (https://github.com/dekoning-lab).

POA-206

Evidencing divergent selection from linked sites while accounting for hierarchical population structure

Marco Galimberti[1], Christoph Leuenberger[2], Simone Fior[3], Matthieu Foll[4], Daniel Wegmann[1]

[1]University of Fribourg (Switzerland), [2]University of Fribourg (Switzerland), [3]ETH Zurich (Switzerland), [4]International Agency for Research on Cancer (France)

Allele frequencies vary across populations and loci, even in the presence of migration. While most differences may be due to genetic drift, divergent selection will further increase differentiation at some loci. Identifying those is key in studying local adaptation, but remains statistically challenging. A particularly elegant way to describe allele frequency differences among populations connected by migration is the F-model, which measures differences in allele frequencies by population specific Fst coefficients. This model readily accounts for multiple evolutionary forces by partitioning Fst coefficients into locus and population specific components reflecting selection and drift, respectively. Here we present an extension of this model to linked loci by means of a hidden Markov model (HMM) that characterizes the effect of selection on linked markers through correlations in the locus specific component along the genome. We further extend the model by accounting for population structure at multiple hierarchies to reflect the complex relationships of populations observed in nature. Importantly, this also allows to identify divergent selection at different levels simultaneously. Using extensive simulations we show that our method has increased statistical power compared to previous implementations that assume sites to be independent. We finally illustrate the power of our novel method by identify loci contributing to the polygenic adaptation to altitude in flowering plants of the genus *Dianthus*.

POA-207

Population structure and vulnerability in mangrove species as a result of geographic barriers and climatic changes

Zhengzhen Wang[1], Haomin Lyu[1], Suhua Shi[1]

[1]Sun Yat-sen University (China)

Avicennia marina is one of the most widely distributed mangrove species, morphologically divided into three distinctive varieties, var. *marina*, var. *eucalyptifolia*and var. *australasica*. This study looked into 577 individuals from 16 global-wide locations across all major distribution regions, representing all three varieties. A total of 94 noncoding nuclear genes were sequenced on Illumina high throughput platform for 16 populations. We constructed haplotype networks with *Avicennia alba* as the outgroup. This large-scale dataset was used to conduct biogeographic analyses and investigate the demographic history among varieties. First, low genetic diversity was prevailing in peripheral populations; while for central populations such as those of var. *eucalyptifolia*, the diversity level was significantly higher, which might be a result of discontinuous gene flow around the geographic barrier of Torres Strait. In addition, the geographic barrier of Malacca Strait had caused high divergence of the var. *marina*populations between west and east sides. Second, the population structure principally supported the differentiation among varieties. However, the status of the population in Southwest Australia presented a more complicated divergence pattern. Its standing morphological distinction from var. *eucalyptifolia* in face of strong genetic contact could be a sign to early speciation. Thirdly, our surveys suggested that populations experienced dramatic contraction at ~20,000 years before present. Due to their sensitivity to sea level fluctuation, *A. marina* could be endangered by fluctuating sea levels during the rapid Quaternary climate changes. In conclusion, geographic barriers and climatic fluctuations would shape the diversification and vulnerability within *A. marina*.

POA-208

Multi-species genetic structure, demography and adaptation of the mangrove genus *Rhizophora in the Atlantic East Pacific and South Pacific region, revealed by re-sequencing data.*

Yoshiaki Tsuda[1], Takashi Yamamoto[2], Ryosuke Imai[1], Takaya Iwasaki[3], Koji Takayama[4], Tadashi Kajita[2]

[1]University of Tsukuba (Japan), [2]University of the Ryukyus (Japan), [3]Kanagawa University (Japan), [4]Kyoto University (Japan)

Mangrove forests have been under severe threat due to human activities, such as coastal development, exploitation for fuel wood, forest products and fishpond operations since the 1980s (Tomlinson 1986). Moreover, mangrove ecosystems are threatened by recent climate change and sea-level rise could be their greatest threat (Gilman et al., 2008). To maintain the high value of ecosystem services of mangroves across several dimensions (ecological, socio-cultural and economic, Mukherjee et al. 2014), conservation genomics is a high research priority for many mangrove species. In this study, we focused on the genus *Rhizophora*, including widely distributed and important species in the Atlantic - East Pacific (AEP) and South Pacific regions. We studied 3 species (*R. mangle*, *R. racemose* and *R. samoensis*, with samples covering each species distributional range) by conducting a resequencing approach based on the draft genome of *R. apiculata* and *R. stylosa* (Xu et al. 2017). In total, 6,585,355 single nucleotide polymorphisms (SNPs) were obtained and the deepest genetic divergence was detected between the Atlantic and Pacific Oceans across the American continent, regardless of species. Although more detailed genetic structure was detected in the Atlantic populations of *R. mangle* than a previous study (Takayama et al. 2013), genetic differentiation was not clear among species in the Pacific populations, probably due to ancestral polymorphism. Historical transbarrier (transoceanic and transisthmian) gene flow, past demographic history and adaptation of these species will be discussed together with conservation implications.

POA-210

The population genomics of the mangrove species Sonneratia alba cast light on the genome shaping from isolation and migration

Qipian Chen[1], Zixiao Guo[1], Ziwen He[1], Suhua Shi[1]

[1]Sun Yat-sen University (China)

How genomic composition is shaped by isolation, gene flow and selection? We sought for answers by re-sequencing 33 genomes of *Sonneratia alba* , which is a widespread mangrove species in the Indo-West Pacific (IWP) region. The results of comparative genomic analyses indicated low genetic diversity within populations and strong sequence divergence among populations from four major oceanic regions, i.e. the South China Sea, the East Indian Ocean, northwestern Australia and eastern Australia. The observed population structure was shaped by the geographic isolation due to the emergence of merged shelves (Sunda and Sahul shelves) during the Pleistocene glacial periods. However, as the admixture in the Kukup population indicated, during the interglacial periods, the rising sea level and opening of the Strait of Malacca also permitted genetic exchange between the East Indian Ocean and South China Sea. The contrasts of population structure pattern among genomic loci revealed by genome scanning may provide insight into how isolation and migration shaped the divergence patterns at the genomic level. Besides, the identified highly divergent genes between the two Hainan populations (Sanya and Wenchang) were under strong positive selection as the MK test indicated. Alleles of partial such genes in the Wenchang population were further found to be introduced from Indian Ocean via gene flow through the Strait of Malacca, which may be a case of adaptive gene flow. The study provides clues that interregional genetic exchanges in mangroves facilitate re-use of standing genetic variation to establish in new habitats.

POA-212

Demographic inferences after a range expansion (can be biased): the test case of the blacktip reef shark (Carcharhinus melanopterus)

Pierpaolo Maisano Delser[6, 2, 1], Shannon Corrigan[3], Drew Duckett[4], Arnaud Suwalski[1], Michel Veuille[1], Serge Planes[5], Gavin Naylor[3], Stefano Mona[1]

[1]Ecole Pratique des Hautes Etudes (France), [2]University of Cambridge (United Kingdom), [3]University of Florida (United States), [4]College of Charleston (United States), [5]CRIOBE-USR 3278 (France), [6]Trinity College Dublin (Ireland)

The evolutionary history of a species is a dynamic rather than a static process. Species modify, expand and contract their spatial distributions over time as a consequence of changes in environmental factors. Range expansions (REs) occur through a series of founder events that are followed by continuous migration among neighbouring demes. The process usually results in structured metapopulations and leaves a distinct genetic signature. Most methods for inferring demographic parameters have been developed for unstructured populations. This poses problems because both theoretical and simulation studies suggest that ignoring the influence of population structure caused by REs can mislead demographic inferences. Here we explore these questions empirically using *Carcharhinus melanopterus*, an abundant reef associated shark. This species has experienced a RE and appears highly structured throughout its range. We used a population genomics approach to statistically confirm the occurrence of a RE and identify its origin. We show that ignoring the patterns induced by the RE would have generated widespread spurious signals of population bottlenecks. We fit metapopulation models to the data and show that the degree of connectivity, associated with habitat availability, better explains the observed variation in effective size through time.

POA-213

Impacts of Recurrent Hitchhiking on Divergence and Demographic Inference in Drosophila

Jeremy D Lange[1], John E Pool[1]

[1]University of Wisconsin (United States)

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

POA-214

Population structure of amphioxus *Asymmetron lucayanum* at West and Central Pacific region

Hsiu-Chin Lin[1], Shun-Yi Fang[1]

[1]National Sun Yat-sen University (Taiwan)

Amphioxus (Subphylum Cephalochordata) are marine chordates encompassing three genera and thirty species. They inhabit soft substratum at warm and shallow seas. Asymmetron lucayanum is the dominant species at the main island of Taiwan. They were recorded from northern Taiwan, southern Taiwan, Lanyu, Green Island, Xiao-liu-kiu, Taiwan Banks, and Dongsha atoll. The highest reported density is 180 per m2 at Kenting, southern Taiwan in June. Asymmetron lucayanum is the only amphioxus species with circumtropical distribution. Based on the DNA sequences of mitochondrial Cytochrome C Oxidase I (COI), A. lucayanum was further divided into three clades, potentially three cryptic species. Clade A is present in the Indo-West Pacific Ocean; Clade B is present in the West and Central Pacific Ocean; Clade C is present in the Atlantic Ocean. Clade A and B are sister taxa and sympatric at the West Pacific where Taiwan locates. In this study, COI sequences of 67 A. lucayanum samples from the coasts of Taiwan were retrieved and all identified as Clade B based on neighbor-joining analyses. Haplotype network analyses of all Clade B samples suggested the existence of three populations. The first population is composed of samples from Taiwan, Ryukyu and the Philippines. The second population is composed of samples from above localities, in addition to Great Barrier Reef and Hawaii. The third population is only composed on samples from Hawaii. The first population is likely ancestral and diverged from Clade A at the West Pacific Ocean, following by distribution expansion into the Central Pacific area.

POA-216

Predicting Cancer Driver Sites and Cancer-specific Selection Pressures under Two-Component Evolutionary Models

Zhan Zhou[1], Jingcheng Wu[1], Wenyi Zhao[1], Zhixi Su[3], Yangyun Zou[3], Xun Gu[2, 4]

[1]Zhejiang University (China), [2]Iowa State University (United States), [3]Fudan University (China), [4]Fudan University (China)

Current cancer genomics databases have accumulated millions of somatic mutations in coding regions of genes, but details remain largely unknown. Simply speaking, there are two types of somatic mutations: driver mutations underlie the process of carcinogenesis, whereas the passenger mutations have little contribution. It is the general thought that only a small portion of cancer somatic mutations are drivers, characterized by high occurrences of recurrent mutations in independent cancer samples. However, the problem is the cutoff arbitrary, i.e., there is no consensus for how many recurrent mutations are sufficient to make the prediction of drivers. Here we develop a statistical framework to predict cancer driver mutation sites and cancer-specific selection pressures on driver sites and passenger sites, respectively. Based on principles of cancer evolution, independent cancer samples are treated as a star-tree, along which somatic mutations occur randomly. We implement two-component model as follows: while the ground component corresponds to passenger mutations, the rapidly-evolving component corresponds to driver mutations. Through the empirical Bayesian rule, one can calculate the posterior probability for each site of a gene being the driver, providing statistically sound driver site predictions under a given cutoff. Using our method to the somatic mutation data from the Cancer Genome Atlas (TCGA) database, we have predicted 10,075 cancer driver sites which distributed in 3,537 genes, with the posterior probability more than 50%. Our results have shown that the new method is practically feasible and efficient.

POA-217

Rules of neutral molecular evolution are only -half right Influences of positive vs. negative selection

Qingjian Chen[1]

[1]Sun-yat san university (China)

Neutral theory is the most popular theory in modern molecular evolution. The neutral theory holds neutral fixation in evolutionary processes. The neutral fixation domination assumption is supported by two important rules. 1) Evolutionary rate is diverse in functional different sites. Conservation changes occur more frequently than radical changes. 2) Purifying selection is ubiquitous whereas positive is rare. Although well accepted, the two rules may fail in certain systems.

To test the functional difference between amino-acids(AAs) pairs, we separate the non-synonymous changes into 75 classes according to their codon exchangeability, named Evolutionary Index(EI). EI is highly correlated with each other in wide taxa, from plants, invertebrates to vertebrates when Ka/Ks<0.2. We convert different EIs into a Universal Index(UI), which is significantly negative correlated with AAs properties. The universal correlation suggests that the strength of negative selection is determined by biochemical properties of AAs and supporting Rule 1. But when Ka/Ks increases, such as human-chimpanzee, the correlation becomes poorer . This suggests radical changes fixed faster, which contradicts Rule 1.

Another question is how ubiquitous positive selection is. Here we define positive selection by Ka/Ks-Pa/Ps (P: polymorphism) and negative selection by 1-Pa/Ps. The positive and negative selection are coherent with each other in Human. This suggests that behind ubiquitous purifying selections lies much positive selection . Rule 2 is disrupted.

In general, neutral theory apply to a wide swath of organisms and evolutionary processes, but not applicable to faster evolving system since Rule 1 and 2 often fail.

POA-218

Low somatic mutational robustness of the human genome

Sofya Garushyants[1, 2], Georgii Bazykin[2, 1]

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

The rate of deleterious mutations could be greatly reduced if the genome possessed higher mutational robustness, i.e., if its composition was biased against mutagenic features at functional sites. In existing genomes, the robustness to germline mutations is low, and this is usually ascribed to weakness of selection in its favor. As the somatic rate of deleterious mutations may be higher than the germline rate, selection in favor of mutationally robust composition is expected to be much stronger in genes and functional elements in which a high rate of somatic mutations will be deleterious. Here, we test this hypothesis using the mutations in human soma that contribute to cancer. 40% of driver mutations associated with cancer occur in CpG dinucleotides, and all these mutations could be avoided without changes to amino acid sequence by using alternative codons. However, we find no evidence of selection in favor of alternative non-CpG codons at these sites, or of conflicting nucleotide-level selection pressure in favor of CpG codons.

POA-219

Considering somatic mutation rate as a measure of genome maintenance capacity in colonial cnidarians

Elora Hayter Lopez[1], Stephen R Palumbi[1]

[1]Stanford University (United States)

Scleractinian corals are a speciose group of cnidarians, most of which are colonial. Whether they possess germline segregation, whether they senesce, and which levels of natural selection are most relevant in these modular organisms is unclear and controversial. Genome maintenance research may answer these questions, and thereby better clarify the evolution of aging and germline-soma distinction. Colony growth is driven by asexual reproduction of polyps, but the extent of selection on the polyp level compared to the colony level is unknown. We analyzed four scleractinian colonies that each had transcriptomes sequenced from 17-22 branches. We called single nucleotide variants that were heterogeneous within a colony, then filtered to identify top candidates for somatic variants. We resequenced the top candidates to verify true somatic mutations. The average coral sample has 10^{-7} to 10^{-8} mutations per nucleotide, an order of magnitude lower than the estimated number of mutations in the average human 15-year old's somatic cells. The base-substitution rate per nucleotide per polyp generation is 5.5×10^{-9} to 2.2×10^{-10} depending on different estimates of the number of polyp generations in a colony. The lower bound is lower than the per-generation mutation rate for other multicellular eukaryotes. Scleractinians may have better genome maintenance mechanisms than other animals. This study provides a framework for how to study genome maintenance and polyp-level selection in colonial organisms with a position in the tree of life that suits them to be key in understanding the evolution of aging and germline-soma distinction.

POA-221

The missing link of cancer evolution - early remnants of tumorigenesis

Bingjie Chen[1], Chung-I Wu[1, 2, 3]

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

Cancer is a problem of selection and clonal expansion, but how a single normal cell transforms to the first aggressive cancer cell(Stage I) and then grows into a macroscopic tumor(Stage II) is blurred. It is impractical to track the transformation and progression of the tumor evolution, while the relic clones which recorded the mutation history of the whole process are crumbled through the whole tumor mass, these intermediate clones may be engulfed by the rapid dividing clones, stay dormant and maintain a relatively constant population size, or accumulate other driver genes and expand slightly. Our multi-regional sequence data, for the first time, detect these intermediate clones and provide direct evidence of mutation accumulation and stepwise selection during normal-tumor transformation. Combined with calculation and simulation result, we conclude that there are substantial intermediate clones mixing within the final dominant tumor clone. Thus, the genetic diversity in a single tumor are made up of two parts: 1) stage II diversity: within main tumor clone diversity which could be extremely high under non-Darwinian evolution, and 2) stage I diversity: the diversity of intermediates clones during tumor transformation which could be equivalent high but may be ignored for a long time. In a rapid growing population, the lower the relative fitness advantage conferred by a driver, the higher population diversity and lower the chance of fixation. Thus, the time(T) and the selection intensity(S) of driver genes influence whether could we detect the intermediate clones when doing retrospective reconstructions from sequencing of final tumors.

POA-222

Measuring DNA mutation rates with Circle-sequencing

Stephan Baehr[1, 2], Lauren Reyes[1, 2], Jean-Francois Gout[1], Michael Lynch[1]

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

DNA mutation is the source of heritable variance in living things. A DNA mutation rate is a core variable measured in evolutionary biology, and is also of concern to medical science: somatic cells have mutation rates too, and appear to have mutation rates one or two orders of magnitude higher than a given species' germline. Recent advances have provided indirect and *in vitro* measurements of DNA mutation rates per cell division, but these methods are temporally demanding, skill-intensive, and therefore generally expensive. Previous publications of circle-sequencing provide a tantalizing alternative, which utilizes multiple (redundant) sampling of individual DNA molecules extracted from tissue. We have extensively updated the protocol to increase its sensitivity, significantly increasing its threshold of detection. We demonstrate the utility of this protocol on prokaryotic and eukaryotic organisms to obtain mutation rate estimates without the need of performing mutation accumulation experiments.

POA-224

Development Of Next-Generation Sequencing (NGS) Techniques For Museum Specimens

Elsa Call[1], Victoria Twort[1], Niklas Wahlberg[1]

[1]Lund (Sweden)

Natural history museums hold a vast amount of biological material that has been amassed over hundreds of years. To date, the majority of the material available has primarily been used for morphological studies. The recent advances in sequencing technologies, has opened up a whole new avenue of research opportunities, including successfully sequencing DNA from fossilized taxa (paleogenomics), such as Neanderthals, mammoths and cave bears. The field of 'museomics' is one that is still developing, but holds a lot of promise. The goal of my PhD project is to apply Next-Generation Sequencing (NGS) techniques to important lepidopteran resources contained in natural history museums. The initial focus has been optimise library preparation protocols, with the subsequent sequencing being used to assess how much of the material recovered is sample vs non-sample contaminate DNA. Preliminary results indicate that more than 60% of the DNA sequenced appears to be from the specimen itself. The development and implantation of these techniques will enable the full use of museum collections around the world. Thereby, enabling interesting avenues of research that were previously inaccessible do to the factors such as the extinction of the species or population, or species that are rare and difficult to collect. Such avenues of research will enable the community as a whole to further develop our understanding of groups of evolutionary interest that until recently were inaccessible.

POA-225

Historic Treponema pallidum genomes from Colonial Mexico

Verena J Schuenemann[1, 2, 3], Aditya Kumar Lankapalli[4], Rodrigo Barquera[4, 5], Elizabeth Nelson[4], Diana Iraiz Hernandez[4, 5], Victor Acuna Alonzo[5], Kirsten I Bos[4], Lourdes Marquez Morfin[6], Alexander Herbig[4], Johannes Krause[4, 2, 3]

[1]University of Zurich (Switzerland), [2]University of Tuebingen (Germany), [3]University of Tuebingen (Germany), [4]Max Planck Institute for the Science of Human History (Germany), [5]National School of Anthropology and History (Mexico), [6]National School of Anthropology and History (Mexico)

Among the worldwide prevalent treponemal diseases, syphilis appears as a global threat that is currently re-emerging. The origins of syphilis and other treponemal diseases are so far unresolved and are subject to an intensive scholarly debate. Until now, assumptions on its origins and evolutionary history could only be drawn from osteological analyses of past cases and genetic analysis of contemporary T. pallidum genomes; contributions from ancient DNA are very rare and have, until now, failed to provide genome-level data. Here we present three historic T. pallidum genomes (two from T. pallidum ssp. pallidum and one from T. pallidum ssp. pertenue) that have been reconstructed from skeletons recovered from the Convent of Santa Isabel in Mexico City, in use between the 17th and 19th century. Our analyses indicate that different T. pallidum subspecies caused similar diagnostic presentations that are normally associated with syphilis in infants, and potential evidence of a congenital infection of T. pallidum ssp. pertenue, the causative agent of yaws. This first reconstruction of T. pallidum genomes from archaeological material opens the possibility of studying its evolutionary history at a resolution previously assumed to be out of reach and thereby establishes a new method that could greatly contribute to uncover the mystery regarding the origins of treponemal diseases.

POA-226

The evidence of cattle domestication in Thailand: Indicated by ancient DNA of cattle specimens in the Bronze and Iron Ages

Sirianong Siripan[1]

[1]Kasetsart University (Thailand)

Animal domestication was one of the most significant achievements of prehistoric human in Neolithic period which our species have changed from hunter gatherer to agricultural culture. Therefore, the study of animal domestication could imply human demographic and cultural evolution. Cattle were selected for this research because they have been associated with human since prehistoric period. They provide us a source of food, use to plough in crop farming. In Thailand, the origin of Thai domestic cattle could not be identified. Thus, the purpose of this study is to provide the first genetic history of cattle domestication in Thailand during Prehistoric time by using ancient DNA analysis. A total of 55 ancient cattle remains excavated from four archaeological sites, aged between 4,000 to 1,730 YBP. Partial D-loop sequences of 25 samples were successfully amplified and sequenced. Then, these sequences were analyzed using blast program and a phylogenetic tree was constructed by Maximum Likelihood. The results showed that all of ancient specimens belong to Bos taurus while native Thai cattle are Bos indicus. Most of the ancient Thai cattle samples shared haplotype with the ancient counterparts from other countries such as Iran, Syria, Greece and China collected from GenBank database. In summary, the result contributed that Bos taurus was domesticated in Thailand in the Prehistoric time and they might originate from domestic cattle from other region. In addition, these ancient Thai cattle didn't show a close genetic relationship to native Thai cattle because they belong to different Bos species.

POA-228

Determining Relatedness of Ancient Individuals

Angela Wieber[1], Joshua Schraiber[1]

[1]Temple University (United States)

The population and family structure of ancient societies can shed light on cultural evolution. With the advent of ancient DNA sequencing, it is now possible to sequence multiple individuals localized spatially and temporally. Thus, we developed a novel algorithm for assessing the familial relationships between ancient samples. Specifically, we conducted this project to determine the relatedness of two ancient individuals. Unlike research that utilizes modern DNA, we do not have direct estimates of allele frequencies in the ancient population, and must project modern allele frequencies backward in time. Moreover, our work needs to account for low coverage data and degradation of ancient DNA. We are able to estimate whether a pair of individuals was unrelated, parent and offspring, full siblings, half siblings, or first cousins. We demonstrate the algorithm by application to data from burial sites across Europe.

POA-231

The genetic history of Africa based on modern and ancient DNA

Carina Maria Schlebusch[1, 2]

[1]Uppsala University (Sweden), [2]University of Johannesburg (South Africa)

In the last few decades, genetics played an increasingly important role in exploring human history. Genetic studies provided conclusive information that helped to answer challenging questions, such as the "Out of Africa" migration of modern humans. Moreover, genetics helped to establish Africa as the birthplace of anatomically modern humans. The history of human populations in Africa is complex and includes various demographic events that influenced patterns of genetic variation across the continent. Several studies based on mitochondrial DNA, Y-chromosomes, autosomal markers and whole genomes contributed to unraveling the genetic sub-structure of African populations. Through these studies, it became evident deep African history is captured by connections among African hunter-gatherers, and that the deepest population divergence date to around 300,000 years before present. Furthermore, it was shown that agriculture had a large influence on the distribution of current-day Africans and that West African agriculturist populations populated the whole of sub-Saharan Africa, replacing and/or assimilating former groups. Other farming groups from Northeast Africa, admixed with Middle Eastern populations and also expanded southwards. These later population movements disrupted pre-existing population distributions and complicate inferences regarding deep human history. With the increased availability of full genomic data from diverse African populations we have more power to infer human demography. Furthermore, the first successful African ancient DNA genomes allow for direct temporal comparisons. With the promise of many more African modern and ancient genomes to come, the next few years will be exciting for investigating our species deep genetic history, rooted in Africa.

POA-233

The Transition to Farming in Eneolithic (Copper Age) Ukraine was Largely Driven by Population Replacement

Ryan William Schmidt[1], Daniel Fernandes[1, 2, 3], Jordan Karsten[4], Thomas Harper[5], Gwyn Madden[6], Sarah Ledogar[7], Mykhailo Sokhatsky[8], Hiroki Oota[9], Ron Pinhasi[1, 2]

[1]University College, Dublin (Ireland), [2]University of Vienna (Austria), [3]University of Coimbra (Portugal), [4]University of Wisconsin-Oshkosh (United States), [5]The Pennsylvania State University (United States), [6]Grand Valley State University (United States), [7]University of New England, Armidale (Australia), [8]Borschiv Regional Museum (Ukraine), [9]Kitasato University (Japan)

The transition to a farming-based economy during the Neolithic happened relatively late in southeastern Europe. Material changes occurred through pottery manufacture, but it wasn't until the sixth millennium BCE that farming was adopted by the Cucuteni-Trypillian archaeological complex (4800-3000 BCE). In many parts of Europe, early farmers who were descended from Anatolian migrants slowly admixed with local hunter-gatherers over the course of the Neolithic. In Eastern Europe and the Balkans, this process may have been more complex since early farmers would likely have admixed with local groups prior to spreading into continental Europe. Studies from the Baltic and Estonia suggest little genetic input from early farmers or continuous admixture with hunter-gatherers. Here, we investigate the impact of Trypillian migrations into Ukraine through the analyses of 19 ancient genomes (0.6 to 2.1X coverage) from the site of Verteba Cave. Ceramic typology and radiocarbon dating of the cave indicate continuous occupation from the Mesolithic to the Medieval Period, with peak occupation coinciding with the middle to late Tripolye. We show that the Trypillians replaced local Ukrainian Neolithic cultures. Also, hunter-gatherers contributed very little ancestry to the Trypillians, who are genetically indistinct from early Neolithic farmers. The one exception is a female that has mostly steppe-related ancestry. Direct radiocarbon dating of this individual places her in the the Middle Bronze Age (3545 years before present). Her lack of farmer ancestry suggests abrupt population replacement resulting perhaps from inter-group hostilities or plague that spread through Europe during the Late Neolithic.

POA-234

Population Dynamics at Late Chalcolithic and Early Bronze Age Arslantepe, Anatolia

Eirini Skourtanioti[1], Choongwon Jeong[1], Yilmaz Selim Erdal[2], Marcella Frangipane[3], Philipp Wolfgang Stockhammer[4], Johannes Krause[1, 6], Wolfgang Haak[1, 5]

[1]Max Planck Institute for the Science of Human History (Germany), [2]Hacettepe University (Turkey), [3]Sapienza University of Rome (Italy), [4]Ludwig-Maximilians-University of Munich (Germany), [5]The University of Adelaide (Australia), [6]University of Tubingen (Germany)

While Anatolia was highlighted as the genetic origin of early Neolithic European farmers, the genetic substructure in Anatolia itself as well as the demographic and cultural changes remain unclear. In eastern Anatolia, the archaeological record reflects influences from North-Central Anatolia, the northeastern sectors of Fertile Crescent and the Caucasus, and suggests that some of these were brought along with the movement of people. Central to this question is the archaeological site of Arslantepe (6th-1st millennium BC), strategically located at the Upper Euphrates, the nexus of all three regions. Arslantepe also developed one of the first state societies of Anatolia along with advanced metal-technologies. Archaeological research suggests that conflicts with surrounding groups of pastoralists affiliated to the Caucasus might have contributed to the collapse of its palatial system at the end of the Chalcolithic period (4th millennium BC). To test if these developments were accompanied by genetic changes, we generated genome-wide data from 18 ancient individuals spanning from the Late Chalcolithic period to the Early Bronze Age of Arslantepe. Our results show no evidence for a major genetic shift between the two time periods. However, we observe that individuals from Arslantepe are very heterogeneous and differentiated from other ancient western and central Anatolians in that they have more Iran/Caucasus related ancestry. Our data also show evidence for an ongoing but also recent confluence of Anatolian/Levantine and Caucasus/Iranian ancestries, highlighting the complexity of the Chalcolithic and Bronze Age periods in this region.

POA-235

Ancient dental calculus: unlocking a high-resolution proxy of past human movement and interaction

Raphael Eisenhofer[1], Atholl Anderson[2], Keith Dobney[3], Scott Fitzpatrick[4], Alan Cooper[1], Alain Froment[5], Laura Susan Weyrich[1]

[1]University of Adelaide (Australia), [2]Australian National University (Australia), [3]University of Liverpool (United Kingdom), [4]University of Oregon (United States), [5]Museum National d'Histoire Naturelle (France)

Historical and contemporary human demographic patterns are the result of past human movements. However, our ability to infer rapid past human movements (lasting 100-200 years) is often limited due to the lack of accumulated genetic changes over a small number of human generations. The comparatively more rapid accumulation of genetic changes in vertically-inherited microbial DNA provides a promising new means of inferring rapid human movements. We can now apply this method in ancient human populations by sequencing ancient microbial DNA preserved within calcified dental plaque (calculus). Here, we reconstruct past human population movements and interactions among Pacific Islands using spatially and temporally diverse ancient dental calculus samples. We identify bacterial species whose evolutionary history recapitulates human movements, and develop and test a novel microbial DNA enrichment panel. Using this method, we identify several new connections between Pacific Islands and investigate bacterial lineage replacement in areas where population interactions resulted in the replacement of existing human populations. This method has the potential to provide unprecedented insights into past rapid human movements and cultural interactions around the world, and, critically, to answer questions that currently cannot be resolved using human genomics.

POA-236

Genetic transition in the Swiss Late Neolithic and Early Bronze Age

Anja Furtwaengler[1], Ella Reiter[1], Gunnar U. Neumann[1], Inga Siebke[2], Noah Steuri[3], Joachim Wahl[4, 5], Juergen Hald[6], Verena J. Schuenemann[1, 7, 8], Philipp Stockhammer[9], Albert Hafner[3, 10], Sandra Loesch[2], Johannes Krause[1, 7, 11]

[1]Eberhard Karls University of Tuebingen (Germany), [2]University of Bern (Switzerland), [3]University of Bern (Switzerland), [4]Ctiy of Constance (Germany), [5]Eberhard Karls University Tuebingen (Germany), [6]Ctiy of Constance (Germany), [7]Eberhard Karls University Tuebingen (Germany), [8]University of Zurich (Switzerland), [9]Ludwig Maximilians University Munich (Germany), [10]University of Bern (Switzerland), [11]Max Planck Institute Jena (Germany)

Recent studies have shown that the beginning of the Neolithic period as well as final stages of the Neolithic were marked by major genetic turnovers in European populations.The transition from hunter-gatherers to agriculturalists and farmers/farming in the 6th millennium BP coincided with a human migration from the Near East. In the 3rd millennium BP a second migration into Central Europe occurred originating from the Pontic steppe linked to the spread of the Corded Ware Complex ranging as far southwest as modern day Switzerland. These genetic processes are well studied for example for the Middle-Elbe-Saale region in Eastern Germany, however, little is known from the regions that connect Central and Southern Europe.

Here, we investigate genomic data from 69 individuals from the Swiss Plateau and Southern Germany that span the transition of the Neolithic to the Bronze Age (5500 to 4000 BP). Our results show a similar genetic process as reported for the Middle-Elbe-Saale region suggesting that the migration from the Pontic steppe reached all the way into the Swiss plateau. The high quality of the ancient genomic data also allowed an analysis of core families within multiple burials, the determination and qualification of different ancestry components and the determination of the migration route taken by the ancestors of the Late Neolithic populations in this region. This study presents the first comprehensive genome wide dataset from Holocene individuals from the Swiss plateau and provides the first glimpse into the genetic history of this genetically and linguistically diverse region.

POA-237

Quantifying performance of admixture detection with ancient DNA

Torsten Gunther[3], Amy Goldberg[2], Joshua G Schraiber[1]

[1]Temple University (United States), [2]UC Berkeley (United States), [3]Uppsala University (Sweden)

Ancient DNA opened a new window into population structure, allowing for a detailed dissection of how population structure changes through time. From this, we have learned many things that would have been impossible to detect from contemporary data alone, including the convoluted make up of modern Europeans. To make these inferences, many groups will use a common set of tools to try to estimate the proportion of each ancestral component in their samples: Admixture and qpAdm. However, there has been no comprehensive assessment of the performance of each tool. In this work, we perform a simulation study to assess the performance of these two methods in the face of the low coverage and sequencing errors that are common in ancient DNA datasets. We find that qpAdm is highly robust to sample sizes, admixture fractions, and coverage. On the other hand, Admixture is particularly impacted by using pseudo-haploid data derived from low coverage ancient DNA. We suspect this is because it is unable to find instances of Hardy-Weinberg equilibrium; we followed up by assessing if a genotype likelihood based approach, such as ngsAdmix, is able to improve the performance of Structure-like approaches.

POA-238

Efficiently integrating ancient DNA into modern Y chromosome trees

Rui Martiniano[1, 2], Lara Cassidy[3], Daniel Bradley[3], Richard Durbin[1, 2]

[1]University of Cambridge (United Kingdom), [2]Wellcome Trust Sanger Institute (United Kingdom), [3]Trinity College Dublin (Ireland)

During the last decade, a huge wealth of ancient Y chromosome data has been generated as part of whole-genome shotgun and target capture sequencing studies. However, given the highly degraded nature of ancient DNA (aDNA) data, post-mortem deamination and often low genomic coverage, combining ancient and modern samples for phylogenetic analyses remains challenging. Most analyses use limited markers and/or extensive manual curation.

Current standard methods for the analysis of Y chromosome data focus on known, gold-standard markers, but these contain only a subset of the total Y chromosomal variation. Examining all polymorphic markers is particularly useful when dealing with low coverage aDNA data because it substantially increases the number of overlapping sites between present-day and ancient individuals and it may also help uncover relationships inaccessible via standard known variation.

We provide an automated workflow for jointly analysing ancient and present-day sequence data using all uniquely mappable regions of the Y chromosome. From a given high-coverage dataset, a maximum likelihood phylogeny is estimated and variants are assigned to each branch of the tree. Next, for each low coverage aDNA sample, we count the number of ancestral and derived alleles at each branch and use this information to map ancient lineages to their most likely place in the phylogeny. We apply this method to a large dataset of novel and publicly available data from ancient Eurasians and characterize patterns of Y chromosomal diversity across time as well as the impact of past migrations on the landscape of present-day paternal lineage distribution.

POA-239

People from Ibiza: an unexpected isolate in the Western Mediterranean

Simone Andrea Biagini[1], Neus Sole-Morata[1], Pierre Zalloua[2], Lisa Matisoo-Smith[3], David Comas[1], Francesc Calafell[1]

[1]Institut de Biologia Evolutiva (CSIC-UPF), Universitat Pompeu Fabra (Spain), [2]The Lebanese American University (Lebanon), [3]University of Otago (New Zealand)

According to history, Ibiza's ancestry finds its roots in the Middle East, North Africa and Europe: before the Catalans conquered the island in 1235, Ibiza already had experienced many different cultures. It was the Phoenician Iboshim, the Carthaginian Ibosium, the Islamic Yabisah, up to the Catalan Eivissa. How all these different civilizations affected the modern genetic structure of the islanders is still unexplored. In this genome-wide study, we dug into the genetic structure of a group made up of individuals coming from different autonomous communities of Spain. Our results pointed to a clear split in two major groups, clearly separating Ibiza from the rest of the samples.

We aimed to find the historical reasons behind this result: is Ibiza separating because of recent historical events, or because of some more well-established historical reasons? We explored the possibility that the modern samples from Ibiza had something to share with the ancient culture from Phoenicia using a sample retrieved in a Phoenician necropolis on the island of Ibiza. Mostly, our analyses pointed out different aspects that seem to link the genetics of the modern samples with the history of the area they lived in, more than to any ancient genetic echo from the past. According to history, Ibiza experienced a series of dramatic demographic changes due to several moments of famine, wars, up to malaria and plague. Interestingly, the ROH analysis is showing a level of homozygosity that might reflect an event of a not so distant founder effect.

POA-240

Into the great wide open: the genomic history of the Greater Caucasus region

Wolfgang Haak[1], Chuanchao Wang[1], Sabine Reinhold[2], Andrej B. Belinskij[3], Alexey Kalmykov[3], Natalia Berezina[4], Alexandra Buzhilova[4], Thomas Higham[5], Thomas Stoellner[6], Lars Fehren-Schmitz[7], Viktor Trifonov[8], David Reich[9, 10, 11], Svend Hansen[2], Johannes Krause[1]

[1]Max-Planck Institute for the Science of Human History (Germany), [2]German Archaeological Institute (Germany), [3]Cultural Heritage Unit (Russian Federation), [4]Lomonosov Moscow State University (Russian Federation), [5]University of Oxford (United Kingdom), [6]Deutsches Bergbau-Museum Bochum (Germany), [7]University of California Santa Cruz (United States), [8]Russian Academy of Sciences (Russian Federation), [9]Harvard Medical School (United States), [10]Harvard Medical School (United States), [11]Broad Institute of Harvard and MIT (Germany)

The Caucasus mountains, bound by the Black and Caspian seas, connect the Near East and the Eurasian steppes. Recent archaeogenetics studies have described the formation of 'steppe ancestry' ultimately as a mixture of Eastern European and Caucasian hunter-gatherers. However, it remains unclear when this ancestry arose and whether cultural innovations originating in the Near East had facilitated the opening of the steppe environment for pastoralist economies. To test whether this also involved gene flow, we generated genome-wide SNP data from 50 prehistoric individuals along a 3000-year transect through time in the North Caucasus region, ranging from the Eneolithic (6300 yBP) to the Late Bronze Age (3400 yBP). We observe a genetic separation between the groups in the northern foothills and south of the Caucasus, and those of the bordering steppe regions in the north. We coin these 'mountain' and 'steppe' Caucasus groups, according to vegetation zones and characteristics of the associated archaeological cultures. Furthermore, 'Steppe Majkop' individuals harbor a distinct ancestry component that relates them to Upper Paleolithic Siberians and Native Americans. In contrast, genomic ancestry profiles of groups from the northern foothills are similar to those in ancient Georgia and Armenia. This suggests that the Caucasus Mountains are not an insurmountable barrier to human movement, and further permits the detection of periods of genetic continuity as well as occasional gene flow. Intriguingly, individuals associated with Yamnaya and subsequent pastoralist cultures show subtle evidence for Anatolian Neolithic-farming-related ancestry, possibly from different sources in the western and southern contact zones.

POA-242

Lipids and the evolution of human diet

Iain Mathieson[1], Sara Mathieson[2]

[1]University of Pennsylvania (United States), [2]Swarthmore College (United States)

The genes *FADS1* and *FADS2* encode fatty acid desaturases that catalyze the synthesis of long chain plasma unsaturated fatty acids (PUFA). Since these essential molecules can also be obtained from dietary sources, variation at the FADS genes contributes to human adaptation to different diets, and has been shown to be under selection in several different human populations. Here, we leverage both modern and ancient DNA to characterize the evolutionary history of this locus over human evolution.

First, we show that functional variation is ancient - dating to before the split of modern humans and Neanderthals. We show that a derived allele that appears to be an adaptation to a plant-based diet was selected in modern humans before the out-of-Africa event. Despite this, it was almost absent in those humans who did leave Africa. The same derived allele was then later selected again in non-Africans but, surprisingly, this selection was not temporally linked to the development of agriculture, which predated it by several thousand years.

For a more general sense of how evolution at *FADS1* relates to the evolution of human diet, we compare patterns of variation at other lipid-associated genes. We find that they show similar patterns of variation, suggesting a more general role for lipid metabolism in dietary adaptation. This contrasts with patterns of copy number variation at the salivary amylase gene *AMY1* , which we show are equally ancient, but not significantly correlated with subsistence strategy in ancient Europeans.

POA-243

Migration and Social Organization in Medieval Europe - a Paleogenomic Approach

Carlos Eduardo Guerra Amorim[1, 2], Stefania Vai[10], Cosimo Posth[3], Daniel Winger[4], Tivadar Vida[6], Dean Bobo[1], Susanne Hakenbeck[5], Guido Barbujani[7], David Caramelli[10], Walter Pohl[8], Caterina Giostra[9], Johannes Krause[3], Patrick J Geary[11], Krishna R Veeramah[1]

[1]Stony Brook University (United States), [2]University of California, Los Angeles (United States), [3]Max Planck Institute for the Science of Human History (Germany), [4]University of Rostock (Germany), [5]University of Cambridge (United Kingdom), [6]Eotvos Lorand University (Hungary), [7]University of Ferrara (Hungary), [8]Austrian Academy of Sciences (Austria), [9]Universita cattolica del Sacro Cuore (Italy), [10]Universita degli Studi di Firenze (Germany), [11]Institute for Advanced Study (United States)

Few topics in European history are as controversial as the nature and impact of the barbarian migrations in the Early Middle Ages. To better understand this key era that marks the dawn of modern European societies, we implemented aDNA analyses (whole genome sequencing and SNP capture) of a large number of samples (N = 63) from two cemeteries historically associated with the barbarian group known as the Longobards. Our dense cemetery-based sampling allowed us to infer key aspects of their social organization, offering novel insights into European medieval history and recent human demography. For instance, we identified a clear pattern of population structure in both cemeteries, involving at least two ancestry groups. These groups are very distinct in what regards their material culture and mortuary practices, suggesting that, although they coexisted, they rarely admixed, and had different social status. Kinship inference shows that individuals were buried next to their kin, suggesting a society organized around biological relationships. Within each cemetery, one single family (composed predominantly by male warriors of northern European genetic ancestry) appears to have represented the core of the corresponding society. Moreover, genetic diversity across and within populations is consistent with the historical barbarian migrations from northern Europe to the heart of the Roman Empire in the South. Beyond this specific application, our study provides a starting point for assessing the dynamics of allele frequency changes in the Middle Ages, a period in which population size in Europe is thought to have dramatically changed.

POA-246

A new targeted-capture method using bacterial artificial chromosome (BAC) as baits exclusively developed for sequencing relatively large loci of ancient DNA

Kae Koganebuchi[1, 2], Takashi Gakuhari[3], Hirohiko Takeshima[4], Satoshi Kasagi[5], Takehiro Sato[6], Atsushi Tajima[6], Hiroki Shibata[7], Motoyuki Ogawa[1, 8], Hiroki Oota[1, 8]

[1]Kitasato University Graduate School of Medical Sciences (Japan), [2]University of the Ryukyus (Japan), [3]Kanazawa University (Japan), [4]Tokai University (Japan), [5]Kitasato University (Japan), [6]Kanazawa University (Japan), [7]Kyushu University (Japan), [8]Kitasato University School of Medicine (Japan)

Whole genome sequencing gets cheaper because of cost down of reagents used in the next generation sequencer (NGS). However, ancient genome analysis is still cost-consuming because 99.0% of DNA extracted from ancient specimens are non-human, mostly bacterial, DNA. A couple of methods have been developed for condensing DNA targeted, but such commercial kits are not optimized to ancient DNA that is chemically modified and damaged. For enrichment of mitochondrial genome, a previous study developed an original targeted capture method that needs only common laboratory reagents and equipment, using PCR amplicons as baits. We have improved a double-capture method which uses bacterial artificial chromosome (BAC) DNA as probes for sequencing a relatively large gene. We applied the BAC double capture (BDC) approach for the 214 kb autosomal region, *ring finger protein 213*, which is the susceptibility gene of moyamoya disease. To evaluate the reliability of BDC, cost and data quality were compared with those of a commercial kit. The test sequencing using BDC showed almost the same mapping ratios as a commercial kit with much better cost performance. While the ratio of duplicate reads was higher, the cost was less than that of the commercial kit. The data quality was sufficiently the same as that of the kit. Here we propose that the BDC could be better applicable to ancient genome sequencing especially for relatively large loci.

POA-248

Adaptive Evolution and Archaic Introgression of Copy Number Variants in Melanesians

PingHsun Hsieh[1], Zev Kronenberg[1], Stuart Cantsilieris[1], Kendra Hoekzema[1], Katherine Munson[1], Francesca Antonacci[2], Mario Ventura[2], Evan Eichler[1, 3]

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

Copy number variants (CNVs) often range up to several mega-bases and are predicted to have larger effect sizes than single nucleotide variants (SNVs). While there is strong evidence for interbreeding between hominin species using SNVs, genetic introgression of CNVs among hominins remains unexplored despite that CNVs have significantly contributed to human evolution and disease. In this study, we systematically searched for signatures of selection and archaic introgression of CNVs using genomes from Simons Genome Diversity Project (SGDP) and three archaic hominin individuals. We identified a conserved set of 19211 autosomal CNVs in the SGDP genomes. Within Melanesians, 162 CNVs differ significantly in copy number from the rest of human samples. Using unique sequences flanking these stratified CNVs, our demography-aware inferences identified 32 and 14 CNVs as candidates for selective and introgressed alleles, respectively. The set includes a >200 kbp duplication with signatures of positive selection and archaic introgression, found exclusively in Melanesian and Denisovan individuals. Interestingly, the duplication integrated into a region of chromosome 16p11 susceptible to genomic rearrangements associated with autism. We also identified two diverged haplotypes at chromosome 8p21 only found in Melanesians, Neanderthals, and a South Asian sample. These haplotypes encompass an upstream deletion and a duplication variant of *TNFRSF10D*. BAC and long-read sequencing analysis of non-human primates suggests the duplication, but not the deletion, is ancestral. The introgression of the two haplotypes likely occurred ~48000 years ago between modern humans and Neanderthals. Our results highlight the importance of CNVs in adaptive evolution within human populations.

POA-250

On the duration of Neandertal admixture

Benjamin Peter[1]

[1]MPI Evolutionary Anthropology (Germany)

Accurately inferring the duration of admixture between archaic and modern humans is a largely open problem. Here, I model admixture into modern humans using a forward-in-time model based on linear branching processes, showing that the problem can be transformed into an allele-age inference problem. I show that the history of an introgressed region can be modelled as a two-phase-process: In the recombination phase, the length of an introgressed chromosome decrease due to recombination. Afterwards, a branching phase models the change in frequency of each introgressed fragment due to genetic drift and selection. Under this model, the distribution of frequency and lengths of introgressed haplotypes can be accurately predicted in scenarios with and without selection against introgressed haplotypes. I find that existing methods based on weighted measures of linkage disequilibrium have very little power to distinguish ongoing from pulse admixture, indicating that the confidence intervals of these methods are systematically too narrow. I also find that selection against introgressed haplotypes will result in them being shorter, causing the date of introgression to be overestimated when selection is not taken into account. My results suggest that admixture date estimates need to be interpreted with great care, and that further work is needed to obtain an accurate understanding of admixture processes.

POA-252

Genome-Wide Ancient DNA Portrays the Forming of the Finnish Population Along a 1400-Year Transect

Kerttu Majander[1, 2, 3], Elina Salmela[3, 1], Kati Salo[4], Thiseas Christos Lamnidis[1], Stephan Schiffels[1], Paivi Onkamo[5, 3], Johannes Krause[1]

[1]Max Planck Institute for the Science of Human History (Germany), [2]University of Tuebingen (Germany), [3]University of Helsinki (Finland), [4]University of Helsinki (Finland), [5]University of Turku (Finland)

The Finnish population has long been a subject of interest for the fields of medical and population genetics, due to its isolation-affected genetic structure and the associated unique set of inherited diseases. Recent advances in ancient DNA techniques now enable the in-depth investigation of Finland's demographic past: the impact of migrations, trade and altering livelihood practices.

Here we analyse genome-wide data from over 30 individuals, representing ten archaeological burial sites from southern Finland, that span from the 5th to 19th century. We find the historical individuals to differ genetically from Finns today. Comparing them with surrounding ancient and modern populations, we detect a transition from genotypes generally connected with prehistoric hunter-gatherers, and specifically resembling those of the contemporary Saami people, into a more East-Central European composition, associated with the established agricultural lifestyle. Starting from the Iron Age and continuing through the Early Medieval period, this transition dates remarkably late compared to the respective changes in most regions of Europe. Our results suggest a population shift, presumably related to Baltic and Slavic influences, also manifested in the archaeological record of the local artefacts from the late Iron Age. Our observations also agree with the archaeological models of relatively recent and gradual adoption of farming in Finland.

POA-253

The first Epipaleolithic Genome from Anatolia suggests a limited role of demic diffusion in the Advent of Farming in Anatolia

Michal Feldman[1], Eva Fernandez-Dominguez[4], Luke Reynolds[3], Raffaela Bianco[1], Cosimo Posth[1], Adrian Nigel Goring-Morris[7], Jessica Pearson[2], Hila May[5, 6], Israel Hershkovitz[5, 6], Douglas Baird[2], Choongwon Jeong[1], Johannes Krause[1]

[1]The Max Planck Institute for the Science of Human History (Germany), [2]University of Liverpool (United Kingdom), [3]Liverpool John Moores University (United Kingdom), [4]Durham University (United Kingdom), [5]Tel Aviv University (Israel), [6]Tel Aviv University (Israel), [7]Hebrew University of Jerusalem (Israel)

Anatolia was home to some of the earliest farming communities, which in the following millennia expanded into Europe and largely replaced local hunter-gatherers. The lack of genetic data from pre-farming Anatolians has so far limited demographic investigations of the Anatolian Neolithisation process. In particular, it has been unclear whether farming was adopted by indigenous hunter-gatherers in Central Anatolia or imported by settlers from earlier farming centers. Here we present the first genome-wide data from an Anatolian Epipaleolithic hunter-gatherer who lived ~15,000 years ago, as well as from Early Neolithic individuals from Anatolia and the Levant. By using a comparative dataset of modern and ancient genomes, we estimate that the earliest Anatolian farmers derive over 90 percent of their ancestry from the local Epipaleolithic population, indicating a high degree of genetic continuity throughout the Neolithic transition. In addition, we detect two distinct waves of gene flow during the Neolithic transition: an earlier one related to Iranian/Caucasus ancestry and a later one linked to the Levant. Finally, we observe a genetic link between Epipaleolithic Near-Easterners and post-glacial European hunter-gatherers that suggests a bidirectional genetic exchange between Europe and the Near East predating 15,000 years ago. Our results suggest that the Neolithisation model in Central Anatolia was demographically similar to the one previously observed in the southern Levant and in the southern Caucasus-Iran highlands, further supporting the limited role of demic diffusion during the early spread of agriculture in the Near East, in contrast to the later Neolithisation of Europe.

POA-254

Robust Reference-Free Archaic Admixture Segmentation Using A Structured Permutation-Equivariant Network

Jeffrey Chan[1], Yun Song[1, 2]

[1]UC Berkeley (United States), [2]UC Berkeley (United States)

Identifying genetic variants in the human genome derived from interbreeding with "ghost" (unknown) archaic hominids sheds light on how humans adapted to past environmental changes. Recent studies have garnered significant interest in the possibility of admixture between present-day populations and a yet-to-be-discovered ghost population. However, current methods lack the statistical power to confidently infer the presence of archaic admixture.

Neural networks provide a powerful likelihood-free inference framework for inferring admixture tracts. Unfortunately, no neural network accounts for the structured permutation-equivariance of admixture tracts; that is, a permutation of genotypes within a population should lead to a permutation of the admixture tracts inferred. On the other hand, this property should not apply to cross-population permutations. In addition, the performance of likelihood-free methods is sensitive to demography misspecification, which is common when data contain many populations with few individuals each. Current methods have not demonstrated robustness to misspecification in the full demography.

In this work, we develop a flexible reference-free Bayesian inference method which can elegantly incorporate any known population genetic information (ancient or modern) while remaining flexible and robust to uncertainty. Our method develops the first structured permutation-equivariant neural network and applies it to two population genetic tasks: hypothesis testing and to inferring admixture tracts (segmentation) of archaic DNA. We significantly outperform the state-of-the-art under the reference-free and reference-dependent regime for both tasks. Furthermore, we demonstrate that our method is robust to demography misspecification.

POA-256

Giant deer (Megaloceros giganteus) phylogeography and population dynamics: Insights from Late Pleistocene and Holocene mitochondrial genomes from Eurasia

Alba Rey-Iglesia[1], Adrian M Lister[2], Paula F Campos[1, 3], Selina Brace[2], Ian Barnes[2], Anders J Hansen[1]

[1]Natural History Museum Of Denmark (Denmark), [2]Natural History Museum Of London (United Kingdom), [3]University Of Porto (Portugal)

The major climatic oscillations that characterized the Quaternary Period had a great influence on the evolution and distribution of many species, including major extinction events of megafauna. One of the iconic species that became extinct during the Holocene was the giant deer (*Megaloceros giganteus*), that survived to around 7,660 calendar years BP in Siberia. For the first time, our study addresses the phylogeography and population dynamics of this extinct species using ancient DNA (aDNA). Here, we have combined in-solution capture enrichment and NGS technologies to generate 31 complete mitochondrial genomes from giant deer specimens spanning from the Late Pleistocene, beyond the ^{14}C radiocarbon limit, to 7,660 calendar years BP from Europe and Western Asia. Bayesian phylogenetic analyses of these complete ancient mitogenomes were used to estimate phylogenetic relationships, divergence dates, as well as population dynamics through the Late Pleistocene and Holocene. Based on the results the species is divided into four main clades: two pre-LGM clades that do not appear in the post-LGM genetic pool, and two major clades that show continuity from before the LGM to the Holocene. Our study has also identified a decrease in genetic diversity starting in Marine Isotope Stage 3 (around 40 ka) and collapsing during the Last Glacial Maximum (ca. 20 ka).

POA-257

A Systematic Investigation of DNA Preservation in Medieval Skeletal Elements

Cody Edward Parker[1], Susanne Friederich[2], Wolfgang Haak[1], Kirsten Bos[1], Johannes Krause[1]

[1]Max Planck Institute for the Science of Human History (Germany), [2]State Office for Heritage Management and Archaeology Saxony-Anhalt (Germany)

One of the major factors influencing the quality of ancient DNA analysis is the choice of which skeletal elements are used in the sampling process. As DNA sampling is destructive, it is in the best interests of both molecular biologists and archaeologists for that sampling to be as efficient as possible. Here we present the first systematic investigation of DNA preservation across skeletal elements both from the same individual and across individuals within a single population, using high-throughput, automated, single stranded library preparation and sequencing technologies. To help identify those skeletal elements DNA is most likely to be preserved in highest abundance, we analyze 249 ancient DNA extracts from twelve separate skeletal elements (and multiple portions of each element) each from eleven individuals excavated from the abandoned medieval village of Krakauer Berg in Sachsen-Anhalt, Germany and evaluate them for human and microbial endogenous DNA content. Specifically, our analyses include estimates of library complexity (target and off-target species), percentage endogenous DNA, and contamination load. The resulting ranked listing of skeletal elements will serve as a useful guide for the selection of potential ancient DNA samples, allowing investigators to choose the most suitable area(s) of the skeleton to sample for a wide range of applications.

POA-258

Historical and modern rabbit populations reveal parallel adaptation to myxoma virus across two continents

Joel M Alves[1, 2, 3], Miguel Carneiro[2, 4], Jade Y Cheng[5, 6], Ana Lemos de Matos[7], Masmudur M Rahman[7], Liisa Loog[8, 9], Anders Eriksson[10], Grant McFadden[7], Rasmus Nielsen[5, 6], Thomas P Gilbert[6, 11], Pedro J Esteves[2, 12], Nuno Ferrand[2, 4, 13], Francis M Jiggins[1]

[1]University of Cambridge (United Kingdom), [2]University of Porto (Portugal), [3]University of Oxford (United Kingdom), [4]University of Porto (Portugal), [5]University of California, Berkeley (United States), [6]University of Copenhagen (Denmark), [7]Arizona State University (United States), [8]University of Manchester (United Kingdom), [9]University of Oxford (United Kingdom), [10]Kings College London (United Kingdom), [11]University Museum (Norway), [12]CESPU (Portugal), [13]University of Johannesburg (South Africa)

In the 1950s the myxoma virus was used as a biological weapon to control the invasive wild European rabbit populations in Australia and Europe. The subsequent pandemic decimated populations and resulted in a remarkable natural experiment, where rabbits in both continents rapidly evolved resistance to the virus. We investigated the genetic basis of this resistance by comparing the exomes of modern individuals with the exomes of historical rabbit specimens collected before the virus release. By replicating our analyses in Australia, France and the United Kingdom we found a strong pattern of parallel selection across the three countries, with the same genetic variants changing in frequency over the last 60 years. Notably, these occurred in genes involved in antiviral immunity and viral replication, and support a polygenic basis of resistance. We experimentally validated the functional role of these genes as viral modulators and showed that selection acting on three amino acids in an interferon protein increased its antiviral effect.

POA-259

Reliable Inference of Genetic Diversity within and between Ancient and Modern Genomes

Vivian Link[1, 7], Zuzana Hofmanova[1, 7], Athanasios Kousathanas[6, 7], Jens Bloechler[2], Christoph Leuenberger[3], Thomas Terberger[4], Detlef Jantzen[5], Joachim Burger[2], Daniel Wegmann[1, 7]

[1]University of Fribourg (Switzerland), [2]Institute of Organismic and Molecular Evolution (iOME), Johannes Gutenberg University Mainz (Germany), [3]University of Fribourg (Switzerland), [4]Lower Saxony State Office for Cultural Heritage (Germany), [5]Landesamt fuer Kultur und Denkmalpflege (Germany), [6]University of Lausanne (Switzerland), [7]Swiss Institute of Bioinformatics (Switzerland)

Comparing ancient and modern genomes provides insights into many aspects of population history. However, such comparison are complicated by particularities of ancient DNA (aDNA) such as post-mortem damage (PMD) and low endogenous DNA content resulting in low sequencing depth. In addition, modern data may be a poor reference for ancient diversity, and the analyses of aDNA should thus not rely on modern data to avoid biases.

Here we introduce a probabilistic framework that accounts for these biases to accurately infer within and between individual genetic diversity even from genomes with median depth <1x, as we show using simulations and by downsampling. We achieve this accuracy by explicitly modeling PMD, and by carefully recalibrating base quality scores with a new method that does not rely on modern data, but exploits homozygous regions in the genome. Importantly, our method also allows for an unbiased clustering of heterochronous individuals using Multi-Dimensional Scaling, rather than by projecting ancient individuals onto a PCA spanned by modern diversity. Finally, our framework also allows for genotype calling, which we found to be unbiased and more accurate for aDNA than GATK.

To illustrate the power of our framework, we studied the diversity among soldiers from a colossal Bronze-age battlefield in norther Europe at the banks of the Tollense River in northern Germany. Our findings suggest that these soldiers, while from a large geographic area, likely represented a single ethnic group.

POA-260

Ancient Fennoscandian genomes reveal origin and spread of Siberian ancestry in Europe

Thiseas Christos Lamnidis[1], Kerttu Majander[1, 2, 4], Choongwon Jeong[1, 3], Elina Salmela[4, 1], Anna Wessman[5], Vyacheslav Moiseyev[6], Valery Khartanovich[6], Antti Sajantila[8], Janet Kelso[7], Svante Paabo[7], Paivi Onkamo[9, 4], Wolfgang Haak[1], Johannes Krause[1], Stephan Schiffels[1]

[1]Max Planck Institute for the Science of Human History (Germany), [2]University of Tuebingen (Germany), [3]Max Planck Institute for the Science of Human History (Germany), [4]University of Helsinki (Finland), [5]University of Helsinki (Finland), [6]Russian Academy of Sciences (Russian Federation), [7]Max Planck Institute for Evolutionary Anthropology (Germany), [8]University of Helsinki (Finland), [9]University of Turku (Finland)

European history has been shaped by migrations of people, and their subsequent admixture. Recently, evidence from ancient DNA has brought new insights into migration events that could be linked to the advent of agriculture, and possibly to the spread of Indo-European languages. However, little is known so far about the ancient population history of north-eastern Europe, in particular about populations speaking Uralic languages, such as Finns and Saami. Here, we analyse genome-wide data from 11 ancient individuals from Finland and Northwest Russia, as well as the complete genome of a modern Saami individual. By modelling these ancient samples as a mixture between European and East Siberian ancestry, we find that Siberian ancestry was most prevalent in our oldest samples from 3,500 years ago. We suggest that these individuals represent early evidence for migrations from Siberia into Europe, and that Siberian ancestry was subsequently admixed into many modern populations in the region, in particular populations speaking Uralic languages today. In addition, we show that ancestors of modern Saami inhabited a larger territory during the Iron Age than today, which adds to historical and linguistic evidence for the population history of Finland.

POA-261

Late Pleistocene North African genomes show deep genetic relationship with ancient Near East and sub-Saharan Africa

Marieke Sophia van de Loosdrecht[1], Abdeljalil Bouzouggar[2], Louise Humphrey[3], Cosimo Posth[1], Nick Barton[4], Ayinuer Aximu-Petri[5], Birgit Nickel[5], Jean-Jacques Hublin[5], Svante Paabo[5], Stephan Schiffels[1], Matthias Meyer[5], Wolfgang Haak[1], Choongwon Jeong[5], Johannes Krause[5]

[1]Max-Planck-Institute for the Science of Human History, Jena (Germany), [2]Institut National des Sciences de l'Archeologie et du Patrimoine, Rabat (Morocco), [3]The Natural History Museum, London (United Kingdom), [4]University of Oxford, Oxford (United Kingdom), [5]Max-Planck-Institute for Evolutionary Anthropology, Leipzig (Germany)

North Africa, connecting sub-Saharan Africa and Eurasia, is important for understanding human history. However, the genetic history of modern humans in this region is largely unknown before the introduction of agriculture. After the Last Glacial Maximum people associated with the Iberomaurusian culture inhabited a wide area spanning from Morocco to Libya. The Iberomaurusian is part of the early Later Stone Age and characterized by a distinct microlithic bladelet technology, complex hunter-gathering and tooth evulsion.

Here we present genomic data from seven individuals, directly dated to ~15,000-year-ago, from Grotte des Pigeons, Taforalt in Morocco. We find a strong genetic affinity of the Taforalt individuals with ancient Near Easterners, best represented by ~12,000 year old Levantine Natufians, that made the transition from complex hunter-gathering to more sedentary food production. This suggests that genetic connections between Africa and the Near East predate the introduction of agriculture in North Africa by several millennia. Notably, we do not find evidence for gene flow from Paleolithic Europeans into the ~15,000 year old North Africans as previously suggested based on archaeological similarities. Finally, the Taforalt individuals derive one third of their ancestry from sub-Saharan Africans, best approximated by a mixture of genetic components preserved in present-day West Africans (Yoruba, Mende) and East Africans (Hadza). In contrast, modern North Africans have a much smaller sub-Saharan African component with no apparent link to Hadza. Our results provide the earliest direct evidence for genetic interactions between modern humans across Africa and Eurasia.

POA-262

The evolutionary history of human cancer genes

Jose Maria Heredia-Genestar[1], David Juan[1], Tomas Marques-Bonet[1, 2, 3], Arcadi Navarro[1, 2, 3]

[1]Pompeu Fabra University (Spain), [2]CNAG-CRG, Centre for Genomic Regulation (CRG), Barcelona Institute of Science and Technology (BIST) (Spain), [3]Institucio Catalana de Recerca i Estudis Avancats (ICREA) (Spain)

Genes associated with cancer in humans are involved in essential cellular processes, especially cell cycle regulation and DNA damage repair. The importance of the role of these genes might suggest they are highly conserved between species, but previous studies have shown that, although the critical domains are conserved, some of these genes have been under selection in the primate lineage, possibly due to their role in viral infection immunity. Despite these intriguing observations, the relevance and extent of the genomic variability generated by these phenomena remains unexplored.

In this project, we intend to investigate the relationship between cancer mutations and the genomic variants in human and great ape populations. To this aim, we studied the landscape of mutations in more than 2,500 human cancers generated by the Pan Cancer Project. We compared these cancer mutations against Chimpanzee, Bonobo and Gorilla population data from the Great Ape Genome Project. First, we analyzed the diversity and functional impact of great apes variants in the orthologues of a panel of human cancer predisposition genes. In addition, we also compared the landscape of passenger mutations across the genome in human tumors and the genomic variability of healthy human and primate populations. Surprisingly, our analyses show that cancer and great apes populations present high genomic variabilities in many regions that remain unchanged in human populations. Our results imply that human tumors resurrect ancient primate mutation hotspots that show low diversity in modern human populations.

POA-264

Demographic processes in Estonia from Bronze Age through Iron Age to Medieval times.

Mait Metspalu[1], Lehti Saag[1, 2], Kristiina Tambets[1], Alena Kushniarevich[1], Liivi Varul[3], Jyri Parik[1], Martin Malve[4], Heiki Valk[4], Lauri Saag[1], Valter Lang[4], Aivar Kriiska[4], Richard Villems[1, 2], Toomas Kivisild[5, 1], Christiana Lyn Scheib[1, 5]

[1]University of Tartu, Institute of Genomics (Estonia), [2]University of Tartu, Institute of Cell and Molecular Biology (Estonia), [3]Tallinn University (Estonia), [4]Institute of History and Archaeology, University of Tartu (Estonia), [5]University of Cambridge (United Kingdom)

N3 and R1a are the two most common Y chromosome haplogroups among modern Estonians. R1a appears with Corded Ware culture but the arrival of hg N has not been determined. To this end we have extracted and studied aDNA from teeth of 18 individuals bracketing the changes in the material culture in the end of the Bronze and early Iron Age. We find N3 in Iron Age but not in Bronze Age. Due to the small sample size we cannot refute the existence of hg N in the latter. In genome-wide analyses the Bronze Age and especially Iron Age samples appear very similar to modern Estonians implying population continuity.

Christianization (13 cc AD) established a new elite of West European origin, which presumably had an impact on the genetic structure of the local population. To investigate this we extracted DNA from teeth of 35 individuals, who have been uncovered from both rural (considered local Estonian population) and town (likely of West European origin) cemeteries of Estonia. We compared the low coverage genomes with each other and with relevant modern and ancient Estonian and other European populations. We find that there is a clear discontinuity between the elite and common people, where the former group genetically with modern German samples and the latter with modern Estonians. We do find three individuals of mixed genetic ancestry. But importantly we do not see a steady shift of either local population strata, which suggests limited contact between the elite and the common people.

POA-265

Unravelling the Estonian genome: the whole is greater than the sum of its parts

Davide Marnetto[1], Francesco Montinaro[1, 2], Lauri Saag[1], Reedik Magi[3], Mait Metspalu[1], Luca Pagani[1, 4]

[1]Institute of Genomics, University of Tartu (Estonia), [2]University of Oxford (United Kingdom), [3]Institute of Genomics, University of Tartu (Estonia), [4]University of Padova (Italy)

Recent advances in the field of ancient human genetics showed that the gene pool of Estonians, as elsewhere in Europe, is the sum of various evolutionary histories that met in the region after the Ice Age. Each of these human groups, however, carried different genetic variants with potentially different consequences on the physiology and genetic makeup of modern Estonians. In this study we introduce the general outline and preliminary results of a recently funded research project, where the whole genome sequences of 2500 modern Estonians generated at the Estonian Genome Center will be subdivided into their ancestral components. These components will be screened for their burden of disease linked variants also taking into account combinatorial effects given by variants occurring on the same haplotype. This annotation will leverage the results of a collection of GWAS, several of which performed on the very population of interest, integrating lower level phenotype annotations as tissue specific eQTL data. Given the amount of data, this approach is expected to shed light on the evolutionary bases that form the present disease burden in the Estonian region.

POA-270

Detecting polygenic adaptation in human history

Fernando Racimo[1]

[1]University of Copenhagen (United States)

An open question in evolution is the importance of polygenic adaptation: adaptive changes in the mean of a multifactorial trait due to shifts in allele frequencies across many loci. In recent years, several methods have been developed to detect polygenic adaptation using loci identified in genome-wide association studies (GWAS). Though powerful, these methods suffer from limited interpretability: they can detect which sets of populations have evidence for polygenic adaptation, but are unable to reveal where in the history of multiple populations these processes occurred. To address this, we created a method to detect polygenic adaptation in an admixture graph, which is a representation of the historical divergences and admixture events relating different populations through time. We apply this method to a dataset containing both present-day genomes and ancient genomes, which allow us to distinguish when and where there were episodes of polygenic adaptation in the complex history of interbreeding among multiple human populations. We provide evidence that variants associated with several traits have been influenced by polygenic adaptation in different populations during recent human evolution.

POA-271

Pseudogenization of PON1 in Marine Mammals Implies Sensitivity to Organophosphate Pesticides

Jerrica Mae Jamison[3], Wynn Meyer[1], Clement Furlong[2], Rebecca Richter[2], Nathan Clark[1]

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

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

POA-273

Phylodynamic assessment of intervention strategies for the West African Ebola virus outbreak

Simon Dellicour[1], Guy Baele[1], Gytis Fudas[2], Nuno R. Faria[3], Oliver G. Pybus[3], Marc A. Suchard[4, 5, 6], Andrew Rambaut[7, 8], Philippe Lemey[1]

[1]KU Leuven - University of Leuven (Belgium), [2]Fred Hutchinson Cancer Research Center (United States), [3]University of Oxford (United Kingdom), [4]University of California (United States), [5]University of California (United States), [6]University of California (United States), [7]University of Edinburgh (United Kingdom), [8]National Institutes of Health (United States)

The recent Ebola virus (EBOV) outbreak in West Africa witnessed considerable efforts to obtain viral genomic data as the epidemic was unfolding. Such data are critical for the investigation of viral molecular epidemiology and can complement contact tracing by public health agencies. Analysing the accumulated EBOV genetic data can also deliver important insights into epidemic dynamics, as demonstrated by a recent viral genome study that revealed a metapopulation pattern of spread. Although metapopulation dynamics were critical for connecting rural and urban areas during the epidemic, the implications for specific intervention scenarios remain unclear. Here, we address this question using a collection of phylodynamic approaches. We show that long-distance dispersal events were not crucial for epidemic expansion and that preventing viral lineage movement to single locations would, in most cases, have had little impact. In addition, urban areas - specifically those encompassing the three capital cities - represented major 'transit centers' for transmission chains, but preventing viral lineage movement to all three simultaneously would have only contained epidemic size to about one third. Using continuous phylogeographic reconstructions we estimate a distance kernel for EBOV spread and reveal considerable heterogeneity in dispersal velocity through time. We also show that announcements of border closures were followed by a significant but transient effect on international virus dispersal. Our study illustrates how phylodynamic analyses can answer specific epidemiological and epidemic control questions and can be used to quantify the hypothetical impact of intervention strategies as well as the impact of barriers on dispersal frequency.

POA-274

Unprecedented genomic diversity of RNA viruses in arthropods reveals the ancestry of negative-sense RNA viruses

Ci-Xiu Li[1]

[1]Zhejiang Provincial Centre for Disease Control and Prevention (China)

Although arthropods are important viral vectors, the biodiversity of arthropod viruses, as well as the role that arthropods have played in viral origins and evolution, is unclear. Through RNA sequencing of 70 arthropod species we discovered 112 novel viruses that appear to be ancestral to much of the documented genetic diversity of negative-sense RNA viruses, a number of which are also present as endogenous genomic copies. With this greatly enriched diversity we revealed that arthropods contain viruses that fall basal to major virus groups, including the vertebrate-specific arenaviruses, filoviruses, hantaviruses, influenza viruses, lyssaviruses, and paramyxoviruses. We similarly documented a remarkable diversity of genome structures in arthropod viruses, including a putative circular form, that sheds new light on the evolution of genome organization. Hence, arthropods are a major reservoir of viral genetic diversity and have likely been central to viral evolution.

POA-275

Single-virion Sequencing of Lamivudine Treated HBV Populations Reveal Population Evolution Dynamics and Demographic History

Yuan Zhu[1], Pauline Aw[1], Paola de Sessions[1], Shuzhen Hong[2], Xian See Lee[2], Lewis Hong[2], Andreas Wilm[1], Chen Hao Li[1], Stephane Hue[3], Seng Gee Lim[4], Niranjan Nagarajan[1], William Burkholder[2], Martin Hibberd[4, 1]

[1]Genome Institute of Singapore (Singapore), [2]Institute of Molecular and Cell Biology (Singapore), [3]London School of Hygiene and Tropical Medicine (United Kingdom), [4]National University Hospital (Singapore)

Viral populations are complex, dynamic, and fast evolving. The evolution of groups of closely related viruses in a competitive environment is termed quasispecies. To fully understand the role that quasispecies play in viral evolution, characterizing the trajectories of viral genotypes in an evolving population is the key. In particular, long-range haplotype information for thousands of individual viruses is critical; yet generating this information is non trivial. Popular deep sequencing methods generate relatively short reads that do not preserve linkage information, while third generation sequencing methods have higher error rates that make detection of low frequency mutations a bioinformatics challenge. Here we applied BAsE Seq, an Illumina based single virion sequencing technology, to eight samples from four chronic hepatitis B (CHB) patients, once before antiviral treatment and once after viral rebound due to resistance. We obtained 248 to 8,796 single-virion sequences per sample, which allowed us to find evidence for both hard and soft selective sweeps. We were able to reconstruct population demographic history that was independently verified by clinically collected data. We further verified four of the samples independently through PacBio SMRT and Illumina Pooled deep sequencing. Overall, we showed that single virion sequencing yields insight into viral evolution and population dynamics in an efficient and high throughput manner.

POA-277

A brief history of papillomaviruses: on the origin and evolution of (onco)genes and genomes

Anouk Willemsen[1], Ignacio G. Bravo[1]

[1]National Center for Scientific Research (CNRS) (France)

Papillomaviruses (PVs) have a wide host range, infecting mammals, birds, turtles and snakes. The recent discovery of PVs in fish has challenged our understanding on the origin of these viruses. We have thus set off on a global analysis of the evolutionary history of the viral family, combining molecular dating, fossil records estimates and reconstruction of large insertion/deletion/recombination events. We have paid special attention to the PV oncogenes, which have followed different evolutionary histories than the PV backbone.

We can date back the most recent common ancestor of the PV backbone to 429 (95% HPD 408-451) million years ago (Mya). By performing common ancestry tests on the origin of the oncogenes, we found that the *E6* and *E7* oncogenes share a common ancestor dating back to 231 Mya (95% HPD 199-264). The *E5* oncogenes, do not appear to share a common ancestor. *E5* rather seems to have evolved through *de novo*gene birth at a specific region in the PV genome. The entrance of *E5* in the PV clade infecting primates concurred with an event that was instrumental for the differential oncogenic potential of present-day PVs infecting humans.

To better understand how PVs infecting primates became oncogenic we are currently combining computational and experimental approaches, including ancestral gene resurrection, where we test different hypotheses on the functions of the oncogenes. Our ultimate aim is to understand why a few PVs are oncogenic for a few host species, while most PVs cause asymptomatic infections in most hosts.

POA-278

The natural evolution of influenza virus hemagglutinin becomes entrenched by a complex epistatic network

Nicholas C. Wu[1], Andrew J. Thompson[2], Jia Xie[3], Chih-Wei Lin[3], Corwin M. Nycholat[2], Xueyong Zhu[1], Richard A. Lerner[3, 4], James C. Paulson[2, 5], Ian A. Wilson[1, 4]

[1]The Scripps Research Institute (United States), [2]The Scripps Research Institute (United States), [3]The Scripps Research Institute (United States), [4]The Scripps Research Institute (United States), [5]The Scripps Research Institute (United States)

The hemagglutinin (HA) receptor-binding site (RBS) in human influenza A viruses is constantly evolving, primarily as a result of the inexorable struggle to evade the immune system. At the same time, the evolution of RBS is constrained by its critical function in virus attachment to host cells. Over the past five decades, an unexpectedly large number of substitutions have emerged and been fixed in the HA RBS of human H3N2 viruses as a result of continuous antigenic drift. From large-scale mutagenesis experiments, we find that RBS substitutions become integrated into an extensive epistatic network that prevents substitution reversion during the natural evolution of the HA RBS. X-ray structural analysis further reveals the mechanistic consequences of this evolutionary trajectory that changes the mode of receptor binding. Whether such evolutionary entrenchment limits future options for immune escape or adversely affect long-term viral fitness is a fascinating ongoing question in the almost 50-year longevity of H3N2 viruses in the human population.

POA-280

Identifying novel viruses associated with Antarctic pinnipeds

Adele Crane[1], Mike Goebel[2], Simona Kraberger[3], Anne Stone[4, 5], Arvind Varsani[3, 6]

[1]Arizona State University (United States), [2]NOAA/National Marine Fisheries Service (United States), [3]Arizona State University (United States), [4]Arizona State University (United States), [5]Arizona State University (United States), [6]University Of Cape Town (South Africa)

Viral diversity in Antarctic wildlife remains largely unknown, and increased anthropogenic activity in this region raises the risk of transmission of new pathogens between humans and wildlife. As such, our investigation aims to characterize novel viral species in an isolated environment and contribute to ongoing surveillance efforts. As a pilot project, we used a viral metagenomic approach to investigate viral diversity in buccal swab samples from Antarctic fur seals (Arctocephalus gazella) breeding on Livingston Island, Antarctica during the 2016/2017 field season. We identified two novel lineages of anelloviruses, which are closely related to an anellovirus previously recovered from California sea lions (Zalophus californianus; ~60% genome-wide pairwise identity). Additionally, a diverse group of anelloviruses were recently identified in Weddell seal (Leptonychotes weddellii) samples around the Ross Island in Antarctica. This research contributes to the identification of viruses associated with pinniped and expands our knowledge on Antarctic animals. In this ongoing study, we aim to build a larger dataset of specific viral groups, and further analyze these viral genomes to determine rates of intra-species recombination, as well as frequency and distribution within coding regions of sites displaying evidence of selection.

POA-281

Domain-based evolutionary analysis of HIV-1 Pol proteins using sequence similarity networks

Shohei Nagata[1, 2], Junnosuke Imai[1], Gakuto Makino[1], Masaru Tomita[1, 2, 3], Akio Kanai[1, 2, 3]

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

Human immunodeficiency virus 1 (HIV-1), the etiological agent of acquired immune deficiency syndrome, have been used as model systems to understand the patterns and processes of molecular evolution because they have high mutation rates and are highly genetically diverse. While HIV is classified into several groups and subtypes, it has been difficult to use its diverse sequences to establish the overall phylogenetic relationships of different strains or the trends in sequence conservation with the construction of phylogenetic trees. Our aims were to systematically characterize HIV-1 subtype evolution and to identify the regions responsible for subtype differentiation at the amino acid level in the Pol protein, which is often used to classify the HIV-1 subtypes. In this study, we systematically characterized the mutation sites in 2,052 Pol proteins from HIV-1 group M (144 subtype A; 1,528 subtype B; 380 subtype C), using sequence similarity networks. Because the Pol protein has several functional domains, we identified the regions that are discriminative by comparing the structures of the domain-based networks. Our results suggest that sequence changes in the RNase H domain and the reverse transcriptase (RT) connection domain are responsible for the subtype classification. By analyzing the different amino acid compositions at each site in both domain sequences, we found that a few specific amino acid residues represent the differences among the subtypes. These residues were located on the surface of the RT structure and in the vicinity of the amino acid sites responsible for RT enzymatic activity or function.

POA-282

Genomic and phylogenetic study of feline paramyxovirus

Shoichi Sakaguchi[1, 2], Satomi Mitsuhashi[3, 4], Makoto Ogawa[5], Takayuki Miyazawa[2], Tadashi Imanishi[3], So Nakagawa[3], Tetsuya Mizutani[1]

[1]Tokyo University of Agriculture and Technology (Japan), [2]Kyoto University (Japan), [3]Tokai University (Japan), [4]Yokohama City University (Japan), [5]Ogawa Pet Clinic (Japan)

Feline paramyxovirus (FPaV) was firstly reported in 2015 in Germany. The research group detected a partial L gene of this virus from cat urine. Because FPaV has a 72-74 % sequence identity with bat's and rodent's paramyxoviruses, they probably emerged by an interspecies transmission. There was no report other than the first report of FPaV. In the present study, we detected FPaV from Japanese cat's urine and revealed a large proportion of this viral genome.

RNA was extracted from a urine sample and treated with Ribosomal RNA Removal Kit. The cDNA obtained by reverse transcription was subjected to metagenomic sequencing. By de novo assembly using CLC Genomics Workbench, one long contig was obtained (16,767 bp). The contig included partial but nearly complete FPaV genome; partial N, P/C/V, M, F, unknown open reading frame (ORF), G, L and trailer sequence. Phylogenetic analysis of whole L gene revealed that FPaV is phylogenetically clustered with Miniopterus schreibersii paramyxovirus, J-virus and Tailam virus. The FPaV genome structure was similar to that of the paramyxoviruses but lacked one or two ORFs.

Since FPaV was detected only from stray cats, this virus is considered to be transmitted outdoors. It is possible that the FPaV is originated from bat or rodent viruses; however, it seems that FPaV is transmitted horizontally between cats because it is shed into the cat urine. In the future, we would like to clarify the detailed evolutionary relationships between FPaV and phylogenetically related viruses.

POA-283

Evolution of influenza virus matrix 2 protein

Hideaki Moriyama[1]

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

Influenza virus can infect various animal species, including humans. Although most influenza virus infections involve mild symptoms, genetic shift, drift, and assortment events have been shown to have resulted in highly pathogenic strains. To date, four influenza virus species have been identified, namely A, B, C, and D. Type A infects several species, including humans as well as porcine, bovine, and canine species. Types B and C infect humans and pigs. Type D is a relatively newly identified type of influenza virus, which has been found to infect cattle and pigs. The influenza A and B virus matrix 2 (M2) protein is a pH-gated proton channel. The M2 protein is involved in the release of viral RNPs from the endosome. In the process, acidification of the external environment around the virus activates the M2 proton channel capability, leading to virus rupture. The influenza A virus M2 protein adopts a homo-tetramer configuration, and the middle of the assembly contains a pore. The valve residues His and Trp face inward within the pore on each monomer. The influenza C and D virus M2 protein is reported serve as a voltage chloride ion channel. The influenza C virus M2 protein seem to take a helical bundle, but 3D-structure is not yet known. The influenza D virus M2 protein shares more sequence similarity with type C virus M2 protein than with sequences of type A or B M2 protein. Predicted evolution process and function-shift of the influenza virus M2 protein will be discussed.

POA-284

Comparing influenza's evolution across within- and between-host scales

Katherine S Xue[1, 2], Jesse Bloom[1, 2]

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

The rapid global evolution of influenza viruses begins with de novo mutations that arise in individual infected hosts. Recent advances in high-throughput deep sequencing have made it increasingly possible to measure within-host genetic diversity, and we have previously shown that influenza viruses can evolve rapidly in chronic infections in ways that mirror global viral evolution. However, major questions remain about how genetic drift, purifying selection, and positive selection combine to shape influenza's evolution within hosts during more typical, acute infections. Moreover, it remains unclear how within-host genetic diversity contributes to influenza's global evolution. Here, we seek to identify the evolutionary forces that act on influenza viruses within hosts by aggregating deep-sequencing data for more than 500 individual infections from four published studies of influenza's within-host genetic diversity. In doing so, we develop the most comprehensive analysis to date of genetic diversity in acute human influenza infections. We compare the spectrum of within-host genetic variation to neutral expectations of genetic diversity to identify the influence of purifying and positive selection upon populations of influenza viruses within hosts. We also compare influenza's within-host genetic diversity with its global genetic variation to determine how transmission bottlenecks and host immunity shape evolution at the within- and between-host scales. Our analyses illuminate how evolutionary forces act across interlocking scales of space and time.

POA-285

Molecular evolutionary analysis of Ebola virus glycoprotein identified two amino acid mutations that affect viral infectivity

So Nakagawa[1, 2], Mahoko Takahashi Ueda[2], Yohei Kurosaki[3], Yusuke Nakano[4], Taisuke Izumi[4], Olamide K Oloniniyi[3], Jiro Yasuda[3], Yoshio Koyanagi[4], Kei Sato[4, 5]

[1]Tokai University School of Medicine (Japan), [2]Tokai University (Japan), [3]Institute of Tropical Medicine (NEKKEN), Nagasaki University (Japan), [4]Institute for Frontier Life and Medical Sciences, Kyoto University (Japan), [5]Japan Science and Technology Agency (Japan)

Ebola virus (EBOV) is extremely virulent, and its glycoprotein is necessary for viral entry. EBOV may adapt to its new host humans during outbreaks by acquiring mutations especially in glycoprotein, which allows EBOV to spread more efficiently. To identify these evolutionary selected mutations and examine their effects on viral infectivity, we adopted experimental-phylogenetic-structural interdisciplinary approaches. In evolutionary analysis of all available Zaire ebolavirus glycoprotein sequences, we detected two codon sites under positive selection, which are located near/within the region critical for the host-viral membrane fusion, namely alanine-to-valine and threonine-to-isoleucine mutations at 82 (A82V) and 544 (T544I), respectively. The transmission dynamics of EBOV Makona variants that caused the 2014-2015 outbreak revealed that A82V mutant was fixed in the population while T544I was not. Reston virus (RESTV), which also belongs to the genus Ebolavirus, causes lethal disease only in non-human primates. Interestingly, RESTV GP possesses A and I at positions 83 and 545, corresponding to positions 82 and 544 of EBOV GP, respectively. To investigate the molecular function of these amino acid replacements, we performed viral pseudotyping experiments with EBOV and RESTV GP derivatives in 10 cell lines from nine mammalian species. We demonstrated that isoleucine at position 544/545 increases viral infectivity in all host species, whereas valine at position 82/83 modulates viral infectivity depending on the viral and host species. Structural modeling suggested that the former residue affects viral fusion, whereas the latter residue influences the interaction with the viral entry receptor, Niemann-Pick C1.

POA-286

Non-retroviral virus-like elements in eukaryotic genomes

Kirill Kryukov[1], Mahoko Takahashi Ueda[2], Tadashi Imanishi[1], So Nakagawa[1, 2]

[1]Tokai University School of Medicine (Japan), [2]Tokai University (Japan)

Retro-transcribing viruses are a known source of eukaryotic DNA, contributing a sizeable fraction of eukaryotic genome via endogenization. However recent reports show examples of non-retroviral viral DNA endogenized in several eukaryotic genomes. Intrigued by these findings, we set out to systematically survey non-retroviral integrations in eukaryotes. We developed a system for detecting endogenous viral element (EVE)-like sequences in eukaryote genomes, and conducted a large scale nucleotide sequence similarity search using all available viral (excluding retroviruses) and eukaryotic genome assemblies stored in the NCBI genome database. We found that more than half of analyzed viruses have strong similarity to eukaryote sequence, and that nearly all eukaryote genomes harbor EVE-like sequences. Other than endogenization of viral sequence in host genome, horizontal transfer can also occur in the other direction - from host to virus. Our molecular phylogenetic analysis confirms the previously reported cases of host-to-virus DNA transfer. We found that viruses often exchange DNA not only with their known hosts, but also with distantly related eukaryotes, suggesting a possibility that a previously undescribed mechanism of horizontal transfer may be involved. We constructed a database, Predicted Endogenous Viral Elements (pEVE, http://peve.med.u-tokai.ac.jp/), which provides comprehensive search results summarized from an evolutionary viewpoint. We believe that our search system and database will facilitate future studies on function and evolution of EVEs.

POA-287

Whole genome diversity of inherited chromosomally integrated HHV-6 derived from healthy individuals of diverse geographic origin

Marco Telford[1], Arcadi Navarro[1, 2, 3], Gabriel Santpere[1, 4]

[1]UPF/Institute of evolutionary biology (IBE) (Spain), [2]Catalan Institution for Advanced Research Studies (ICREA) (Spain), [3]Center for Genomic Regulation (Spain), [4]Yale school of medicine (Spain)

Human herpesviruses 6 -A and -B (HHV-6A, HHV-6B) are ubiquitous in human populations worldwide. These viruses have been associated with several diseases such as multiple sclerosis, Hodgkin's lymphoma or encephalitis. Despite of the need to understand the genetic diversity and geographic stratification of these viruses, the availability of complete viral sequences from different populations is still limited. Here, we present nine new inherited chromosomally integrated HHV-6 sequences from diverse geographical origin which were generated through target DNA enrichment on lymphoblastoid cell lines derived from healthy individuals. Integration with available HHV-6 sequences allowed the assessment of HHV-6A and -6B phylogeny, patterns or recombination and signatures of natural selection. Analysis of the intra-species variability showed differences between A and B diversity levels and revealed that the HHV-6B reference (Z29) is an uncommon sequence, suggesting the need for an alternative reference sequence. Signs of geographical variation are present and more defined in HHV-6A, while they appear partly masked by recombination in HHV-6B. Finally, we conducted a scan for signatures of selection in protein coding genes that yielded at least 6 genes (4 and 2 respectively for the A and B species) showing significant evidence for accelerated evolution, and 1 gene showing evidence of positive selection in HHV-6A.

POA-291

Host network structure and fitness effects shape the emergence and spreading of new mutations in viruses

Kent Kawashima[1, 2], Hiroshi Akashi[1, 2]

[1]SOKENDAI (The Graduate University for Advanced Studies) (Japan), [2]National Institute of Genetics (Japan)

We present a model that attempts to bridge the gap between population genetics and epidemiology to show the importance of host networks on the molecular evolution of infectious disease pathogens. Within-host mutation, selection, genetic drift, and recombination have often been cited as the major forces that shape the evolution of viruses. The effect of migration via pathogen transmission between hosts has received less attention. One reason for this bias is the lack of models that combine population genetic theory with epidemiological models of disease spreading. Here, we introduce a general model for virus evolution to study both the effect of fitness differences and effects of contact and transmission network topology on the frequency of new mutations. Through computer simulations, we found that host connection variance affects the emergence and spreading of new mutations, especially under neutral evolution. By tracking mutation dynamics, we found the evolutionary trajectories of new mutations are affected not only by their fitness effects, but also by the location of their host in the contact network: mutations appearing during infection of hosts near or within susceptible high-contact hosts were more likely to spread than mutations emerging from infections at sparsely connected areas of the host network. These results demonstrate how "superspreaders" - hosts that disproportionately spread the disease - can play an important role in the fixation of new virus mutations and the creation of new strains.

POA-293

Mapping the drivers of within-host pathogen evolution using massive data sets

Duncan Palmer[1, 2, 3], Isaac Turner[1, 2], Sarah Fidler[4], John Frater[3, 5, 6], Dominique Goedhals[7], Philip Goulder[8, 9], Kuan-Hsiang Gary Huang[10, 5], Annette Oxenius[11], Rodney Phillips[12, 3, 5], Roger Shapiro[13, 14], Cloete van Vuuren[7], Angela McLean[3, 15], Gil McVean[1, 2, 16]

[1]University of Oxford (United Kingdom), [2]University of Oxford (United Kingdom), [3]University of Oxford (United Kingdom), [4]Imperial College (United Kingdom), [5]University of Oxford (United Kingdom), [6]University of Oxford (United Kingdom), [7]University of KwaZulu-Natal (South Africa), [8]University of the Free State (South Africa), [9]University of Oxford (United Kingdom), [10]Einstein Medical Center Philadelphia (United States), [11]Swiss Federal Institute of Technology Zurich (Switzerland), [12]UNSW (Australia), [13]Harvard TH Chan School of Public Health (Botswana), [14]Harvard TH Chan School of Public Health (United States), [15]University of Oxford (United Kingdom), [16]University of Oxford (United Kingdom)

Differences among hosts, resulting from genetic variation in the immune system or heterogeneity in drug treatment, can impact within-host pathogen evolution. Identifying such interactions can potentially be achieved through genetic association studies. However, extensive and correlated genetic population structure in hosts and pathogens presents a substantial risk of confounding analyses. Moreover, the multiple testing burden of interaction scanning can potentially limit power. To address these problems, we have developed a Bayesian approach for detecting host influences on pathogen evolution that makes use of vast existing data sets of pathogen diversity to improve power and control for stratification. The approach models key processes, including recombination and selection, and identifies regions of the pathogen genome affected by host factors. Using simulations and empirical analysis of drug-induced selection on the HIV-1 genome we demonstrate the power of the method to recover known associations and show greatly improved precision-recall characteristics compared to other approaches. We build a high-resolution map of HLA-induced selection in the HIV-1 genome, identifying novel epitope-allele combinations.

POA-294

Non-stationary evolution of Influenza A surface proteins

Anfisa Popova[2], Alexey Neverov[2], Georgii Bazykin[1]

[1]Institute for Information Transmission Problems of the Russian Academy of Sciences (Russian Federation), [2]Central Research Institute for Epidemiology (Russian Federation)

Influenza A virus is a major public health problem and a pandemic threat. Its evolution is largely driven by diversifying positive selection, so that relative fitness of different amino acid variants changes with time due to changes in herd immunity or genomic context, and novel amino acid variants attain fitness advantage. However, diversifying selection is also expected to have another manifestation: the fitness associated with a particular amino acid variant should decline with time since its origin as the herd immunity adapts to it. Here, we study the evolution of antigenic sites of Influenza A surface proteins, and show that an amino acid variant becomes progressively more likely to become replaced by another variant with time since its origin - a phenomenon we term senescence. Senescence is particularly pronounced at experimentally validated antigenic sites, implying that it is largely driven by host immunity. By contrast, at internal sites, existing variants become more favourable with time due to arising compensatory mutations. Our findings reveal a previously undescribed facet of adaptive evolution, and suggest novel approaches for prediction of pathogen evolutionary dynamics.

POA-295

FAVITES: A framework for the simulation of compatible viral transmission networks, phylogenetic trees, and sequences

Niema Moshiri[1], Siavash Mirarab[2]

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

Motivation: Reconstructing HIV transmission networks can greatly enhance epidemic intervention, but transmission network reconstruction methods have various limitations, and their accuracies are poorly understood, largely because it is difficult to obtain "truth" sets on which to test them and properly measure their performance.

Results:
We introduce FAVITES, a robust framework for simulating realistic contact networks, transmission networks, phylogenetic trees, and sequences. We then perform simulation experiments to study the accuracy and scalability of multiple transmission network reconstruction methods.

Availability and implementation: FAVITES was written in Python 3, but its modules wrap around multiple Linux tools. They have all been tested in Linux and Mac OS X. FAVITES can be found on GitHub (https://github.com/niemasd/FAVITES), and a Docker image with all dependencies can be found on DockerHub (https://hub.docker.com/r/niemasd/favites).

POA-299

Neanderthal ancestry in modern-day humans provide clues for the pattern of Neanderthal-human admixture in the past

Fernando A. Villanea[1], Joshua G. Schraiber[1]

[1]Temple University (United States)

Neanderthals and humans overlapped geographically for a period of over 30,000 years following human migration out of Africa. During this period, Neanderthals and humans interbred as evidenced by Neanderthal portions of the genome carried by non-African individuals today. A key observation is that the proportion of Neanderthal ancestry is different between European and East Asian populations. Here, we explore various demographic models that could explain this observation. These include distinguishing between a single admixture event and multiple Neanderthal contributions to either population, and the hypothesis that reduced Neanderthal ancestry in modern Europeans resulted from more recent admixture with a ghost population that lacked a Neanderthal ancestry component (the "dilution" hypothesis). We approach this problem by applying a combination of analytical theory and simulation models. We use machine learning to distinguish between these demographic models based on the outputs of our simulations, compared to empirical data of Neanderthal introgression based on the 1000 genome panel. We find that there is not statistical power to distinguish between a single admixture model and the dilution model. However, we strongly reject a model of a single admixture. Thus, our results support a complex population history of both Europeans and East Asians. The asymmetric pattern of Neanderthal ancestry observed today is the result of secondary gene flow from Neanderthals in East Asia, although we cannot rule out the dilution of Neanderthal ancestry in Europeans by the contribution of a yet-to-be-identified population carrying little to no Neanderthal genomic elements.

POA-301

CLADES: A Classification-based Machine Learning Method for Species Delimitation from Population Genetic Data

Jingwen Pei[1], Chong Chu[1, 2], Xin Li[1], Bin Lu[3], Yufeng Wu[1]

[1]University of Connecticut (United States), [2]Harvard Medical School (United States), [3]Chinese Academy of Science (China)

Species are considered to be the basic unit of ecological and evolutionary studies. Since multilocus genomic data is becoming increasingly available, there has been considerable interests in the use of DNA sequence data to delimit species. In this paper, we show that machine learning can be used for species delimitation. To the best of our knowledge, there exists no species delimitation methods that are based on machine learning. The main idea of our method is viewing the species delimitation problem as a classification problem. It is a problem of identifying the category of a new observation on the basis of training data. Extensive simulation is first conducted over a broad range of evolutionary parameters for training purpose. Each pair of known populations are combined to form training samples with a label of same species or different species. We use Support Vector Machine or SVM to train a classifier using a set of summary statistics computed from training samples as features. The trained classifier can classify a test sample to two outcomes: same species or different species. Given multi-locus genomic data of multiple related organisms or populations, our method, called CLADES, performs species delimitation by first classifying pairs of populations, then delimiting species by maximizing the likelihood of species assignment for multiple populations. CLADES is evaluated through extensive simulation and also tested on real genetic data. We show that CLADES is both accurate and efficient for species delimitation when compared with existing methods.

POA-302

SeleDiff: A fast and scalable tool for testing and estimating selection differences between populations

Xin Huang[1]

[1]Shanghai Institutes for Biological Sciences (China)

Genome-wide scan for natural selection is a classical challenge in human genetics. Currently, there are two popular kinds of approaches for detecting signals of natural selection from human genomes. One is based on extended haplotype homozygosity (EHH); the other is based on genetic diversity. The time complexities of these approaches, however, are quadratic with respect to the number of variants or the number of samples. As more and more variants and samples become available, this quadratic time complexity would limit research community to detect signals of natural selection quickly. Moreover, neither of these approaches could quantify the differences of the strength of natural selection between populations. Here, we implemented a fast and scalable tool called SeleDiff for testing and estimating selection differences between populations. SeleDiff integrates with t-digest, which is a recent developed on-line algorithm in machine learning, in order to detect natural selection efficiently. Using SeleDiff, we can analyze a simulated dataset containing 300,000 variants and 100,000 samples in 15.7 minutes (wall time) with less than four gigabytes of random access memory (OS: Red Hat Enterprise Linux Server release 6.3; CPU: AMD OpteronTM 6174). Running time analysis showed SeleDiff has linear time complexity with respect to both the number of variants and the number of samples. This linear time complexity would make SeleDiff a fast tool for quantifying signals of natural selection in genome-wide scan.

POA-304

Rates of molecular evolution suggest life history and a post-K-Pg nocturnal bottleneck of Placentals

Jiaqi Wu[1], Takahiro Yonezawa[2], Hirohisa Kishino[1]

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

Life history and behavioural traits are often difficult to discern from the fossil record, but evolutionary rates of genes and their changes over time can be inferred from extant genomic data. Under the neutral theory, molecular evolutionary rate is a product of mutation rate and the proportion of neutral mutations. Mutation rates may be shared across the genome, whereas proportions of neutral mutations vary among genes because functional constraints vary. By analysing evolutionary rates of 1,185 genes on a phylogeny of 89 mammals, we extracted historical profiles of functional constraints on these rates in the form of gene-branch interactions. By applying a novel statistical approach to these profiles, we reconstructed the history of 10 discrete traits related to activity, diet and social behaviours. Our results indicate the ancestor of placental mammals was solitary, seasonally breeding, insectivorous and likely nocturnal. The results suggest placental diversification began 10-20 million years before the K-Pg boundary (66 Mya), with some ancestors of extant placental mammals becoming diurnal and adapted to different diets. However, from the Palaeocene to the Eocene-Oligocene transition (EOT, 33.9 Mya), we detect a post-K-Pg nocturnal bottleneck where all ancestral lineages of extant placentals were nocturnal. While diurnal placentals may have existed during the elevated global temperatures of the Paleocene-Eocene Thermal Maximum, we hypothesize that diurnal placentals were selectively extirpated during or after the global cooling of the EOT whereas some nocturnal lineages survived due to preadaptations to cold environments.

POA-306

Fast Approximate Inference for Phylogenetic Reconstruction via Stochastic Variational Inference in Large Data Sets

Tung Thanh Dang [1]

[1] The University of Tokyo (Japan)

Bayesian mixture models for modeling across site variation of the substitution process are now used in a wide variety of applications in phylogenetic reconstruction. Although Monte-Carlo Markov chain (MCMC) sampling techniques make approximate inference possible for both finite and infinite mixture models, the computational burden is prohibitive on the large modern data sets. To overcome this problem, we developed new algorithms for fast and accurate inference of the model underlying the PhyloBayes MPI program approaching variational Bayesian procedures. Variational frameworks convert the problem of approximating posterior distributions into solving a sequence of unconstrained optimization problems. We analyzed empirical large-scale datasets to compare time estimates produced by the variational algorithm with those reported by using MCMC approaches in PhyloBayes MPI. We demonstrated that variational methods achieve accuracy competitive with Markov chain Monte Carlo approaches while requiring orders of magnitude less computational time.

POA-307

Probabilistic modeling of genetic variation reveals protein-protein interactions and the effects of mutations on interactions

Anna Gustafson Green[1], Debora Marks[1]

[1]Harvard Medical School (United States)

Molecular evolution is constrained by specific, high-fidelity interactions between biological molecules. To understand the effects of genetic variation on molecular and organismal phenotype, we must uncover both which proteins interact and which residues mediate these interactions. Recent work has shown that pairwise models of biological sequences can resolve three-dimensional protein structures and predict the effects of mutations. This method, called EVCouplings, infers constraints between pairs of residues solely from evolutionary sequence data. When applied pairs of proteins, this method resolves which proteins interact at residue resolution using only genomic sequence data. I will present recent methodological advances and their application to predict interactions and co-constrained interface residues for large numbers of protein complexes. I will feature inferred interactions in the bacterial elongation and division machinery, which are the targets of beta-lactam antibiotics. I will demonstrate that though the two systems share an evolutionary history, they are characterized by different protein-protein interactions. I will also show how this statistical model successfully predicted mutations that perturb protein-protein interactions in the elongation machinery. This works demonstrates that statistical modeling of genetic variation can infer protein-protein interactions, elucidate the structure of particular protein complexes at residue resolution, and reveal mutations that perturb the interaction of proteins.
This software is freely available on github: github.com/debbiemarkslab

POA-309

Prediction model to infer degree of functionalization based on protein and expression divergence rate in Arabidopsis

Akihiro Ezoe[1], Kazumasa Shirai[1], Kousuke Hanada[1]

[1]Kyushu institute technology (Japan)

There is a large variation of functionalization in duplicate genes. However, we do not know what kinds of duplicate genes with either high or low degree of functionalization (HDF or LDF) tend to be retained in evolution at genomic scale. Based on genes with phenotypic data throughout Arabidopsis knock-out analyses, we generated a prediction model to infer the degree of functionalization throughout protein and expression divergence rate. Among 4017 duplicated genes, we identified 1052 HDF and 600 LDF. To validate the prediction model, we examined the overlap of either Gene Ontology (GO) annotations. Duplicate genes with LDF tend to have a higher rate of overlapping in GO than those with HDF at a significant level, indicating that our prediction works well. To characterize HDF and LDF duplicate genes, we examined overrepresented GO and domains in HDF and LDF. As expected, HDF tends to be associated with immune responses but LDF tends to be associated with ubiquitous function such as ribosome and cell cycle. Interestingly, HDF tends to have fewer number of functional domains than LDF, indicating that HDF and LDF tend to have simple and complicated protein structures, respectively. To address the stability of LDF and HDF duplicate genes, we examined the retention rate of anciently duplicated genes in two close species. The duplicate genes with LDF tends to be retained in relative species of Arabidopsis than HDF. Thus, HDF tends to be species-specific with a simple protein structure. Taken together, genome evolution prefers minor changes of duplicate genes.

POA-310

Learning and interpreting the evolution of the gene regulatory grammar in a deep neural network framework

Ling Chen[1]

[1]Vanderbilt (United States)

Our work uses deep neural networks (DNNs) to address two fundamental questions. The first is scientific: what are the rules of the gene regulatory grammar and how do they evolve? The second is methodological: how can we extract biological insight from accurate, but complex DNNs?

Using genome-wide maps of tens of thousands of enhancer-associated histone modifications, we developed convolutional neural networks (CNNs) to predict liver enhancers across six diverse mammals from their DNA sequences. The CNNs accurately distinguished enhancers from background in each species, and performed substantially better than support vector machines (average auROC 0.85 vs. 0.78). This suggests that higher-order interactions learned by internal layers improve prediction. Next, we applied the human CNN to other species; it performed well across species, indicating that much of the regulatory code is shared among mammals. However, we identified differential usage of many neurons by other species enhancers, which suggests that certain rules of the regulatory grammar vary across species.

The interpretation of features learned by CNNs is challenging. Using simulated and real sequences, we evaluated existing and novel techniques for interpreting sequence patterns learned by the CNN. A novel algorithm that integrates techniques from image processing and sequence analysis best interpreted the model. First layer neurons learned essential liver TFs, like HNF4, and CEBPB. Higher-layer neurons learned different complex patterns that suggest the importance of transposable elements.

These results demonstrate the promise of deep learning in revealing the conservation of the complex combinatorial patterns in regulatory sequences across species.

POA-311

Machine reasoning with phenotypes: enhancing expert knowledge about the genetics of an ancient evolutionary transition

Todd Vision[1], Dahdul Wasila[2], James Balhoff[1], Alex Dececchi[4], Pasan Fernando[2], Hilmar Lapp[3], Paula Mabee[2], Prashanti Manda[5], Kellen Mastick[2], Monte Westerfield[6], Erliang Zeng[2]

[1]University of North Carolina at Chapel Hill (United States), [2]University of South Dakota (United States), [3]Duke University (United States), [4]Queens University (Canada), [5]University of North Carolina at Greensboro (United States), [6]University of Oregon (United States)

Semantically-annotated genetic and phenotypic data can be used for machine reasoning about the genetic basis of phenotypic innovation in the history of life. As a test case for this application of machine reasoning, we examine the Devonian-era transition from aquatic fins to terrestrial limbs in tetrapodamorph vertebrates, a well-studied transition in the fossil record for which at least 162 different candidate genes have been reported in the evo-devo literature with some evidentiary support. We asked to what extent an expert system would recover the same set of candidate genes using only knowledge about the phenotypes of (i) the relevant fossil taxa and (ii) genetic perturbations in vertebrate model organisms. We used the semantic similarity between the two classes of phenotypes as a measure of the strength of a candidate gene association. The similarity between fossil and gene phenotypes was generally higher for candidates relative to non-candidates, but the distributions overlap and numerous exceptions suggest possible refinements to the candidate gene list. To understand why some genes performed strongly counter to expectation, and study the functional relationships among the genes with the highest cross-domain similarities, we examined the clustering of candidates and non-candidates within protein interaction networks. Our results demonstrate the potential of machine reasoning to accurately rank the strength of evidence for candidate genes when presented with a large volume of descriptive phenotype information. This approach could in principle be used to refine candidate gene hypotheses culled from the literature.

POA-312

Proteome-wide evidence for evolutionary signatures of function in highly diverged disordered regions

Taraneh Zarin[1], Alan Michael Moses[1, 2, 3]

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

Intrinsically disordered regions (IDRs) are regions of proteins that do not autonomously fold into stable secondary or tertiary structures. Though they defy the classical view of proteins as rigidly structured macromolecules, IDRs are widespread in eukaryotic proteomes, and are associated with a diverse range of functions in the cell. Interestingly, the majority of IDRs appear to be evolving rapidly at the level of the primary amino acid sequence. This seemingly rapid evolution is facilitated by insertions and deletions, which makes it difficult to quantify evolutionary conservation and infer function in these regions through standard sequence analysis. Recently, we found that highly diverged amino acid sequences can encode conserved phenotypes in an IDR in Saccharomyces cerevisiae, showing that sequence divergence does not necessarily imply functional divergence in these regions. By quantifying bulk molecular features in IDRs and comparing them to our null expectation of disordered region evolution, we are able to identify the molecular features underlying these conserved phenotypes. We apply this method proteome-wide, and find that most IDRs in the budding yeast proteome show evolutionary signatures that deviate from our null expectation of disordered region evolution. Clustering analysis reveals that many IDRs share sets of evolutionary signatures, and that these signatures are associated with highly specific functions. This provides insight into the abundance and persistence of highly diverged IDRs in the proteome, and offers a framework for their classification.

POA-313

Machine learning identifies signatures of host adaptation in the bacterial pathogen Salmonella enterica

Nicole E Wheeler[1, 2], Paul P Gardner[2, 3], Lars Barquist[4]

[1]Wellcome Sanger Institute (United Kingdom), [2]University of Canterbury (New Zealand), [3]University of Otago (New Zealand), [4]University of Wuerzburg (Germany)

Emerging pathogens are a major threat to public health, however understanding how pathogens adapt to new niches remains a challenge. New methods are urgently required to provide functional insights into pathogens from the massive genomic data sets now being generated from routine pathogen surveillance for epidemiological purposes. Here, we measure the burden of atypical mutations in protein coding genes across independently evolved *Salmonella enterica* lineages, and use these as input to train a random forest classifier to identify strains associated with extraintestinal disease. Members of the species fall along a continuum, from pathovars which cause gastrointestinal infection and low mortality, associated with a broad host-range, to those that cause invasive infection and high mortality, associated with a narrowed host range. Our random forest classifier learned to perfectly discriminate long-established gastrointestinal and invasive serovars of *Salmonella*. Additionally, it was able to discriminate recently emerged *Salmonella* Enteritidis and Typhimurium lineages associated with invasive disease in immunocompromised populations in sub-Saharan Africa, and within-host adaptation to invasive infection. We dissect the architecture of the model to identify the genes that were most informative of phenotype, revealing a common theme of degradation of metabolic pathways in extraintestinal lineages. This approach accurately identifies patterns of gene degradation and diversifying selection specific to invasive serovars that have been captured by more labour-intensive investigations, but can be readily scaled to larger analyses.

POA-314

Associating the microbiome with experimental treatment groups, using a random forest, in a model of inflammatory bowel disease

Gurdeep Singh[1], Sheena Cruickshank[1], Andrew Brass[1], Christopher Knight[2]

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

Understanding how the microbiome contributes to host health is a key area of investigation. However, in exploring the relationship between the microbiome and host health, much analysis involves drawing conclusions based on single snapshots of either species or phyla, with limited consideration across taxonomic levels.

Here we develop a community-orientated approach to exploring the role of the microbiome in discrete niches of the gut in a model of colitis. We constructed a phylogenetic tree using 16S ribosomal RNA (rRNA) sequence data derived from the stools and colonic mucus from 40 littermate-controlled, cohoused, healthy wildtype mice and mice that spontaneously develop colitis over time (mdr1a-/- mice), in order to examine the relationships between the taxa present in these samples. We then employed a random forest (RF) model incorporating taxa at all levels of the phylogenetic tree to determine which taxa are most important in distinguishing the different treatment groups: age, genotype and microbial location (mucus versus stools).

We found that the RF discriminated between the age and location of samples with >90% accuracy and has allowed us to identify important taxa that are associated with our different treatment groups. Mucus and stool microbiomes are primarily distinguished by differences in abundant high-level taxa, but more subtle differences among taxa at intermediate phylogenetic scales, for instance the Erysipelotrichaceae family, tend to distinguish different ages of mice. Thus, our methods support more traditional forms of microbiome analysis in the identification of microbial taxa associated with other conditions of interest.

POA-316

Rates of mutation and recombination in Siphoviridae phage genome evolution over three decades

Anne Kupczok[1], Horst Neve[2], Kun D. Huang[1], Marc P. Hoeppner[3], Knut J. Heller[2], Charles M.A.P. Franz[2], Tal Dagan[1]

[1]Kiel University (Germany), [2]Max Rubner-Institut (Federal Research Institute of Nutrition and Food) (Germany), [3]Kiel University (Germany)

The evolution of asexual organisms is driven not only by the inheritance of genetic modification but also by the acquisition of foreign DNA. The contribution of vertical and horizontal processes to genome evolution depends on their rates per year and is quantified by the ratio of recombination to mutation. Here we delineate the contribution of mutation and recombination to dsDNA phage genome evolution. Analyzing 34 isolates of the 936 group of Siphoviridae phages that were sampled using a constant Lactococcus lactis strain from a single dairy over 29 years, we estimate a constant substitution rate of 1.9×10^{-4} substitutions per site per year due to mutation. This substitution rate is within the range of estimates for eukaryotic viruses. The reconstruction of recombination events reveals a constant rate of five recombination events per year and 4.5×10^{-3} nucleotide alterations due to recombination per site per year. The recombination rate thus exceeds the substitution rate to a great extent, resulting in a relative effect of recombination to mutation (r/m) of ~24 that is homogenous over time. Especially in the early transcriptional region, we detect frequent gene loss and regain due to recombination with other phages of the 936 group, demonstrating the role of the 936 group pangenome as reservoir of genetic variation. The observed substitution rate homogeneity conforms to the neutral theory of evolution; hence, we demonstrate that the neutral theory can be applied to phage genome evolution and also to genetic variation brought about by recombination.

POA-317

Resolving ultrametric phylogeny of prokaryotic strains with frequent homologous recombination from the variation of local SNP density on their genomes

Tin Yau Pang[1], Martin Lercher[1]

[1]Heinrich Heine University Duesseldorf (Germany)

Homologous recombination among closely related strains is a major source of DNA sequence variation in prokaryotes. This can lead to large uncertainties in phylogenies inferred using algorithms based on models of purely vertical inheritance. A popular approach to recombination-aware phylogeny reconstruction is to consider the ancestral recombination graph, which identifies recombined genomic segments as outliers. However, this approach requires that the phylogenetic signal is dominated by the clonally inherited part, an assumption frequently violated for closely related strains.

Here, we propose an alternative, fast, recombination-aware algorithm for the reconstruction of ultrametric phylogenetic trees. For each pair of strains, the algorithm calculates the local density of SNPs in short segments of the nucleotide alignment. It then searches for an ultrametric phylogenetic tree together with recombination probabilities that generates theoretical distributions of SNP densities that best match the observed distributions. A corresponding strategy can be formulated for amino acid sequences.

We test the accuracy of our algorithm against current state-of-the-art algorithms, including ClonalFrameML and Gubbins, on simulated and real genomes. For genomes with high levels of recombination, our new algorithm outperforms alternative approaches in terms of branch length prediction, while tree topologies are at least as accurate. For real E. coli genomes, phylogenetic trees calculated with our approach are significantly more consistent with the phylogenetic signal of gene gains and losses than alternative algorithms.

POA-318

Experimentally informed site-specific substitution models deepen phylogenetic estimates of the divergence of viral lineages.

Sarah K. Hilton[1, 2], Jesse D. Bloom [1, 2]

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

Molecular phylogenetics is commonly used to estimate the time since the divergence of modern gene sequences. Such phylogenetic techniques often estimate substantially shallower divergence times than other methods. For instance, in the case of viruses, there is independent evidence, such as viral insertions into the host genome or historical records, that molecular phylogenetics can underestimate deep divergence times. This discrepancy is thought to be caused in part by inadequate models of purifying selection that cause branch-length underestimation. Here we investigate the effect of site-specific substitution models on long branch estimation using the phylogenies of influenza virus hemagglutinin. Hemagglutinin is well suited for questions of long branch estimation because the 18 hemagglutinin subtypes are upwards of 60% diverged on the amino acid level. We show that substitution models informed by experimental measurements of site-specific purifying selection lengthen the branches compared to site-independent Goldman-Yang models. This increase in branch lengths is due to better modeling how site-specific amino-acid preferences affect the stationary state of the substitution models. Furthermore, this increase is independent of the branch-length-extension due to modeling site-to-site variation in substitution rate, such as the common gamma-distributed omega value. However, the improvements from these site-specific models are limited by the inherent tension between the enhanced accuracy of accounting for site-specific amino-acid preferences and the fact that these preferences shift over long evolutionary times. Overall, our work underscores the importance of modeling how site-specific purifying selection affects the stationary state as well as the substitution rate when estimating deep divergence times.

POA-319

The Molecular Clock Winder: Assessing the Effects of Life-history Traits and Reproductive Biology on Substitution Rates in Primates

Lucas Henriques Viscardi[1], Vanessa Rodrigues Paixao-Cortes[2], Guillermo Reales[1], Maria Catira Bortolini[1], Carlos Eduardo Guerra Amorim[3]

[1]Federal University of Rio Grande do Sul (Brazil), [2]Federal University of Bahia (Brazil), [3]University of California, Los Angeles (United States)

Primates are known to present a marked variation in yearly substitution rates, tending to be lowest in great apes, violating some of the assumptions of the Molecular Clock. This is commonly credited to life history trait differences across species. Additionally, differences in life histories across sexes are also thought to impact the pace of the Molecular Clock. Here, we made use of multi-species alignments to estimate genome-wide neutral substitution rates for 16 primate species. We further sought to evaluate possible effects of 27 life history traits (sex-specific or not) on the Molecular Clock to test long standing hypotheses of neutral evolution and divergence. We show that variation in substitution rates in the autosomes correlates negatively with female-to-male sex maturity ratios and generation-time ($p = 0.0043$ and 0.0197 respectively), among other life history traits, even after controlling for phylogenetic relationship with Independent Contrast Analysis. Our data supports the long-standing hypothesis of the *hominoid rate slowdown* and points out to the effect of differences in reproductive biology between sexes on the Molecular Clock. The observed ratio of X-to-autosomal substitution rates for gorillas and humans differs markedly from the expected based on a robust analytical model, but not for chimps. These discrepancies can be explained by changes in life history traits across time. In particular, our data suggest that sperm competition in humans and gorillas may have been more intense in the past, comparable to that of present-day chimps. Finally, we show that male mutation bias is pervasive across primates.

POA-324

Dating the emergence of DNA by dating the origin of the ribonucleotide reductase protein family

Adrien Jules Boniface[1], Timothy M. Vogel[1], Catherine Larose[1]

[1]Ecole Centrale de Lyon, Universite de Lyon (France)

The RNA world hypothesis provides a conceptual framework for the emergence of life from abiotic chemistry and its early evolution but numerous questions remain unanswered, such as the mechanisms and timing of the emergence of DNA and its adoption as the carrier of genomic information. The transition from RNA to DNA likely involved a protein able to synthesize dNTPs (the building blocks of DNA), a process carried by members of the Ribonucleotide Reductase (RNR) protein family found in every modern cell studied to date.

All known RNR proteins are homologous and perform ribonucleotide reduction through the use of a protein-derived cysteinyl free radical. Furthermore, the RNR protein family belongs to the 10-stranded β-α barrel superfamily, whose topologies are generally considered to be ancient. Three different classes of RNR proteins are distinguished based on the mechanism by which the protein-derived radical is generated. All three classes of RNR proteins are found in the three domains of life, therefore the RNR protein family appears to have an ancient origin, probably predating LUCA. We hypothesized that the RNR protein family was responsible for the emergence of dNTPs, and subsequently DNA, in a primitive RNA + protein world.

The aim of this study was to use modern phylogenetic tools and molecular clock methods to provide a date estimate for the origin of the RNR family, which we hypothesized as playing a critical role in the emergence of the DNA world.

POA-329

RepetDB: a resource for unified transposable element references with classification

Joelle Amselem[1], Guillaume Cornut[1], Nathalie Choisne[1], Michael Alaux[1], Francoise Alfama-Depauw[1], Veronique Jamilloux[1], Florian Maumus[1], Thomas Letellier[1], Isabelle Luyten[1], Cyril Pommier[1], Anne-Francoise Adam-Blondon[1], Hadi Quesneville[1]

[1]INRA Universite Paris Saclay (France)

The ability of Transposable Elements (TEs) to move and replicate throughout the genomes makes them perhaps the most important contributors to genome evolution. Their detection and annotation are considered essential, and must be undertaken in the frame of any genome sequencing project.

Only a fully automated TE *de novo* annotation process is able to face the sequence deluge rapidly increasing with the improvement of high-throughput sequencing technologies. This process generally relies on pipelines to *de novo* detect, build and classify consensuses using similarity with sequences already identified and/or according to their structure. However, as any automated procedure, TE identification and classification is intrinsically an error prone process. Consequently, there was a crucial need to provide TE consensuses with evidences able to justify their classification, and this on the up to thousands TE consensus for one genome that such an approach may provide.

Few TE databases already exist focusing generally on different TE biology views (Species, TE families) but biological information on the sequences remains globally poor.

Here, we will present RepetDB developed in the frame of GnpIS, a genetic and genomic Information System. RepetDB was designed to store and retrieve TEs homogeneously detected and classified with their TE evidences by REPET pipelines. RepetDB is an implementation of Intermine, a public data warehouse framework used here to store, search, browse, analyze and compare all the data recorded with each TE consensus allowing simple to very complex queries. Finally, TE data are exposed through a worldwide data discovery system.

POA-330

Transposable elements lineage-specific activity and genome content during the evolution of branchiopod crustaceans.

Andrea Luchetti[1], Barbara Mantovani[1]

[1]University of Bologna (Italy)

Transposable elements (TEs) are ubiquitous component of eukaryotic genomes and may play a major role in genome evolution and gene regulation. Although main TEs lineages are shared by all eukaryotes, they are often differentially represented even in related genomes; this can be due to either genomic mechanisms or to host organism life-history traits. The class Branchiopoda (Crustacea) includes, among others, the renowned model organism *Daphnia* (order Cladocera) and the so-called living fossil *Triops cancriformis* (order Notostraca). Branchiopods exhibit very small genomes (<300 Mb), but they attracted little attention regarding genome sequencing and consistent data on TEs complement and its evolution are lacking. A clear lineage-specific activity pattern emerged from TE analyses performed in four published and four de novo assembled genomes: DNA elements dominates the Notostraca genomic landscape, while LTRs are majorly represented in *Daphnia* and LINEs in the only available Spinicaudata species. The TEs content also varies extensively among genomes (9.5% - 40%) and even if it appears correlated to the genome size, the spinicaudatan taxon shows an unexpectedly high TEs amount. TEs content variation and activity pattern have been linked to host organisms' evolutionary history or life-history traits, such as the mating system. In the present analysis, the comparison between two *Triops cancriformis* genomes, from a bisexual and a parthenogenetic population, shows only limited variation. Overall, the small branchiopod genome and their peculiar life-history traits may represent a nice framework where to address specific studies about the TEs role in genome functions and evolution.

POA-331

Multiplatform assembly of a bird-of-paradise genome reveals rapid turnover of repetitive sequences on W chromosomes and near centromeres of birds

Valentina Peona[1], Mozes Blom[2, 1], Luohao Xu[3], Ignas Bunikis[4], Qi Zhou[3], Knud Jonsson[5], Martin Irestedt[2], Alexander Suh[1]

[1]Uppsala University (Sweden), [2]Swedish Natural History Museum (Sweden), [3]Wien University (Austria), [4]Uppsala University (Sweden), [5]Natural History Museum of Denmark (Denmark)

Birds-of-paradise are iconic for their colorful and unique feathers as well as for elaborate male display dances. They have undergone a rapid evolutionary radiation in the past ~15 my and despite the high sexual selection, they are able to hybridize even between genera. Here we sequenced the deep-branching species Lycocorax pyrrhopterus (paradise crow) with a genome size of 1Gb to elucidate the structure of highly repetitive genomic regions in birds-of-paradise.

We used 1) PacBio long reads as the assembly backbone; 2) Illumina reads to correct PacBio errors; 3) 10XGenomics linked reads to scaffold and orient contigs; and 4) Dovetail Chicago and 5) PhaseGenomics Hi-C physical maps to obtain a chromosome-level assembly of this non-model organism.

In line with the expected karyotype, we obtained 38 chromosome-level scaffolds that include more than 96% of the assembly. PacBio reads resolved significantly more repetitive elements than Illumina alone, revealing an additional 20 TE subfamilies as well as potentially centromeric satellites. We obtained high-quality W chromosome, comparable with the chicken W both in size (~19 Mb vs. ~7 Mb) and repeat content (~70%). Notably, over 50% of the W is comprised of endogenous retroviruses (ERVs). Our assembly highlights that frequent rearrangements, probably caused by the great ERV content, lead to a lack of synteny except for short conserved regions.

Thanks to our multiplatform approach, we were able to investigate highly repetitive and complex regions that revealed to be in rapid evolution in contrast to the stable gene-rich regions of bird genomes.

POA-332

Transposable elements affect the transcriptional regulation of stress response genes in Drosophila and humans

Josefa Gonzalez[1], Vivien Horvath[1], Jose Villanueva[1]

[1]Institute of Evolutionary Biology (CSIC-UPF) (Spain)

Although transposable elements (TEs) are an important source of regulatory variation, their genome-wide contribution to the transcriptional regulation of stress response networks has not been studied in detail, mostly because of technical limitations. TEs are repetitive sequences and as such are difficult to annotate and to study. However, we have developed a computational pipeline that allows us to accurately estimate the frequency of TEs in populations: T-lex2. In this work, we take advantage of the wealth of information available for Drosophila melanogaster and humans to quantify the role of transposable elements in stress regulatory networks. We used in silico predictions and ChIP-seq data to localize transcription factor binding sites (TFBS) for several stress-response transcription factors involved in six different stress regulatory networks. For the transposable elements annotated in the D. melanogaster reference genome, we then used T-lex2 to estimate their population frequencies in 61 worldwide natural populations. Besides population frequencies, we also considered other lines of evidence such as the presence of histone marks in TEs, and the presence of signatures of natural selection, to pinpoint those TEs more likely to be affecting the expression of nearby genes. For a representative subset of the candidate TEs, we performed transgenic reporter assays in different stress conditions. Overall, our results showed that TEs are relevant contributors to the transcriptional regulation of stress response genes.

POA-333

The impact of repetitive DNA on speciation rates in teleost fish

William Reinar[1], Ole Toerresen[1], Michael Matschiner[2, 1], Jostein Starrfelt[1], Alexander Nederbragt[1], Kjetill Jakobsen[1], Sissel Jentoft[1]

[1]University of Oslo (Norway), [2]University of Basel (Switzerland)

Repetitive DNA, such as transposable elements (TEs) and simple repeats, are sometimes referred to as non-functional genomic dark matter. TEs take advantage of the DNA replication and transcription machinery in host cells to facilitate self-propagation within genomes. Simple repeats expand in a different manner, mainly by replication slippage and recombination. It has been proposed that genetic variation caused by high mutation rates linked to repetitive DNA can facilitate adaptation. However, it is known that if left uncontained, active TEs can be hazardous for genome integrity. Further, large length expansions in simple repeats are in many cases linked to disease. Still, TE propagation has been associated with adaptive radiation events and variation in simple repeats have been shown to regulate gene expression, protein function and phenotypic traits. In this study, we take advantage of genome sequencing resources to conduct a full characterization of the repetitive DNA content in genomes of more than a hundred species of teleost fish, elucidating general patterns and lineage-specific shifts. Moreover, we demonstrate correlations between rates of change in repetitive DNA content and clade-specific speciation rates.

POA-334

REPET: a tool for revealing the secrets of transposable elements

Veronique Jamilloux[1], Hadi Quesneville[1]

[1]INRA French National Institute for Agricultural Research (France)

REPET: a tool for revealing the secrets of transposable elements

Transposable elements (TEs) are the most dynamic component of eukaryote genomes. They play important roles in genome structures and gene expressions that both impact genome evolution. Consequently, they are important biological entities, having particular evolutionary dynamics that need to be annotated to fully understand this impact for the host genomes. To rigorously study them, it is necessary to use tools taking into account TEs evolutionary dynamics.

Today, the recent developments of new sequencing technologies improve by far their reconstruction in genome sequences, filling what was assembly gaps with previous sequencing technologies. This recent advance combined with the high pace of the new sequenced species, makes their systematic annotation challenging and require to automatize the TEs discovery process.

We developed the REPET package for these aims for more than 12 years. Our last release 3.0 implements new methods and tools reducing computing resource requirement, that allows to overcome the most difficult cases such as complex and/or large genomes. We are going to present novelties and strategies to help biologists to work out a TEs detection and annotation adapted to their biological questions.

POA-337

Comparative analysis of genomic repeat content in acridid grasshoppers reveals phylogenetic similarities as well as unexpected differences

Abhijeet Shah[1, 2], Holger Schielzeth[1], Joe Hoffman[2]

[1]Friedrich-Schiller-Universitaet Jena (Germany), [2]Universitaet Bielefeld (Germany)

Large parts of eukaryotic genomes consists of repetitive elements and the repeat content is variable among species and correlates tightly with genome variation. Genome size varies tremendously across species with similar level of cellular and developmental complexity and effects fitness-related traits such as gene expression, metabolic rate and cell and body size . This phenomenon has been insufficiently examined in insects.The highly diverse orthopteran group exhibits genome gigantism amongst the insects. However, it was not known what may contribute to this phenomenon. The recently published Locusta migratoria genome revealed that repetitive elements constituted about 60% of the assembled genome, of which DNA transposons and LINE retrotransposons where the most abundant elements. Recent genomic repeat content analysis of the acridid grasshopper, Gomphocerus sibiricus, suggests that satellite DNA dominates (estimated 9-10%) the genome as the largest single repeat class with an estimated genomic repeat content of about 86-89%. This distribution of repeat elements differs substantially from other published distributions. This existence of one predominant class of repeats argues for a recent expansion of this repeat sequence type. However this pattern does not hold for other acrididae grasshoppers. We investigated 6 acrididae grasshopper species for potential gains and losses of repeat elements. Our preliminary results suggests that estimated repeat content strongly correlates with estimated genome size, with estimated genome content ranging from 79-94%. We provide evidence of satellite DNA to the likely driver of genome expansions in the observed genomic gigantism in specific Acridid grasshoppers.

POA-338

Detecting structural variations in human genome using nanopore sequencer

Satomi Mitsuhashi[1], Martin C Frith[2, 3, 4], Naomichi Matsumoto[1]

[1]Yokohama City University (Japan), [2]National Institute of Advanced Industrial Science and Technology (Japan), [3]University of Tokyo (Japan), [4]National Institute of Advanced Industrial Science and Technology (Japan)

In Mendelian disease studies, DNA sequencers have been widely used to detect the pathogenic variations in the human genome. However, current detection rate of the pathogenic variants is approximately 30% in many studies. There are many possible reasons but part of this may be a difficulty in detecting structural variations or in completely covering the highly similar repetitive regions by short-read sequencers (100-300bp). DNA sequencers are the important technologies to accurately detect the pathogenic variations in the diseased individuals and there has been a great improvement in a few years including long-read sequencers (more than thousands to ten thousands bp). Long-read sequencers are proposed to be applicable to Mendelian diseases in detecting pathogenic structural variations or alteration of repeats, however, it is still challenging partly due to a lack of our knowledge of normal structural variations or repetitive regions in human reference genome.

We are aiming to detect possible pathogenic structural variations using nanopore sequencer at a relatively low coverage and cost, which could be missed or overestimated by the current technologies. To detect the true structural variations, we applied the method that compares the ape reference genomes to human genome reference to omit the falsely detected structural variations in the patients' sequence reads. Obtaining sequence reads long enough to encompass the whole structural variation or repeats was crucial to accurately detect the variation. Although there are still a number of limitations, long-read sequencer may provide essential information to understand the association between structural variations and the human diseases.

POA-339

Accumulation of repeated elements during dog domestication: insight from grey wolf and dhole genomes

Guo-Dong Wang[1], Xiu-Juan Shao[2], Bing Bai[3], Jue Ruan[2], Ya-Ping Zhang[1]

[1]Chinese Academy of Sciences (China), [2]Chinese Academy of Agricultural Sciences (China), [3]the first hospital of Yunnan province (China)

Structural variations played a prominent role in phenotypic evolution, disease susceptibility, and environmental adaptation during domestication of animals. Here, we present high-quality draft genome sequences of the grey wolf (Canis lupus) and the dhole (Cuonalpinus), and report a comprehensive analysis of the evolutionary dynamics of structural variation of dogs based on three Canine genomes. Functional annotations of structural variations of dog harboring genes indicate their involvement in energy metabolism, neurological process, and immune system. Interestingly, we identified and verified at population level, an insertion fully covering a copy of AKR1B1 transcript. Transcriptome analysis revealed a high level of expression of the new copy in the small intestine and liver, implying that dogs tend to increase de novo fatty acids synthesis and antioxidant ability compared to grey wolf, in response to a change in dietary composition during the agricultural revolution. Our findings demonstrate that retroposition can offer raw material for new genes for domestication, and affirm the importance of large-scale genomic variants in domestication studies.

POA-340

Horizontal transfer of transposable elements between parasitic nematodes and their hosts

Sonja Maria Dunemann[1], James Wasmuth[1]

[1]University of Calgary (Canada)

Horizontal transfer (HT) of transposable elements (TE) is more frequent than HT of genes, and HT is more likely to occur between organisms that live in close association, such as parasites and their hosts. Lack of rigorous testing however leads to accumulation of false HT claims. In this study, we used transposable elements of parasitic nematodes to quantify the occurrence of horizontal transposon transfer and found a single case of putative HT.

We used RepeatMasker to screen for nematode transposons within all eukaryotic Refseq genomes. We found a single case of putative horizontal transposon transfer in the common shrew, ¥textit{Sorex araneus}. We tested for contamination of the shrew genome on both assembly and sequence read level, searching for traces of contamination in insertion flanking regions and read pair taxonomy. Finally, we reconstructed the phylogeny with MrBayes and RAxML.

We identified a single nematode transposon to be putatively transferred to a different organism. We found the transposon, RTE1_Sar, to be of nematode origin though annotated as ¥textit{Sorex araneus} transposon in Repbase. Construction of a phylogenetic tree suggests close relation of the transposon in the shrew and in parasites of rodents. This suggests that horizontal transfer, although often suggested for species in close association, does not happen excessively between parasitic nematodes and their hosts.

POA-341

Network analysis of bacterial genes to predict horizontal co-transfer and mobile genetic elements

Yu Wan[1, 4], Ryan R. Wick[1], Danielle J. Ingle[2, 3], Michael Inouye[4], Justin Zobel[5], Kathryn E. Holt[1]

[1]Bio21 Molecular Science and Biotechnology Institute, University of Melbourne (Australia), [2]The Australian National University (Australia), [3]The University of Melbourne, The Peter Doherty Institute for Infection and Immunity (Australia), [4]Baker Heart and Diabetes Institute (Australia), [5]University of Melbourne (Australia)

Mobile genetic elements (MGEs) in bacteria are common vectors evolved to capture, retain and horizontally transfer genes between bacterial strains. They are of particular interest in pathogen evolution for their substantial contribution to the emergence and dissemination of clinically important traits including antimicrobial resistance (AMR) and virulence. The detection of genes horizontally co-transferred between species is relatively straightforward, leveraging the presence of near-identical gene sequences within otherwise highly divergent species; however, co-transfer within species poses analytical challenges. The horizontal co-transfer of genes localised to the same MGE leaves a signature of positive gene-gene associations across bacterial populations, which can be leveraged to identify MGEs. To this end, we developed a method for constructing gene co-occurrence networks within a bacterial species using draft genome data, correcting the associations for population structure in order to increase statistical power and reduce false positives. The association network can be further filtered based on evidence for physical genetic linkage in assembly graphs, to enrich for probable MGEs in the resulting subnetworks. Our method is implemented in GeneMates, an R package that takes as input an SNP matrix, gene profiles and optionally, a table of physical gene distances, and then produces an association/co-transfer table that can be viewed in Cytoscape. We demonstrate utilities of GeneMates for the construction of AMR gene networks and identification of associated MGEs in two important bacterial pathogens, *E. coli* and *Salmonella* Typhimurium. We also show the approach can be used to identify MGEs mediating virulence gene transmission in *E. coli*.

POA-342

Selection against LTR retrotransposons is balanced by locally adapted transposable element alleles in Arabidopsis thaliana

Michelle C Stitzer[1, 2], Jeffrey Ross-Ibarra[1, 2, 3]

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

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

POA-343

Transposon activity in the Arabidopsis thaliana 1001 genomes

Luz Mayela Soto Jimenez[1], Magnus Nordborg[1]

[1]Gregor Mendel Institute (Austria)

Transposon activity in the *Arabidopsis thaliana* 1001 genomes

Transposable elements (TEs) conform about 20% of the *Arabidopsis thaliana* reference genome. Variation in genome size and active copies of transpons, as well as their global impact on gene expression and DNA methylation have been reported before. Complementing these studies, we perform an in-depth analysis of TE variation in the world wide population, focusing more on genetically distinct subpopulations. Furthermore, we investigate the extent of the variation and how it correlates to other phenotypes, aiming to understand more comprehensively the dynamics and regulation of TEs in this species and uncover functions not yet described.

We identified TE insertions de novo, using short-read sequencing data, and observed an extreme TE copy number variation (CNV), where evidence of recent activity is apparent and some individuals have twice as many TEs than others. This TECNV, as in other organisms, is reflected in variation in genome size, accounting for 30% of it. The TECNV follows a geographical pattern that appears to be correlated with another phenotype, DNA methylation. Differences in TE activity and superfamily content are more obvious when contrasting different subpopulations, few of those differences fixed in either subpopulation perhaps helping with or leading to local adaptation. Additionally, we are corroborating our findings with de novo assembled genomes, from PacBio data, of individuals with distinct TECN and trying to identify common causative variants using GWAS.

POA-344

Genome size change and transposon dynamics in the allotetraploid *Arabidopsis suecica*

Robin Burns[1], Polina Yu. Novikova[2, 1], Magnus Nordborg[1]

[1]Vienna BioCenter (Austria), [2]University of Ghent (Belgium)

The merging of two diverged genomes is hypothesized to cause a hybrid breakdown in transposable element (TE) regulation. Additionally, the mating system of a species can alter the strength of selection acting on TEs. We are studying *Arabidopsis suecica* a self-fertile allopolyploid, whose parent genomes are the selfing *A.thaliana* and the outcrossing *A.arenosa*.

Using population data of *A.suecica* from its native Fenno-Scandinavian range we observed that the TE copy number in the *A.arenosa* subgenome of *A.suecica* is significantly lower than in *A.arenosa*. Aligning high quality PacBio assemblies we observed that the *A.arenosa* subgenome of *A.suecica* apparently shrank in size. This streamlining of the genome may be a consequence of the shift to selfing in *A.suecica*.

We resynthesized *A.suecica* lines and observed an increase of TEs in the *A.thaliana*subgenome of *A.suecica* after just a few generations. This swift increase suggests the hybrid breakdown is real. Interestingly, we did not see this increase of TE copy number reflected in our natural populations. This suggests the TE mis-regulation may be limited to the early generations following allopolyploidy.

These findings suggest that the two sub-genomes of *A.suecica* have experienced drastic changes in their repeat content and size, and despite co-existing in the same cell, are evolving asymmetrically.

POA-345

Transcriptome analysis to identify genes derived from endogenous retrovirus that mediate cell-cell fusion during myoblast differentiation

Mahoko Takahashi Ueda[1], Satomi Mitsuhashi[2], Hiroaki Mitsuhashi[3], Tadashi Imanishi[4], So Nakagawa[1, 4]

[1]Tokai University (Japan), [2]Yokohama City University Graduate School of Medicine (Japan), [3]School of Engineering, Tokai University (Japan), [4]Tokai University School of Medicine (Japan)

Endogenous retrovirus (ERV) accounts for approximately 10% of mammalian genomes. The ERVs belong to the retrotransposon, the majority of which are considered as junk DNA. However, some ERVs, especially containing viral envelope genes, were reported to be functional in mammalian development. The most famous envelope gene was found in placenta; *Syncytins* are associated with trophoblast cell differentiation by mediating cell-cell fusion to generate multi-nucleated syncytiotrophoblasts of placental villi. Like syncytiotrophoblasts, skeletal muscle cells are also multinucleated, which are formed by the fusion of undifferentiated myoblasts into myotubes. Although *syncytin* was reported to have fusogenic effect in myogenesis of male mice, genes that directly mediate myoblast fusion are unclear. Our sequence analyses of mammalian genomes revealed that more than thousands of ERVs still have open reading frames (ORFs) of envelope gene. This led us to hypothesize that new envelope genes may be involved in cell-cell fusion during myogenesis (myoERV). To identify myoERV genes, we performed RNA-seq analyses of mouse myoblasts at three different time points of differentiation. Using in-house developed annotations of ERV sequences, we identified envelope-like sequences showing different expression patterns during differentiation. ORFs of these myoERV candidates were cloned and overexpressed in mouse muscle cell line. Two of them were confirmed to localize to plasma membrane, and were further analyzed to identify their functions. In this meeting, I will present and discuss our latest results including the development of new method for RNA-seq analysis specialized in targeting ERV sequences as well as the evolution of myoERV in mammals.

POA-347

Evaluating genome and transcriptome variation across the Antarctic Notothenioid fish radiation to explore causes and consequences of adaptive speciation.

Iliana Bista[1, 2], Shane McCarthy[2], Eric Miska[3], Thomas Desvignes[4], Melody Susan Clark[5], John Postlethwait[4], C.-H. Christina Cheng[6], Walter Salzburger[7], H. William Detrich III[8], Karen Oliver[1], Jason Skelton [1], Michelle Smith[1], Petr Danecek [1], Richard Durbin[1, 2]

[1]Wellcome Trust Sanger Institute (United Kingdom), [2]University of Cambridge (United Kingdom), [3]University of Cambridge (United Kingdom), [4]University of Oregon (United States), [5]Natural Environment Research Council (United Kingdom), [6]University of Illinois at Urbana-Champaign (United States), [7]University of Basel (Switzerland), [8]Northeastern University (United States)

The Notothenioid radiation of Antarctic fish presents one of the most dramatic examples of marine adaptive radiations, with over 120 species, dominating the Southern Ocean in fish species richness and biomass. Through the Vertebrate Genomes Project, we are sequencing >20 species (5 families), producing reference assemblies through long read sequencing (PacBio, 10X Genomics). Mapping against long read assemblies will allow study of highly repetitive gene families (e.g. antifreeze glycoproteins, AFGPs) and detection of structural variation. We will measure transposon copy number variation for each repeat family, to investigate the role of increased activity of transposable elements (TEs) in the observed expansion of genome size across the Notothenioids, increasing from most basal to recently divergent species. Furthermore, we are sequencing the transcriptomes of 16 notothenioid species (4 families) from piRNA and total RNA. piRNAs are a type of endogenous small RNAs, involved in protecting the genome from TE activity, and are mostly found in the germline. We will compare the expression levels of piRNAs throughout the different species, and investigate piRNA activity relevant to specific transposons in relation to the potential genome expansion. Overall the study will provide a deep genomic characterization of this important fish group, and is one of the most extensive studies using whole genome data to decipher the mechanisms of fish genome evolution.

POA-349

Pilot studies of transposable elements in Bronze Age skeletal human DNA

Oliver Piskurek[1]

[1]University of Goettingen (Germany)

In the last two decades more than 60 Bronze Age individuals were investigated from the Lichtenstein cave in Lower Saxony, central Germany. Skeletal human remains in this cave were evolutionary preserved in a constant temperature of 8 degree C and coated with a gypsum layer for 3,000-years. STR analyses were shown to be successful and just recently presence/absence patterns of TEs were investigated in these prehistoric samples as well. To test the excellent conditions of the aDNA, we showed in a pilot study that it is possible to amplify Alu loci, including flanking regions with fragment lengths up to 500bp. The presence/absence situation of 30 Alu loci was illustrated for three members of a prehistoric family. Additionally, we reduced the amount of needed aDNA and applied a new developed assay on several samples from the Lichtenstein cave with varying states of DNA preservation. We developed two duplex Alu PCR assays, which consist of two flanking Alu primer sets and one internal Alu primer. In general, amplification success decreased with increasing degree of DNA degradation. Since TEs can cause problems due to their similar sequences these new assays will help for controlling NGS assembling of Alu elements. First whole genome library preparation and shotgun sequencing efforts of Lichtenstein cave samples showed that the endogenous DNA content can be higher than 40%. With this amount of human DNA it will be possible to investigate Alu insertions in the 3,000-year-old Bronze Age samples on a genomic scale.

POA-351

The PIWI/piRNA response is relaxed in a rodent that lacks mobilizing transposable elements

Michael W Vandewege[1, 2], Roy N. Platt[2], Aliceanne Szeliga[3], Dana Merriman[4], David A Ray[2], Federico G. Hoffmann[5, 6]

[1]Temple University (United States), [2]Texas Tech University (United States), [3]Harvey Mudd College (United States), [4]University of Wisconsin (United States), [5]Mississippi State University (United States), [6]Mississippi State University (United States)

Transposable elements (TEs) are genomic parasites that can propagate through host genomes by inserting additional copies of themselves. TEs are major players in the evolution of genomes and TE mobilization is generally deleterious. Animals have evolved defense mechanisms against TEs, where Piwi-interacting RNAs (piRNAs) and PIWI/Argonaute proteins have emerged as key players in the repression of TE mobilization. Because the set of active TEs changes over time, the relationship between TEs and piRNAs resembles the evolutionary arms race between pathogens and the immune system. The PIWI/piRNA system has been investigated largely in animals with actively mobilizing TEs and it is unclear how the system functions in the absence of mobilizing TEs. The ground squirrel provides an excellent opportunity to examine PIWI/piRNA and TE dynamics within the context of undetectable current TE accumulation. To do so, we sequenced RNA and small RNAs pools from ground squirrels and compared patterns with rabbit and mouse. Contrary to our expectations, we found that TEs are actively expressed in the ground squirrel. Our analyses suggest that the PIWI/piRNA system reduces TE expression in rabbit and mouse, but not in squirrels. We also discovered that PIWIL4 is expressed all the way into adulthood in the squirrel, an important difference with rabbit and mouse, as PIWIL4 plays an upstream role in TE methylation. These observations suggest that in the ground squirrel the piRNA response to TE expression has relaxed and that DNA methylation might have played a significant role in suppressing TE activity.

POA-352

Analysis of the red seaweed Gracilariopsis chorda genome provides insights into genome size evolution in Rhodophyta

JunMo Lee[1], Debashish Bhattacharya[2], Hwan Su Yoon[1]

[1]Sungkyunkwan University (Republic of Korea), [2]Rutgers University (United States)

Red algae (Rhodophyta) underwent two phases of massive genome reduction during their early evolution. The derived red seaweeds did not attain genome sizes or gene inventories typical of other multicellular eukaryotes. We generated a high quality 92.1 Mbp draft genome assembly from the red seaweed *Gracilariopsis chorda*, as well as methylation and small (s)RNA data. We analyzed these and other Archaeplastida genomic data to address three questions: 1) what is the role of repeats and transposable elements (TEs) in explaining Rhodophyta genome size variation, 2) what is the history of genome duplication and gene family expansion/reduction in these taxa, and 3) is there evidence for TE suppression in red algae? We find that the number of predicted genes in red algae is relatively small, in particular when compared to plants, with no evidence of polyploidization. Genome size variation is primarily explained by repeat and TE expansion with the red seaweeds having the largest genomes. LTR and DNA repeats are the major contributors to genome growth. About 8.3% of the *G. chorda*genome undergoes cytosine methylation among gene bodies, promoters, and TEs, and 71.5% of TEs contain methylated DNA with 57% of these regions associated with sRNAs. These latter results suggest a role for TE-associated sRNAs in RNA-dependent DNA methylation to facilitate silencing. Concomitant with TE spread was the rise of epigenetic suppression that we postulate, in combination with other factors such as changes in population size help explain genome size evolution in red algae.

POA-353

Transposable elements and resistome analysis of Staphylococcus lugdunensis isolates from diverse hospital and community sources in Hong Kong

Melissa Chunjiao LIU[1], Huiluo CAO[1]

[1]The University of Hong Kong (China)

Background: Multidrug resistance of Staphylococcus lugdunensis have been reported among certain clones. Here we analyze its gene pool to provide a deeper understanding of its transposable elements and resistome.

Materials/methods: A number of 93 S. lugdunensis isolates collected from diverse hospital and community sources during 1998 to 2016 were sequenced using Illumina platforms. The sequencing data was trimmed using Trimmomatic, de novo assembled by the CLC Workbench and annotated in the RAST server.

Results: A total of 211 acquired resistance genes were identified in 72 isolates. Among them, 13 unique genes were identified as conferring resistances to aminoglycosides, lincosamides, tetracyclines, chloramphenicol, mupirocin, macrolides, beta-lactams and some disinfectants. Of note, all the genes were associated with specific transposable elements spreading among a wide range of bacterial hosts. Specifically, aacA-aphD was disseminated by a Tn4001-IS257 hybrid between Staphylococcal plasmids. An aadD-rep22-lnuA gene cluster and the co-carry of tetK, cat and rep7 were mobilized by a nearby IS257. The IS257 also carried ileS2, qacA and qacC among the other plasmid contigs. The ermC gene was found in a rep10 plasmid. Furthermore, co-existing of aadE and aphA-3 was characteristically spotted on a phage-like fragment. The chromosomal blaZ was uniquely identified in the Tn552 and the mecA was harbored by a type IV or V SCCmec element.

Conclusions: Our analysis reveals that the resistance genes in S. lugdunensis isolates are strongly associated with specific transposable elements for gene transfer.

POA-354

Reconstructing the evolutionary history of endogenous retroelements

Laura F. Campitelli[1], Mathieu Blanchette[2], Timothy R. Hughes[1, 3]

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

The human genome contains over 750 C2H2 zinc finger proteins (C2H2s), comprising the largest class of transcription factors (TFs). The sequence preferences of C2H2s evolve rapidly and vary dramatically, due to their modular and flexible DNA recognition mechanism. About half of these proteins (345) contain a transcription-silencing KRAB domain (KZFPs). The expansive KZFP collections in mammalian genomes are disproportionately lineage-specific, demonstrating positive selection in recent history. Additionally, the majority of KZFPs bind at least one class of transposon, the majority of which are endogenous retroelements (EREs). Investigation of the KZFP-ERE interaction is limited by a need for full-length consensus models for many human EREs, including virtually all LINE elements. However, establishment of full-length models from the modern human genome is impeded by tens of millions of years of neutral decay since the EREs first amplified. I have established a unified approach to delineate the stepwise evolution of sequences across all ERE subfamilies throughout the last 100M years of human evolution by exploiting a maximum-likelihood imputation of whole ancestral genomes (Ancestors 1.1) to create a timeline of ERE amplifications and reverse ERE decay to build full-length consensus models from EREs in pre-human genomes. My goal is to produce full-length sequence models for all human EREs. The accuracy and utility of this approach is demonstrated by thorough evaluation of a well-studied internal control: the L1 superfamily. Future work will exploit this model to dissect the evolutionary interaction driving and maintaining KZFP-ERE coevolution.

POA-357

The expansion of a ligand transporter related gene family in cluster specific to teleost fishes

Langyu Gu[1, 2], Canwei Xia[3]

[1]Southwest University (China), [2]University of Basel (Switzerland), [3]Beijing Normal University (China)

Gene and genome duplication play an important role in the evolution of new gene functions. Especially for gene clusters regarding to their associations with innovation and adaptation. Here, we are the first time that report the expansion of a ligand transporter related gene family in cluster specific to teleost fishes. The only one gene in tetrapod was expanded in two clusters via genome duplication and tandem gene duplication, showing a dynamic evolutionary pattern in different fishes. Based on comparative genomic and transcriptomic analyses, protein 3D structure simulation, evolutionary rates detection and genome structure detection, subfunctionalization and neofunctionalization after duplication were observed at both protein and expression level, especially for lineage specific duplicated genes that were under positive selection. Different duplicated genes expressed in tissues that are related to sexual selection and adaptation. Especially for a cichlid fish, Astatotilapia burtoni, whose paralogs in different clusters showed highly expression in egg-spots (a sexual selection related trait) and lower pharyngeal jaw (related to feeding strategy), respectively. These two novelties are famous for adaptive radiation of cichlid fishes. Interestingly, the gene clusters are located near/at the breaking point of genome rearrangement. Since genome rearrangement can capture adapted genes or antagonist sex determining genes to protect them from introgression by reducing recombination, it can promote divergence and reproductive isolation. This further suggest the importance of the expansion of this ligand transporter related gene family for speciation and adaptation in teleost fishes, especially for cichlid fishes.

POA-359

Revisiting Ohno's hypothesis of dosage compensation by using the neo-sex chromosomes in *Drosophila*

Masafumi Nozawa[1, 2]

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

In his seminal book entitled "*Sex chromosomes and sex-linked genes*" published in 1967, Susumu Ohno proposed the concept of dosage compensation (DC) in which a single male X chromosome is up-regulated to equalize the gene expression level to that on the two Xs in females and that on autosomes. Since then, many biologists have evaluated this hypothesis and confirmed that many organisms indeed have a global DC mechanism operating along the entire male X. At the initial stage of sex chromosome evolution, however, gene-by-gene DC on individual X-linked genes may also be necessary because Y-linked genes are expected to become individually pseudogenized at different times. I therefore tested whether the up-regulation of X-linked genes depends on the status of their Y-linked homologs, using the young sex chromosomes, neo-X and neo-Y, in *Drosophila*. In support of the presence of gene-by-gene DC, the extent of up-regulation in males was higher for neo-X-linked genes with pseudogenized neo-Y-linked homologs than for neo-X-linked genes with functional neo-Y-linked homologs. Further analyses indicated that neo-X-linked genes first acquired the potential of up-regulation, which enabled the pseudogenization of neo-Y-linked homologs, without serious deleterious effects on male fitness. Surprisingly, Ohno predicted the presence of gene-by-gene DC in his book more than 50 years ago. He apparently connected degeneration of Y chromosomes and genome duplication in terms of gene dosage, although these events result in opposite outcomes. In this presentation, I therefore would like to emphasize how dosage sensitivity affects the evolutionary patterns of genomes as well as organisms.

POA-360

Faster evolving primate genes are more likely to duplicate.

Aine Niamh O'Toole[1, 3], Laurence D Hurst[2], Aoife McLysaght[1]

[1]Trinity College Dublin (Ireland), [2]University of Bath (United Kingdom), [3]University of Edinburgh (United Kingdom)

An attractive and long-standing hypothesis regarding the evolution of genes after duplication posits that the duplication event creates new evolutionary possibilities by releasing a copy of the gene from constraint. Apparent support was found in numerous analyses, particularly, the observation of higher rates of evolution in duplicated as compared with singleton genes. Could it, instead, be that more duplicable genes (owing to mutation, fixation, or retention biases) are intrinsically faster evolving? To uncouple the measurement of rates of evolution from the determination of duplicate or singleton status, we measure the rates of evolution in singleton genes in outgroup primate lineages but classify these genes as to whether they have duplicated or not in a crown-group of great apes. We find that rates of evolution are higher in duplicable genes prior to the duplication event. In part this is owing to a negative correlation between coding sequence length and rate of evolution, coupled with a bias toward smaller genes being more duplicable. The effect is masked by difference in expression rate between duplicable genes and singletons. Additionally, in contradiction to the classical assumption, we find no convincing evidence for an increase in d_N/d_S after duplication, nor for rate asymmetry between duplicates. We conclude that high rates of evolution of duplicated genes are not solely a consequence of the duplication event, but are rather a predictor of duplicability. These results are consistent with a model in which successful gene duplication events in mammals are skewed toward events of minimal phenotypic impact.

POA-361

Prevalence, Mechanisms and Importance of Duplicate Gene Divergence in Exon-Intron Structure

Xuehao Fu[1, 2], Hongyan Shan[1], Peipei Wang[1], Hongzhi Kong[1, 2], Guixia Xu[1]

[1]Institute of Botany, Chinese Academy of Sciences (China), [2]University of Chinese Academy of Sciences (China)

Gene duplication plays key roles in organism and genome evolution. Understanding how duplicate genes evolve and diverge through time is critical for elucidating the mechanisms underlying the origins of new characters and new organisms. Previous studies have shown that, at least in plants, some duplicate genes have diverged in the exon-intron organization, suggestive of structural divergence. However, because the species and gene pairs sampled were limited, it is yet unclear whether this phenomenon is widespread. In this study, by conducting a genome-wide study on closely related duplicate genes from four representative species of plants (Arabidopsis thaliana), animals (Drosophila melanogaster), fungi (Saccharomyces cerevisiae), and protists (Paramecium tetraurelia), we found that structural divergence occurred prevalently in every examined species. Exon/intron gain/loss, exonization/pseudoexonization, and intra-exonic insertion/deletion, are detected to be responsible for structural divergence, the occurrence of which increases with evolutionary time. The fact that the Pearson correlation coefficient and Euclidean distance of expression patterns between a pair of genes is respectively lower and higher in diverged than in undiverged duplicates suggests that structural divergence may be coupled with expression divergence. Using dN/dS value as a measurement of functional constraint, we found that duplicate genes with structural changes have higher dN/dS values, indicative of weaker functional constraint on these genes. This is in concordance with their lower expression levels. Our findings show that the modes of structural divergence of duplicate genes are generally consistent in different eukaryotic species, implying that structural divergence is an important contributor to the evolution of duplicate genes.

POA-362

Testing the Ortholog Conjecture for whole genome duplications

Tina Begum[1, 2], Marc Robinson-Rechavi[1, 2]

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

There is a long standing interest in quantifying the impact of whole genome duplication (WGD). Much attention has been focused on the relative role of sub- and neo-functionalization of ohnologs, paralogs from WGD. Yet there is a recent debate on the question whether paralogs do diverge in function more than orthologs, the "Ortholog Conjecture". It is important to answer this question in the case of ohnologs, if we are to understand the impact of WGD. Here we analyse expression tissue-specificity, as in Kryuchkova-Mostacci 2016, with an improved phylogenetic analysis which builds on Dunn et al 2018, for expression data from vertebrates with an emphasis on fishes. We show, contra Dunn et al 2018, that the ortholog conjecture holds for fish WGD. We combine this with previous results on biased gene retention after WGD to show how fish WGD have contributed to shaping their genome function.

Kryuchkova-Mostacci & Robinson-Rechavi 2016 Tissue-Specificity of Gene Expression Diverges Slowly between Orthologs, and Rapidly between Paralogs. PLOS Comput Biol 12:e1005274

Dunn et al. 2018 Pairwise comparisons across species are problematic when analyzing functional genomic data PNAS 115:E409-E417

POA-363

Mechanisms of loss and preservation of whole-genome duplicates revealed by population genomics

Parul Johri[1], Jean-Francois Gout[2], Michael Lynch[2]

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

Gene duplications serve as a primary source of novelty in evolution. Whole-genome duplications (WGDs) that have occurred in the ancestor of many eukaryotes, have also pre-dated the species radiation of a group of 15 species belonging to the genus *Paramecium*, a unicellular lineage. It is not currently understood which genes are preferentially retained versus lost post-WGD, neither are the molecular mechanisms of these processes well-defined. In order to address this gap, we closely examined ongoing loss of whole-genome duplicates in *Paramecium* populations, by sequencing 10-13 individuals (sampled worldwide) from each of 5 *Paramecium* species. We show that orthologous copies of gene duplicates that have been lost in one taxa are much more likely to have segregating nonfunctional alleles in sister taxa. We also show that gene duplicates that have segregating non-functional polymorphisms accumulate more deleterious variants and thus appear to be headed for future loss. We also observe different mutations causing inactivation of the same copy in individuals from different populations. Our results suggest that one of the two copies created by a WGD gets marked for degradation early, resulting in loss of the same copy across multiple taxa. These results also imply that the process of gene loss following WGD in some cases occur slowly; such copies under relaxed selection may thus provide the raw material for sequence changes that may result in acquisition of a novel function.

POA-364

Tissue specific ploidy variation in sexual and apomictic seeds

Dorota Paczesniak[1], Marco Pellino[1], Devan Guenter[1, 4], Siegfried Jahnke[2], Andreas Fischbach[2], John T. Lovell[3], Timothy F. Sharbel[1]

[1]University of Saskatchewan (Canada), [2]Forschunscentrum Juelich (Germany), [3]HudsonAlpha Genomic Sequencing Center (United States), [4]Government of Canada (Canada)

The effects of polyploidy are often examined at the level of a whole organism or across species, however ploidy variation occurs also at the intra-individual level and can be limited to a specific tissue, e.g. endosperm tissue in the seeds of flowering plants.
Endosperm is the tissue providing nutrition to the developing embryo, and in sexually-reproducing plants it is typically triploid. In the genus *Boechera*, a wild relative of *Arabidopsis*, both sexual and apomictic (i.e. reproducing asexually via seed) lineages are found. Diploid apomictic lineages produce a meiotically-unreduced diploid egg cell (apomeiosis) which develops parthenogenetically (without fertilization) into an embryo that is genetically identical to the mother plant, whereas endosperm is pseudogamous (i.e. requires fertilization) and typically hexaploid. Interestingly, both sexual (triploid) and apomictic (hexaploid) endosperm have a 2:1 maternal to paternal genome ratio. Deviations from this pattern are also commonplace: variation in endosperm ploidy and the ratio of maternal to paternal genomes exists aamong and within apomictic lineages.
We are interested in understanding effects of ploidy variation and maternal/paternal genome dosage on endosperm function and plant life history traits, in the context of sexual and apomictic reproduction. We have examined the effects of genotype, ploidy and parental genome dosage as potential factors underlying phenotypic variation in seed size, an agriculturally important trait.

POA-365

Ancient polyploidy in orchids

Zhen Li[1, 2], Rolf Lohaus[1, 2], Guo-Qiang Zhang[3], Zhong-Jian Liu[3], Yves Van de Peer[1, 2, 4]

[1]Ghent University (Belgium), [2]VIB (Belgium), [3]The National Orchid Conservation Center of China and The Orchid Conservation and Research Center of Shenzhen (China), [4]Genomics Research Institute (South Africa)

With more than 25,000 species, Orchidaceae represents a staggering 8-10% of the world's plant species and is therewith one of the largest angiosperm families. Orchids are renowned for their specialized flowers, show a huge diversity of epiphytic and terrestrial growth forms, and have successfully colonized almost every habitat on Earth. Ever since Darwin published his book entitled *Fertilization of Orchids* in 1862, orchids have attracted great interest from botanists and evolutionary biologists. The genomes of the previously-sequenced orchids *Dendrobium catenatum* and *Phalaenopsis equestris*contained signatures of ancient whole-genome duplication (WGD) events. Here we analyzed the newly-sequenced genome of the early diverging orchid *Apostasia shenzhenica* to investigate whether these WGDs were independent events, or whether all orchids share an ancient polyploidy. Also the genome of *A. shenzhenica* shows clear evidence of a WGD event, and seems to have occurred prior to the divergence of all orchids. Although phylogenomic analyses of duplicated genes on collinear blocks mostly support a shared WGD in Orchidaceae, a large number of such duplicates from *A. shenzhenica* coalesce on the Apostasioideae stem branch. The incongruent phylogenetic signals found across orchid gene trees result from the probably very short time interval between the shared WGD event and the divergence of Orchidaceae. This could also suggest that the diploidization of the ancient polyploid occurred after the divergence of extant orchids, potentially facilitating orchid diversification.

POA-367

Specialization is the main route of regulatory evolution following whole genome duplication in vertebrates

Ferdinand Marletaz[1], Panos Firbas[2], Ignacio Maeso[2], Juan Tena[2], Jon Permanyer[4], Hector Escriva[3], Jose Luis Gomez-Skarmeta[2], Manuel Irimia[4]

[1]University of Oxford (United Kingdom), [2]CSIC Universidad Pablo de Olavide (Spain), [3]UPMC Univ Paris (France), [4]Centre for Genomic Regulation (CRG) (Spain)

All vertebrates share multiple morphological and genomic novelties. The most prominent genomic difference from non-vertebrate chordates is the reshaping of the gene complement that followed two rounds of whole genome duplication (WGD or 2R), which likely occurred at the base of the vertebrate lineage. These large-scale mutational events are hypothesized to have facilitated the evolution of vertebrate morphological innovations, at least in part through the preferential retention of developmental gene families and transcription factors after duplication. However, duplicate genes and their associated regulatory elements were initially identical, and could not drive innovation without regulatory and/or protein-coding changes. To unravel the major routes of regulatory evolution following WGD at the origin of vertebrates, we compare developmental and tissue-specific transcriptomes of amphioxus, zebrafish, frog and mouse. We found that over 80% of gene families with multiple paralogs in vertebrates have members that restricted their ancestral expression. However, we found that specialization of specific paralogs rather than subfunctionalization (as defined by the DDC model) is the major fate for multi-gene families. Counter-intuitively, vertebrate genes that underwent expression restriction increased the complexity of their regulatory landscapes. Taken together, our results indicate that the two rounds of WGD not only caused an expansion and diversification of gene repertoires in vertebrates, but also allowed functional and expression specialization of the extra copies by increasing the complexity of their gene regulatory landscapes.

POA-368

Subgenome-Enriched Transposable Elements Reveal Allopolyploid Origins Following Whole-Genome Duplications

Adam M Session[1], Daniel S Rokhsar[1, 2]

[1]Joint Genome Institute (United States), [2]UC Berkeley (United States)

Polyploid organisms can arise due to genome doubling without cell division (autopolyploidy) or doubling after interspecific hybridization (allopolyploidy), resulting in more than two chromosome sets per somatic nucleus. Similarity between such homeologous chromosomes presents a barrier to accurate polyploid genome assembly. In recent years, better assembly methods have overcome this hurdle, enabling quantitative exploration of hypotheses concerning the mechanism of individual polyploid formation (allo vs auto), as well as the initial response of genomes to polyploidy.

We hypothesized that one positive marker of allopolyploidy would be transposable elements that differentially expanded in the two diploid progenitors prior to hybridization. To find such markers, we developed a method of using k-mers (specifically 15-mers identified by Jellyfish) to identify commonly occurring words that distinguish homeologous chromosomes. In the polyploid genomes where we found this positive signal, the words that split pairs of homeologous chromosomes overlap, allowing the identification of coherent subgenomes that are each descended from each diploid progenitor. When mapped to the genome, these differentiating 15mers align to the long-terminal repeats (LTRs) of retrotransposons in plants, and often align to DNA transposons in animals. By studying the sequences divergence of these LTRs we can estimate the timing of the hybridization of polyploidy. In contrast, the divergence of progenitor species typically predates polyploidization.

We use this method to show that multiple grass and vertebrate genomes are allopolyploid. We also investigated the evolution of protein-coding gene expression following polyploidy, and its relationship with expanding transposable elements.

POA-369

Beyond Ohno's gene duplication: evolution of enzyme substrate specificity driven by gene loss and horizontal gene transfer

Francisco Barona-Gomez[1]

[1]Cinvestav-IPN (Mexico)

The connection between gene loss (GL) or horizontal gene transfer (HGT) and the functional adaptation of retained proteins is still poorly understood. We apply phylogenomics, comparative genomics and metabolic modeling to detect bacterial species that are evolving by GL or HGT, with the finding that Actinobacteria are undergoing rapid genome dynamics, including loss and gain of L-histidine and L-tryptophan biosynthesis. We observe that the dual-substrate phosphoribosyl isomerase A (PriA), at which these pathways converge, appears to coevolve with the occurrence of trp and his genes (1,2). Characterization of two dozen PriA homologs shows that these enzymes adapt from bifunctionality in the largest genomes, to monofunctional yet not necessarily specialized forms, in genomes undergoing reduction and expansions (3,4). These functional changes are accomplished via mutations, which result from relaxation of purifying selection (GL) or positive selection (HGT), in residues structurally mapped after sequence and X-ray structural analyses. Our results show how GL and HGT can drive enzyme evolution independent of Ohno's gene duplication.

1. Barona-Gomez (2003) Occurrence of a putative ancient-like isomerase involved in histidine and tryptophan biosynthesis. EMBO rep, 4: 296-300.
2. Verduzco-Castro (2016) Co-occurrence of analogous enzymes determines evolution of a novel isomerase sub-family after non-conserved mutations in flexible loop. Biochem J. 473(9): 1141-52.
3. Noda-Garcia (2013) Evolution of substrate specificity in a recipient's enzyme following horizontal gene transfer. Mol Biol Evol. 30(9): 2024-34.
4. Juarez-Vazquez (2017) Evolution of substrate specificity in a retained enzyme driven by gene loss. eLife, 10.7554/eLife.22679

POA-370

Reshaped Patterns of Alternative Splicing after Allopolyploidy in Brassica napus

Keith Adams[1], David Tack[1]

[1]University of British Columbia (Canada)

In allopolyploids, when two genomes from the parental species come together in a common nucleus, there are often novel expression profiles with respect to parental patterns. Alternative splicing is a fundamental aspect of gene expression that generates more than one type of final transcript from a single type of mRNA by differential splicing. We investigated changes in transcript levels and alternative splicing in three resynthesized and one natural allopolyploid Brassica napus line compared to their parental species, B. oleracea and B. rapa, using transcriptome sequencing (RNA-seq). The majority of the expression level changes are repeated among the resynthesized allopolyploids, with a few changes being further paralleled in natural B. napus. Expression changes specific to an allopolyploid line highlight the variability in outcomes of allopolyploidization. We investigated changes to the frequency of alternative splicing among the allopolyploids, revealing myriad qualitative and quantitative differences in alternative splicing events among the genotypes. Decreases in IR event frequency are the most common type of change in the resynthesized polyploids. In each of the three resynthesized allopolyploids, the CT subgenome (derived from the B. oleracea parent) IR event frequencies show significantly more change than the AT subgenome (derived from the B. rapa parent) IR event frequencies. We also examined alternative donor and acceptor events. Our results show that changes in alternative splicing patterns are a common occurrence after allopolyploidy and they contribute to the transcriptome shock experienced by newly formed allopolyploids.

POA-371

Are variable constraints on gene function the only cause of asymmetric fates of paralogs?

Yuichiro Hara[1], Miki Takeuchi[2], Kaori Tatsumi[1], Yuka Kageyama[3], Masahiko HIbi[2, 4], Hiroshi Kiyonari[1], Shigehiro Kuraku[1]

[1]RIKEN (Japan), [2]Nagoya University (Japan), [3]Kwansei Gakuin University (Japan), [4]Nagoya University (Japan)

Phylogenetic reconstruction of a wealth of gene families has shed light on asymmetric molecular evolutionary rates between paralogs occurring through whole genome duplications. Interestingly, fast-evolving duplicates sometimes underwent secondary loss in multiple taxa including traditional experimental animals, which had not been recognized until recently. So far, such asymmetric paralog fates have been attributed to variable functional constraints on those genes, but our recent study focusing on gene loss has provided clues for a different cause. In this study, we reconstructed a phylome, a comprehensive set of gene phylogenies, of broad vertebrate taxa and identified the paralog pairs separated in early vertebrates with asymmetric gene retention: a gene, namely elusive gene, whose orthologs were absent from multiple taxa and its counterpart retained by almost all the vertebrates used. The comparisons among paralogs revealed high nucleotide substitution rates and uneven intra-genomic distribution of the elusive genes, as well as increased gene density, repeat element density, and GC-content within the genomic regions flanking them. This trend of the elusive genes was demonstrated at two different taxonomic levels, that is, Amniota, featuring the Madagascar ground gecko (*Paroedura picta*) whose genome was sequenced by ourselves, as well as Eutheria. Our findings demonstrate that the fates of duplicated genes can be affected by intrinsic properties of their genomic regions, which can persist for hundreds of millions of years after the whole genome duplications, in addition to constraints on gene functions.

POA-373

Using genetic duplication and machine learning to explore important features of genes associated with both heritable, and somatic disease.

Alexandra Claire Martin-Geary[1], Mark Reardon[1], David Wells Newman[1], David L Robertson[1, 2]

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

Using genetic duplication and machine learning to explore important features of genes associated with both heritable and somatic disease.

Understanding factors and features present in a gene's evolutionary history, that lead to heightened disease association is imperative in order to predict both currently unknown, and de-novo mutations as they arise, and rapidly provide targeted medical intervention. Currently our knowledge of the underlying factors leading to a heightened propensity for a gene to be disease associated is limited, however, with the recent development of novel analytical methods and machine learning algorithms we have been able to extend the breadth of comprehension of this important area, and have identified that a mechanistic evolutionary understanding of diverse genetic properties is imperative to the targeting and future prediction of human genetic disease.

Here we present the findings of our analysis of large-scale data collated from a number of open-source datasets. Using both statistical and machine learning algorithms we explored associations between key factors in a gene's history and disease. Categorizing data into a number of sets of significance to both evolution and disease, we explored those factors that have significant associations within and between each group, and subsequently the mechanistic and evolutionary underpinnings of different disease associations therein.

POA-374

Reciprocally retained genes in the angiosperm lineage show the hallmarks of dosage balance sensitivity

Setareh Tasdighian[1, 2, 3], Michiel Van Bel[1, 2, 3], Zhen Li[1, 2, 3], Yves Van de Peer[1, 2, 3], Lorenzo Carretero-Paulet[1, 2, 3], Steven Maere[1, 2, 3]

[1]Ghent University (Belgium), [2]VIB (Belgium), [3]Ghent University (Belgium)

In several organisms, particular functional categories of genes, such as regulatory and complex-forming genes, are preferentially retained after whole-genome multiplications but rarely duplicate through small-scale duplication, a pattern referred to as reciprocal retention. This peculiar duplication behavior is hypothesized to stem from constraints on the dosage balance between the genes concerned and their interaction context. However, the evidence for a relationship between reciprocal retention and dosage balance sensitivity remains fragmentary. We identified which gene families are most strongly reciprocally retained in the angiosperm lineage and studied their functional and evolutionary characteristics. Reciprocally retained gene families exhibit stronger sequence divergence constraints and lower rates of functional and expression divergence than other gene families, suggesting that dosage balance sensitivity is a general characteristic of reciprocally retained genes. Gene families functioning in regulatory and signaling processes are much more strongly represented at the top of the reciprocal retention ranking than those functioning in multiprotein complexes, suggesting that regulatory imbalances may lead to stronger fitness effects than classical stoichiometric protein complex imbalances. Finally, reciprocally retained duplicates are often subject to dosage balance constraints for prolonged evolutionary times, which may have repercussions for the ease with which genome multiplications can engender evolutionary innovation.

POA-375

Dosage-sensitive ohnologs have restricted evolutionary trajectories

Alan M Rice[1], Pauric Donnelly[1], Aoife McLysaght[1]

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

Dosage-sensitive genes are often seen to be refractory to variation. Verebrate ohnologs, paralogs produced from whole genome duplication events at the base of the vertebrate lineage, have been shown to be refractory to small-scale duplication, depleted on human benign copy number variants (CNVs) but enriched on pathogenic variants. This intolerance to copy number change is likely due to a CNV giving rise to a violation of an expression constraint that exists in one or more tissues. While CNVs will impact expression across all tissues, expression quantitative trait loci (eQTLs), genomic regions harbouring sequence variants that influence the expression level of one or more genes, can act in a tissue-specific manner. Expression changes in unconstrained tissues due to the presence of an eQTL will be neutral or potentially beneficial and allow dosage-sensitive genes to vary expression while obeying constraints in unaffected tissues. We find that ohnologs are enriched for being affected by eQTLs and that these eQTLs are biased towards having narrow tissue specificity with ohnologs having fewer eQTL-affected tissues than nonohnologs. Additionally, we find that ohnologs are depleted for being affected by broad tissue breadth eQTLs, likely due to the increased chance of these eQTLs conflicting with expression constraints in one or more tissues and being removed by purifying selection. These patterns suggest that dosage-sensitivity shapes the evolution of ohnologs by precluding copy number evolution and restricting their evolutionary trajectories to changes in expression regulation compatible with their constraints.

POA-376

Preferential retention of homeologs from a single parental subgenome after polyploidy is shaped by functional interactions and dosage-based intrinsic selective constraint

Yue Hao[1], Marianne Emery[2], M. Madeline Willis[3], Jacob D. Washburn[4], Jacob Rosenthal[5], Brandon Nielsen[6], Kerrie Barry[7], Khouanchy Oakgrove[7], Yi Peng[7], Jeremy Schmutz[7], Patrick P. Edger[8, 9], Eric Lyons[10], J. Chris Pires[2, 11, 12], Gavin C Conant[1, 13, 14]

[1]North Carolina State University (United States), [2]University of Missouri - Columbia (United States), [3]University of Missouri - Columbia (United States), [4]Cornell University (United States), [5]Oberlin College (United States), [6]Clarion University of Pennsylvania (United States), [7]Lawrence Berkeley National Laboratory (United States), [8]Michigan State University (United States), [9]Michigan State University (United States), [10]University of Arizona (United States), [11]University of Missouri - Columbia (United States), [12]University of Missouri - Columbia (United States), [13]North Carolina State University (United States), [14]North Carolina State University (United States)

Polyploidy is a driving force of both evolutionary innovation and ecological success. In allopolyploid, it has been claimed that post-polyploidy gene retention often favors one of the two subgenomes. However, most analyses of *biased fractionation* are limited to single or pairwise genome comparisons, potentially giving rise to artifactual estimates. Using our likelihood-based tool POInT (Polyploid Orthology Inference Tool), we model the resolution of At-α allopolyploidy in the *Arabidopsis thaliana* and its relatives by phasing syntenic regions of six genomes with respect to each other. We find statistically robust evidence for the existence of biased fractionation. More importantly, we show that this bias was not confined to the earliest phases of post-WGD evolution. We also show that a driver of this pattern of biased losses is the co-retention of genes that are members of co-evolved functional complexes from the same parental genome. Meanwhile, to better understand the evolutionary pressures acting on surviving duplicates from recent plant polyploidies (the At-α duplication and more recent *Brassica* hexaploidy Br-α), we compare the strength and direction of selection acting at the species and population levels. We show that genes retained after polyploidy are intrinsically more constrained; even though genetic redundancy appears to relax the selective pressure to some degree. Importantly, the intensified purifying selection acting on retained duplicates are still detectable in populations at the present day. We conclude that biased fractionation and preferential retention are long-term forces shaping the evolution of paleopolyploid genomes, suggesting even yesterday's polyploids still have distinct evolutionary trajectories.

POA-377

Human Segmental Duplications Revisited

Marina Braso-Vives[1], Diego Andres Hartasanchez[2], David Juan[1], Arcadi Navarro[1]

[1]Universitat Pompeu Fabra - CSIC (Spain), [2]Universite Claude Bernard (France)

Duplications have had a major role in the evolution of eukaryotic genomes. Recent (>90% of identity between copies) and large (>1Kb in length) duplications, known as segmental duplications, conform around 5% of the human genome. Despite their importance, very little is known about their origin, molecular evolution and contribution to adaptation. Here we revisit them in order to elucidate the mechanisms through which they arise, evolve, and conduce to the birth of novel functional regions in the human genome. We combine two major duplication detection methods benefiting from their complementary advantages. We find huge differences between different types of segmental duplications that point to different origins and evolutionary dynamics. Most of the duplicated sequence of our genome is conformed by complex duplicated regions generated by multiple rounds of duplication. Moreover, intrachromosomal duplications are mainly generated through retrotransposon-mediated duplication; interchromosomal duplications are smaller and include most of the duplications generated through retrotransposition; and tandem duplications are younger and mainly generated through non-allelic homologous recombination. Furthermore, human segmental duplications are a scale-free network and, importantly, different types of segmental duplications have distinct dispositions within the network. Isolated segmental duplications tend to spread copies in much complex duplicated regions with which they maintain similarity through time. These complex duplicated regions might be acting as evolutionary repositories. By expanding our knowledge of the characteristics and evolution of segmental duplications we are taking a step forward in the understanding of the processes through which new human genes and other functional regions arise and evolve.

POA-379

The dark side of duplications: what to expect, what to look for, and their collapse

Diego A Hartasanchez[1, 2, 3], Marina Braso-Vives[2, 3], Txema Heredia[2, 3], Marc Pybus[2, 3], Arcadi Navarro[2, 3, 4]

[1]Universite de Lyon 1 (France), [2]Universitat Pompeu Fabra - CSIC (Spain), [3]Universitat Pompeu Fabra (Spain), [4]CRG (Spain)

The study of Segmental Duplications (SDs) and Copy-Number Variants (CNVs) is of great importance in the fields of genomics and evolution. However, SDs and CNVs are usually excluded from genome-wide selection scans since they are a source of confounding factors for most neutrality tests. Due to high identity between copies, low frequency CNVs, which are usually not in the reference, are prone to be collapsed when aligning sequence data from single individuals to the reference, even if repeat regions are masked. Such collapsed regions, which will be considered as single-copy when analyzing data, are challenging, because concerted evolution between duplications alters their site frequency spectrum and linkage disequilibrium patterns. Thus, summary statistics traditionally used to detect the action of natural selection on DNA sequences cannot be applied to SDs and CNVs. To investigate the potential effect of collapsed duplications upon natural selection scans we have obtained expectation values for ten summary statistics for duplications evolving under a range of interlocus gene conversion and crossover rates. We observe that in some cases values for known duplications mimic selective signatures. However, both known and collapsed duplications can be differentiated from single-copy regions or regions under selective pressure with test statistics that measure levels of nucleotide and haplotype diversity. Contrary to our expectations, we find that regions with low (and not high) nucleotide and haplotype diversity are enriched in duplications in a human African populations. This pattern might be due to the strict filtering applied by SNP-calling algorithms.

POA-380

Inferring ancestral state before WGD from enhancers and CTCF/cohesin loops in developmental genes.

Kenta Sumiyama[1]

[1]RIKEN (Japan)

When estimating traits of vertebrate common ancestor, the most outgroup of extant vertebrates is the cyclostomes such as lamprey. Since the lamprey has already experienced WGD (Whole Genome Duplication), it was difficult to estimate traits of vertebrate ancestor before WGD from genome information. We focused on the chromatin three-dimensional structure (CTCF/cohesin-loop) information and tissue-specific enhancer activity of duplicated developmental genes by WGD, and attempted to estimate the state of their common ancestor. One example is that the Dlx gene clusters has a characteristic TAD structure and we found that the 3D structure of chromatin which controls branchial arch-specific enhancer activity is shared among paralogs. Based on this observation, it is likely that a common ancestor before WGD already had a mechanism to control the expression in branchial arches. Thinking together with the lamprey Dlx expression pattern in arches, it seems that the common ancestor had rather uniform expression pattern in the branchial arches, but not like nested expression found in mammals. Another example is the Gsx genes (parahox genes), where enhancer / chromatin loop structure that controls Gsx expression in LGE, which is a novel trait of vertebrate, is shared among paralogous TADs. Thus it is likely that ancestral vertebrate before WGD already possessed LGE structure in its forebrain. We propose that cis-regulatory architecture together with 3D genome information is quite useful tool inferring ancestral state before WGD.

POA-381

Genomic Responses to Fungal Allopolyploidy in a Natural Biological Experiment

David Winter[1], Nikki Charlton[2], Carolyn Young[2], Murray Cox[1], Austen Ganley[3]

[1]Massey University (New Zealand), [2]Noble Research Institute (United States), [3]University of Auckland (New Zealand)

Polyploidy involves changes in the number of genome sets carried by an organism, and is observed in many eukaryote lineages. As outlined in Ohnos classic *Evolution by gene duplication* book, polyploidy is thought to be an important evolutionary process because it provides superfluous copies of genes that can be repurposed. Consistent with this, correlations between major evolutionary diversifications and polyploidy events have been documented, and polyploids often have increased fitness. Despite this potentially pivotal role in evolution, the genomic consequences of polyploidy remain poorly understood. Here we utilize high-throughput sequencing approaches to investigate the genomic consequences of allopolyploidy in an emerging model system for studying polyploidy, the *Epichloe* fungal endophytes. We investigated how the genome and gene expression respond to allopolyploidy using a natural biological experiment with two *Epichloe* species formed by independent allopolyploidy events from the same parental species. We find little change in expression for most genes following allopolyploidy. Remarkably, however, in almost every case where a gene is only expressed in one of the two parental species, there is no expression from either copy in both allopolyploids. This suggests that gene-silencing mechanisms present in one parent are dominant in the allopolyploids. We will also report on what our long-read sequence data reveal about chromosomal rearrangements between the two genomes and gene loss in the allopolyploids. Our work builds a picture for how the genome responds at the genomic and transcriptomic levels to allopolyploidy, and provides the basis for understanding what underlies phenotypes characteristic of polyploid species.

POA-382

Concerted Divergence after Gene Duplication in Polycomb Repressive Complexes

Yichun Qiu[1], Shao-Lun Liu[2], Keith Adams[1]

[1]University of British Columbia (Canada), [2]Tunghai University (Taiwan)

Duplicated genes are a major contributor to genome evolution and phenotypic novelty. There are multiple possible evolutionary fates of duplicated genes. Here, we provide an example of concerted divergence of simultaneously duplicated genes whose products function in the same complex. We studied POLYCOMB REPRESSIVE COMPLEX2 (PRC2) in Brassicaceae. The VERNALIZATION (VRN)-PRC2 complex contains VRN2 and SWINGER (SWN), and both genes were duplicated during a whole-genome duplication to generate FERTILIZATION INDEPENDENT SEED2 (FIS2) and MEDEA (MEA), which function in the Brassicaceae-specific FIS-PRC2 complex that regulates seed development. We examined the expression of FIS2, MEA, and their paralogs, compared their cytosine and histone methylation patterns, and analyzed the sequence evolution of the genes. We found that FIS2 and MEA have reproductive-specific expression patterns that are correlated and derived from the broadly expressed VRN2 and SWN in outgroup species. In vegetative tissues of Arabidopsis thaliana, repressive methylation marks are enriched in FIS2 and MEA, whereas active marks are associated with their paralogs. We detected comparable accelerated amino acid substitution rates in FIS2 and MEA but not in their paralogs. We also show divergence patterns of the PRC2-associated VERNALIZATION5/VIN3-LIKE2 that are similar to FIS2 and MEA. These lines of evidence indicate that FIS2 and MEA have diverged in concert, resulting in functional divergence of the PRC2 complexes in Brassicaceae. This type of concerted divergence is a previously unreported fate of duplicated genes. In addition, the Brassicaceae-specific FIS-PRC2 complex modified the regulatory pathways in female gametophyte and seed development.

POA-384

Exploring the genomic structure of the root-knot nematode Meloidogyne enterolobii

Marine Poullet[1, 2], Georgios Koutsovoulos[2], Etienne Danchin[2]

[1]Nice Sophia Antipolis University (France), [2]Institute of National Agronomical Research (France)

Root-knot nematodes (genus *Meloidogyne*) are obligatory plant endoparasites that cause huge economic loss in the agricultural industry and impact the global food supply. The most virulent and widely distributed *Meloidoyne* species worldwide reproduce asexually. *M. enterolobii* is an obligatory asexual species and is able to reproduce on tomato and pepper cultivars that are resistant to the other members of the genus *Meloidogyne*. In order to further understand the high parasitic success despite the absence of sexual reproduction, we will explore and analyse its genomic structure.

Using PacBio and Illumina sequencing technologies, we assembled the most contiguous *Meloidogyne* genome to date. Using MCScanX, we found that 55.66% of the genes are part of duplicated collinear blocks, possibly as a result of hybridization or multiple segmental duplications. We will look at how collinear genes evolve in these regions and their possible effects on the species parasitic success.

In some instances, the duplicated blocks form palindromic structures within the same scaffold. The genome structure is consistent with the allopolyploid structure recently described by our team in other *Meloidogyne* species with obligatory asexual reproduction (e.g *M. arenaria*, *M. javanica* and *M. incognita*). The palindromes, also observed in *M. incognita* and *M. arenaria*, are consistent with the absence of segregation between homologous chromosomes during meiosis.

These results could shed light into the surprising parasitic success and adaptation of *Meloidogyne* asexual species in multiple environments and hosts.

POA-385

Multi-Species Comparison of Potential Enhancers Found Across the Paramecium Aurelia Species Complex

Timothy James Licknack[1, 3], Weibo Zheng[2, 3], Michael Lynch[1, 3]

[1]Arizona State University (United States), [2]Ocean University (China), [3]Arizona State University (United States)

The Paramecium Aurelia species complex contains several morphologically similar species that differ vastly in their genome size due to two, ancient whole-genome duplications (WGDs). Much effort has gone into understanding the evolutionary fate of these duplicated genes, with particular focus on whether gene copies are retained over time, lost over time, acquire new functions, or partition ancestral functions. Equally interesting is how the regulatory regions of these genes have changed over time. Our lab has sequenced the genome of 15 of these species and computationally identified dozens of motifs located in promoters which may serve as possible enhancers. We are in the process of injecting some of these motifs as reporter constructs into the Paramecium tetraurelia macronucleus to determine if these motifs are actually enhancers. After this, we will manipulate these motifs to see if their position or copy number influences transcriptional output. This will be done in several species to determine how the nuclear environment in each cell differs across the complex, particularly between species that did and did not undergo the same ancient WGD. We will also look at the genomic distribution of these potential enhancers to determine if they lay in promoters of genes which serve similar functions.

POA-387

Adaptation of A-to-I RNA editing in Drosophila

Jian Lu[1]

[1]Peking University (China)

Adenosine-to-inosine (A-to-I) editing is hypothesized to facilitate adaptive evolution by expanding proteomic diversity through an epigenetic approach. However, it is challenging to provide evidences to support this hypothesis at the whole editome level. In this study, we systematically characterized 2,114 A-to-I RNA editing sites in female and male brains of D. melanogaster, and nearly half of these sites had events evolutionarily conserved across Drosophila species. We detected strong signatures of positive selection on the nonsynonymous editing sites in Drosophila brains, and the beneficial editing sites were significantly enriched in genes related to chemical and electrical neurotransmission. The signal of adaptation was even more pronounced for the editing sites located in X chromosome or for those commonly observed across Drosophila species. We identified a set of gene candidates that had nonsynonymous editing events favored by natural selection. We presented evidence that editing preferentially increased mutation sequence space of evolutionarily conserved genes, which supported the adaptive evolution hypothesis of editing. We found prevalent nonsynonymous editing sites that were favored by natural selection in female and male adults from five strains of D. melanogaster. We showed that temperature played a more important role than gender effect in shaping the editing levels, although the effect of temperature is relatively weaker compared to that of species effect. We also explored the relevant factors that shape the selective patterns of the global editomes. Altogether we demonstrated that abundant nonsynonymous editing sites in Drosophila brains were adaptive and maintained by natural selection.

POA-388

Newly-originated A-to-I RNA Editing Events Rapidly Evolve as Functional Regulator of RNA Subcellular Localization in Primates

Ni A. An[1], Jiguang Peng[1], Xin-Zhuang Yang[1], Jia-Yu Chen[1], Chuan-Yun Li[1]

[1]Peking University (China)

Recent studies have revealed thousands of A-to-I RNA editing events in primates, while the functions of these events are not well addressed. Here we performed comparative editome study in human and rhesus macaque, and uncovered a group of 317,938 species-specific A-to-I editing sites. These species-specific RNA editing events are selectively constrained in general, indicating a potential role in species-specific traits. Comparative transcriptome studies in fractional human and macaque brain tissues further revealed their contribution to cross-species differences in mRNA subcellular localization. Strikingly, knockout of ADAR1 in human cells attenuated the cross-species differences for the subcellular localization of mRNAs with solely species-specific RNA editing, strengthening the functional relevance of these editing events in transcript subcellular localization. Overall, we reported a new model for species-specific A-to-I editing events in primates and demonstrated that these specific RNA editing events may quickly acquire functionality through modulating mRNA subcellular localization.

POA-389

Evolution of structural and abundancy profiles in vertebrate mitochondrial mRNAs

Yao Sun[1], Masaki Kurisaki[1], Yasuyuki Hashiguchi[2], Yoshinori Kumazawa[1]

[1]Nagoya City University (Japan), [2]Osaka Medical College (Japan)

Genes encoded in vertebrate mitochondrial DNAs are transcribed as a polycistronic transcript for both strands, which is later processed into individual mRNAs, rRNAs and tRNAs, followed by modifications, such as polyadenylation at the 3' end of mRNAs. Although mechanisms of the mitochondrial transcription and RNA processing have been extensively studied using some model organisms, the variability of mitochondrial mRNAs in their structure and abundancy across different groups of vertebrates is poorly understood. We used the high-throughput RNA sequencing data to identify major polyadenylation sites for mitochondrial mRNAs in sixty species representing diverse vertebrate groups. We also inferred approximate locations for the 5' end of these mRNAs using the RNA Sequencing data. The results showed that nearly a half of the species had distinct polyadenylation sites from human counterparts in mRNAs for ND5, ND6, ND1, etc, revealing some structural plasticity of mitochondrial mRNAs during vertebrate's evolution. Some changes in the polyadenylation profiles appeared to be due to mitochondrial gene rearrangements and others to gene overlaps that triggered to create di-/tri-cistronic mature mRNAs. Relative abundancy of mitochondrial mRNAs estimated using the RNA Sequencing data showed higher mRNA abundancy for cytochrome oxidase and ATPase subunits than those for NADH dehydrogenase subunits in agreement with the relative abundancy of respiratory chain complexes on the mitochondrial membranes.

POA-390

What is the selective advantage of the widespread nonsynonymous A-to-I RNA editing in coleoids?

Daohan Jiang[1], Jianzhi Zhang[1]

[1]University of Michigan (United States)

A-to-I RNA editing converts the base adenosine (A) in RNA molecules to inosine (I), which is recognized as G in translation. Pervasive A-to-I editing and signatures of positive selection for nonsynonymous editing were reported in coleoids (octopuses, squids and cuttles), but the editing's benefit is unknown. We propose that it is the ancestral protein sequence and function before the emergence of widespread editing that are selected for. Specifically, in the presence of a high editing activity, some genomic positions that used to accept G but not A now accept A, because A can be edited back to G at the RNA level. The fixation of G-to-A mutations at these sites is likely slightly deleterious because the editing level (i.e., proportion of RNA molecules edited) is usually much lower than 100%. Consequently, mutations increasing the editing levels of these sites are compensatory and selected for. To test this hypothesis, we compare two classes of nonsynonymous editing events, restorative editing, which converts the corresponding amino acid back to the ancestral state, and diversifying editing, which converts the amino acid to a non-ancestral state. Both the editing level and fraction of sites edited are higher for restorative editing than synonymous editing, demonstrating positive selection for restorative editing. By contrast, diversifying editing shows signals of purifying rather than positive selection. Because restorative editing merely restores the ancestral state prior to the origin of pervasive editing, the reported positive selection for RNA editing in coleoids doesn't mean that RNA editing is adaptive.

POA-393

Evolutionary landscape and spatiotemporal dynamics of A-to-I RNA editing across metazoan species

Li-Yuan Hung[1], Yen-Ju Chen[1, 2], Te-Lun Mai[1], Chia-Ying Chen[1], Min-Yu Yang[1], Tai-Wei Chiang[1], Yi-Da Wang[1], Trees-Juen Chuang[1, 2]

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

Adenosine-to-inosine (A-to-I) editing, which converts genetically transcribed adenosine into inosine at the RNA level, is widespread across the kingdom Metazoa. Here we describe a knowledge-based framework to identify high-confidence editing sites by analyzing RNA sequencing data alone, without using single-nucleotide polymorphism (SNP) information, which remain largely absent for most non-model species. We thereby unveil the evolutionary landscape of A-to-I editing maps across 3 distant phyla, including 4 nematodes (*Caenorhabditis* species), 4 fruit flies (*Drosophila*species), and 12 vertebrates (from zebrafish to human). The evolutionary landscape provides unprecedented evidence on how A-to-I editing gradually expands its substrates and increases its influence along the history of evolution, and suggests that cross-species shared nonsynonymous A-to-I editing events are beneficial. Our result revealed that highly clustered and conserved editing sites tended to have a higher editing level and a higher magnitude of the ADAR motif. The ratio of the frequencies of nonsynonymous editing to that of synonymous editing remarkably increased with increasing the conservation level of A-to-I editing. These results thus suggest potentially functional benefit of highly clustered and conserved editing sites. In addition, spatiotemporal dynamics analyses reveal a conserved enrichment of editing and ADAR expression in the central nervous system throughout more than 300 million years of divergent evolution in complex animals and the comparability of editing patterns between invertebrates and between vertebrates during development. Our study thus offers an evolutionary explanation for the answer why the orchestra of transcriptome needs its editing repertoire, in addition to the blueprint of life.

POA-395

Diversification of transcription factor pulsing dynamics is driven by phosphorylation site evolution in intrinsically disordered regions

Ian Shen Hsu[1], Alan Moses[1]

[1]University of Toronto (Canada)

Several examples of transcription factors that show stochastic, unsynchronized pulses of nuclear localization have been recently described. The pulsing dynamics of these transcription factors have been shown to affect function in signaling pathways. Some pulsing transcription factors contain a high proportion of intrinsically disordered regions that contain nuclear localization signals, nuclear export signals, and a large number of experimentally confirmed phosphorylation sites. This sequence property is consistent with the Conformational Switch Model: the conformational change in intrinsically disordered regions is affected by posttranslational modification on the large number of phosphorylation sites, and the conformation of these regions affects the localization of pulsing transcription factor and leads to the stochastic pulses because the regions contain the signals of nuclear transport. If the Conformational Switch Model is correct, then we expect to see a large number of phosphorylation sites conserved through evolution around signals of nuclear transport in the homologs of pulsing transcription factors. Furthermore, several pairs of paralogous transcription factors have been shown to pulse but differ in their pulsing dynamics. Because the phosphorylation sites on a pulsing transcription factor are modified by different kinases, an evolution by duplication may happen when only the phosphorylation sites of a certain kinase are conserved in one paralog but not the other. Therefore, we are interested in understanding how well the number of phosphorylation sites of different kinases is conserved between paralogous pulsing transcription factor, and experimentally manipulate the phosphorylation sites to observe the effect on pulsing dynamics.

POA-397

Coupling between sequence and function in evolution of the binding sites of the male-specific lethal complex in Drosophila

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

[1]Sun Yat-sen University (China)

The male-specific lethal (MSL) complex mediates dosage compensation in Drosophila. The MSL complex comprises five proteins (MOF, MSL1, MSL2, MSL3 and MLE) and two non-coding RNAs (roX1 and roX2). In addition to the MSL complex, MOF also resides in the NSL (nonspecific lethal) complex that acts as a genome-wide transcriptional regulator in Drosophila. Adaptive evolution in all proteins of the MSL complex has occurred specifically in D. melanogaster. Here we report a comparative analysis of the binding sites of MOF, MSL1 and MSL2 between D.melanogaster and D.simulans. We found MOF and MSL1 gained many lineage-specific binding sites in D. melanogaster that are not overlapped with MSL2. Despite the increases of binding affinity and binding length for the overlapped binding sites in D. melanogaster, the non-overlapped sites specific to D. melanogaster have no preference in chromosome distribution and are mainly from retrotransposons especially for MOF. Population genetic analysis showed the overlapped and non-overlapped binding sites followed different evolutionary trajectories. Non-overlapped MOF binding sites is associated with increased expression divergence between D. melanogaster and D. simulans, which cannot be attributable to transposable element (TE) insertion polymorphisms. Our results suggest evolutionary plasticity of the binding sites of the MSL complex, wherein sequence variation is coupled with the changes of pleotropic functionality during the evolution of closely related Drosophila species. TE co-option might play a role in this process.

POA-398

The unreasonable effectiveness of population genetic inference via image recognition

Daniel Schrider[1]

[1]University of North Carolina (United States)

The availability of population-scale genomic datasets has given researchers a new avenue toward answering questions about populations' recent evolutionary histories. Such work has included efforts to infer demographic events such as population size changes and gene flow among closely related populations/species, the construction of genetic maps from genetic polymorphism data, and scans for loci underlying recent adaptation. In recent decades a host of theoretical and methodological advances have addressed these problems. Typically these methods summarize patterns of genetic variation using a summary statistic designed to be sensitive to the phenomenon of interest. More recently, machine learning approaches have proved successful in simultaneously examining many of these statistics in order to make far more accurate inferences. The rationale for these approaches is that any single statistic will capture only a subset of the discriminatory information present in the original data, and thus a set of complementary statistics will perform better. A more fruitful approach would thus be to perform inference directly on the input sequence data rather than digesting it into a set of numbers. Here we attempt to accomplish this be representing a population genetic alignment as an image and using modern deep learning techniques for image processing. We apply this approach to the problems listed above, and find that in each case it matches or exceeds the accuracy of current state-of-the-art methods. Thus, when applied to images of alignments, modern image recognition algorithms outperform expert-derived statistics and even collections thereof.

POA-400

Phylonumerics: A New Mass-Based Molecular Evolution Approach Investigates the Emergence of Antiviral Resistance in the Influenza Virus

Kevin Downard[1], Elma Akand[1]

[1]University of New South Wales, Sydney (Australia)

Molecular based approaches to phylogenetic analysis, driven by technological advances in gene or whole genome sequencing and bioinformatics, have revolutionized our view of evolution. We have developed a new and novel mass based phylonumerics approach and algorithm to study the evolution of any organism from the protein perspective using datasets commonly employed in proteomics. Sets of peptide masses, known as mass maps, rather than gene or protein sequence data, can be employed to construct phylogenetic trees and these mass trees used to trace and study the evolutionary history of organisms from which the proteins are derived. Furthermore, mass differences associated with single amino acid mutations can be charted and interrogated across the tree.

Understanding the mechanisms by which antiviral drug resistance mutations manifest and are compensated for remains elusive despite its importance to improving responses to the influenza virus. The approach is shown to be able to investigate the emergence of antiviral resistance mutations in influenza neuraminidase. Frequent ancestral and descendant mutations to antiviral resistance mutations are identified in N2 neuraminidase. The majority occur in the head region around the active site and drive hydrophilicity changes, primarily through the incorporation or loss of hydroxyl groups. The mass tree phylonumerics approach allows the evolution of influenza to be viewed from a global protein perspective and putative epistatic and compensatory mutations, remote in their sequence and structure, to be proposed.

POA-401

Classifying ENU induced mutations from spontaneous germline mutations in mouse with machine learning techniques

Yicheng Zhu[1], Cheng Soon Ong[2], Gavin Huttley[1]

[1]ANU (Australia), [2]CSIRO (Australia)

Understanding sequence diagnostic signatures associated with mutagenesis mechanisms can facilitate the development of more accurate models for identifying disease causing mutagens. In most genetic variation catalogs, variant data are derived from a mixture of mutagenic processes. This presents a challenge to assigning the mechanistic origins of an individual variant. Information regarding the mechanistic origin of point mutations is present in surrounding DNA sequence. These motifs can reflect the combination of chemical and biochemical influences of neighbouring bases on mutagenesis. We assess whether information from sequence neighborhood can be used to identify mutations resulting from the potent chemical mutagen, ENU. ENU is a synthetic chemical employed in mutagenesis studies, introducing novel point mutations to genomes. We developed a machine learning classifier to discriminate between ENU-induced and spontaneous point mutations in the mouse germline. Our classification results reveal that a combination of k-mer size and representation of second-order interactions among nucleotides was able to improve classification performance in comparison to the naive classifier approach.

POA-402

Orthology assignment of 5 novel Sparidae proteomes and their phylogenetic position among teleosts

Paschalis Natsidis[1, 2], Pavlos Pavlidis[3], Costas Tsigenopoulos[1], Tereza Manousaki[1]

[1]Hellenic Centre for Marine Research (Greece), [2]University of Crete (Greece), [3]Foundation for Research and Technology (Greece)

The Sparidae are a family of teleosts constituted of fish with high commercial value, such as seabreams and porgies. The phylogenetic relationships among Sparidae species, as well as between Sparidae and other teleost families has been studied, only with the use of single or few markers with controversial results. Here, we integrated 5 recently established Sparidae gene sets with other publicly available well-annotated teleost proteomes, comprising a comprehensive dataset of whole-genome information for 31 species. This dataset was then used to infer orthologous genes using two different algorithms, and conduct a phylogenomic analysis to determine the position of Sparidae within the teleost phylogeny.

POA-403

Excision-reintegration at a pneumococcal phase-variable restriction-modification locus drives within- and between-strain epigenetic differentiation and inhibits gene acquisition

Min Jung Kwun[1], Marco R Oggioni[2], Stephen D Bentley[3], Nicholas J Croucher[1]

[1]Imperial College London, London (United Kingdom), [2]University of Leicester (United Kingdom), [3]Wellcome Sanger Institute (United Kingdom)

Restriction-modification systems (RMS) have an important role in cellular defense against infection by mobile genetic elements, and can also affect phenotypes, such as virulence, through epigenetic regulation. Here we characterize the phase-variable SpnIV type I R-M system, encoded by the translocating variable R-M (tvr) locus.We demonstrate a novel mechanism by which segments of DNA encoding target recognition domains, which determine the R-M system's specificity, are moved between ~70 bp direct repeats within the locus via a circular intermediary form. This process typically occurs over hours, at a rate limited by a transcriptional attenuator and toxin-antitoxin-type system within the locus. By characterizing a panel of 'locked' mutants using methylation-sensitive sequencing. Genomic data also allows us to explore the extensive sequence variation present within this locus across pneumococcal populations, revealing an extensive repertoire of specificity subunits. Hence the *tvr*locus can drive many different patterns of methylation both within and between different strains, giving it the potential to substantially influence both gene expression and dynamics of mobile genetic element transmission between pneumococci.

POA-404

Convergent somatic mutations in asexual pathogen *Phytophthora ramorum* NA1 contributes to population genetic diversity

Jennifer D. Yuzon[1], Renaud Travadon[1], Madhu Malar C.[2], Sucheta Tripathy[2], Nathan Rank[3], Heather Mehl[1], Richard Cobb[1], Tedmund Swiecki[4], Elizabeth Bernhardt[4], Corinn Small[1], Tiffany Tang[1], David Rizzo[1], Matteo Garbelotto[5], Takao Kasuga[6, 1]

[1]University of California, Davis (United States), [2]CSIR Indian Institute (India), [3]Sonoma State University (United States), [4]Phytosphere Research (United States), [5]University of California, Berkeley (United States), [6]USDA Agricultural Research Service (United States)

Asexual reproduction implies limited genetic diversity compared to sexual reproduction. Yet, somatic mutations can rapidly increase an asexual lineage's genetic diversity. *Phytophthora ramorum* NA1 is an invasive plant pathogen that was introduced to the Pacific Coast of the United States. Since its introduction in the mid 1990s, *P. ramorum* has propagated clonally and successfully spread to forest ecosystems throughout the north coast of California. Previous genomic analyses of *P. ramorum* have identified the presence of genetic diversity in the form of Structural Variants (SVs). To understand how SVs contribute to the evolution of *P. ramorum*, we reconstructed genealogical relationship between 47 isolates and observed the distribution of SVs in the phylogeny. First, SVs were identified using read depth, split-read, and paired-end mapping methods. All SVs called by read depth method were included. SVs called by split-read and paired-end methods had to intersect with at least one alternative SV caller to be included in our final SV dataset (57 copy number increases, translocations, and deletions). To estimate SVs at ancestral nodes, phylogenetic relationships were reconstructed using Single Nucleotide Variants (SNVs). On average, each SV arises 2.6 times independently throughout the phylogeny. Only two internal nodes had a convergent SV that was transmitted to extant lineages. The extant lineages occur in geographically distant locations (Monterey and Sonoma counties). The present study represents another trajectory for the evolution of asexual organisms and how somatic mutations contribute to genetic variation.

POA-405

Excess of movement out of the X chromosome across 250 million years of Dipteran evolution

Melissa Toups[1], Beatriz Vicoso[1]

[1]Institute of Science and Technology (Austria)

The presence of differentiated sex chromosomes is accompanied by an excess of gene movement out of the X chromosomes in several taxa. Hypotheses for this movement have involved both mechanistic (excess of retroposition in testes-expressed genes) and selective explanations (sexually antagonistic selection and escape from MSCI), yet taxon-specific analyses have not yet fully disentangled these effects. Here, we develop a pipeline to infer gene movement on draft genomes, even in the absence of direct chromosomal information. This allows us to examine gene movement across 10 transitions in sex chromosomes over 250 million years of Dipteran evolution. We find excess out of the X chromosome gene movement to be ubiquitous among Dipterans, independent of the chromosome used for sex-determination. We test for effects of duplication mechanism by assessing DNA-mediated and RNA-mediated duplicates separately, yielding insights into the role played by the different mutational processes. Finally, we take advantage of the well-described phylogeny of flies to infer ancestral gene expression, allowing us to assess the importance of sexual antagonism and other putative selective pressures that are hypothesized to have driven the out-of-X migration.

POA-406

Feathered Feet Are Just Winging It: Shifts in Pigeon Limb Identity Reveal Conserved Regulatory Networks

Elena F Boer[1], Hannah F Van Hollebeke[1], Carlos R Infante[2, 4], Douglas B Menke[3], Michael D Shapiro[1]

[1]University of Utah (United States), [2]University of Arizona (United States), [3]University of Georgia (United States), [4]University of Arizona (United States)

Deciphering the genetic mechanisms of morphological variation remains a critical challenge in biology. The domestic pigeon is an outstanding model to study the molecular changes that underlie morphological variation, as it displays striking phenotypic variation within a single species and is amenable to genetic crosses, embryonic studies, and genomic analyses. While most pigeons have scales on their feet, some breeds have feet covered with feathered skin. We recently showed that cis-regulatory alleles of the limb-identity genes PITX1 and TBX5 play major roles in feathered feet, and that this striking trait results from a partial change in limb identity. To understand the gene regulatory networks downstream of PITX1 and TBX5 that cause foot feathering, we generated forelimb (FL) and hindlimb (HL) transcriptomes from scale, grouse, and muff pigeon embryos. Comparative analyses of HLs reveal a differentially expressed (DE) gene set that is enriched for transcription factors, extracellular matrix genes, and signaling pathways with key roles in limb development. A subset of the DE genes that distinguish scale, grouse, and muff HLs are also DE between pigeon FL and scale HL buds, suggesting a specific set of genes that is misregulated in the partial transformation from HL to FL identity. In addition, we compared pigeon limb bud transcriptomes to chicken, anole lizard, and mammalian datasets to identify TBX5- and/or PITX1-regulated components of an evolutionarily conserved limb GRN. Our analyses reveal a set of subtle gene expression changes that are conserved across amniotes to regulate the development of morphologically distinct limbs.

POA-407

Molecular evolution and functional diversity of doublesex, a master regulator of polymorphisms in insects

Saurav Baral[1], A Gandhimathi[1], Riddhi Deshmukh[1], Krushnamegh Kunte[1]

[1]National Centre for Biological Sciences (India)

doublesex is a well-known transcription factor that controls early sexual differentiation in insects. Apart from this conserved role, recent studies show that dsx also regulates sex-specific, sex-limited or morph-specific adaptive phenotypes during late developmental stages. How does a gene that is highly conserved and carries out a critical function also remain functionally dynamic to gain new, ecologically important functions that are divergent sometimes even in sister species? We performed an exon-level analysis to trace the molecular evolution and structural diversity of dsx from 144 insect species from orders Lepidoptera, Coleoptera, Diptera and Hymenoptera, representing 350 million years of divergent evolution. We discovered that the rate of evolution was variable in different genic regions: the functional domains were highly conserved, whereas the sex-specific domains and non-domain regions showed high variability and many sites under positive selection. Male-specific sequences evolved faster than the female-specific sequences. Homology modeling further revealed conserved domain regions and female-specific structures, whereas male-specific structures were highly divergent, which may be associated with the evolution of novel functions across sexes and lineages. Thus, highly conserved but dynamic master regulator genes may partition their specific conserved and novel functions with differential rates of molecular and protein evolution at deep timescales.

POA-408

Composite Likelihood Estimation of Phylogenies from Genomic Data using Coalescent Theory

Geno Guerra[1], Rasmus Nielsen[1, 2]

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

When comparing genetic data from multiple species, it is common to observe that different regions of the genome can produce conflicting phylogenetic tree topologies. One ubiquitous source of this discrepancy, incomplete lineage sorting (ILS), can be well characterized using the multispecies coalescent (MSC). Many methods have been developed in the past few years to address the presence of ILS using multi-gene sequences in place of the classical, however statistically-inconsistent, gene-concatenation method. Most of these methods take the two-step approach of reconstructing gene trees from multiple genomic segments by some phylogenetic method and then treating the constructed gene trees as observed data, while not properly accounting for the uncertainties. There has been criticism of this two-step approach finding that uncertainty in gene tree estimation from the mutational process can have large downstream effects on the accuracy of phylogenetic tree reconstruction.

We present COAL-PHYRE, a composite maximum likelihood coalescent-based method for inferring rooted species trees along with estimates of population sizes and divergence times from multiple gene sequences. COAL-PHYRE jointly models coalescent variation from the MSC and error in gene tree reconstruction using the bootstrap. We compare our method against ASTRAL, a statistically consistent leading quartet-based method, as well as against concatenation using neighbor joining. We show that we outperform both ASTRAL and concatenation under moderate to high levels of ILS in the presence of increasing levels of gene tree reconstruction error. We are also able to recover ancestral population sizes and divergence times with high accuracy under many settings.

POA-409

Repeated adaptive evolution of an enzyme in plant specialized metabolism

Arunraj Saranya Prakashrao[1], Elisabeth Kaltenegger[1], Dietrich Ober[1]

[1]University of Kiel (Germany)

Plants produce a vast variety of specialized chemical compounds also known as secondary metabolites to adapt to specific environmental conditions. This great variety of compounds is a result of the diversity of specialized metabolic pathways. Majority of these novel metabolic pathways originate via gene duplication of already existing genes from primary metabolisms and then evolve through natural selection. **Pyrrolizidine alkaloids (PA)** are specialized metabolites that are toxic compounds synthesized in plants to defend against insect herbivores. Homospermidine synthase (HSS) catalyzes the first and critical step in PA biosynthesis. HSS originated from deoxyhypusine synthase (DHS) through a gene duplication event. Here, we analysed the HSS recruitment to PA biosynthesis and the evolutionary path taken by duplicated genes using morning glory family as a model system. We identified both HSS and DHS sequences from 33 different species of morning glory using degenerative PCR. Phylogenetic analysis of these sequences showed that a single gene duplication event gave rise to extant HSS sequences in many species. Further, our molecular evolutionary analyses showed that duplicated HSS paralogs underwent varying selection pressures throughout their evolution including purifying, neutral, and positive Darwinian selections. In two different lineages of the morning glory family, we observed the above similar pattern of selections as above including identical amino acid replacements in the analysed HSS. Reconstruction of ancestral HSS sequences shows a pattern of repeated adaptive evolution of HSS in PA biosynthesis in the unrelated species of the morning glory.

POA-410

High resolution analysis of emerging mutations at very short timescale

Han Mai[1], Anton Nekrutenko[1]

[1]The Pennsylvania State University (United States)

In bacterial experimental evolution studies genomic changes are often determined via analysis of several timepoints -- aliquots that are drawn from the experimental populations, amplified, and subjected to DNA sequencing. While illustrative, such sampling can be viewed as an extreme case of bottleneck in which only genetic changes at high frequency are revealed but no information about the underlying mutational dynamics is retained. Here we present results of a turbidostat experiment in which at every timepoint approximately 2/3 of the evolving population is sampled and, without amplification, subjected to duplex sequencing designed to reveal genetic changes at very low frequencies. The high sensitivity of our mutation detection strategy allows us to trace variants as they occur through time within unprecedented resolution.

POA-411

c-genie - Assessing the impact of recombination on phylogenomic inference

Michael Matchiner[1], Milan Malinsky[1]

[1]University of Basel (Switzerland)

Genealogical relationships in multi-species datasets often vary along the genome as a result of incomplete lineage sorting (ILS) and recombination. The variation in genealogies may generate challenges for species tree reconstruction; however, the extent of these challenges is a matter of controversy. Some authors [e.g. Springer and Gatesy (2016)] argue that, in a typical phylogenomic dataset, stretches of the genome with a single underlying genealogy (c-genes) are so short that they invalidate fundamental assumptions of many species tree estimation methods. Others [e.g. Edwards et. al. (2016)] argue that changes in genealogies are not important if they do not change the tree topology and that within-locus recombination and ILS have minimal impact on species tree reconstruction. Despite compelling theoretical arguments on both sides, empirical evidence is limited to a study by Lanier and Knowles (2012), showing that recombination has minimal impact on the accuracy of reconstructing a simulated eight taxon tree using between 1 and 9 genes.

To address the above questions we developed the software c-genie, built on top the ultra-fast msprime coalescent simulator. Given a species tree, an N_e value, and a recombination rate/map, c-genie generates msprime code, simulates genealogies, finds topology breakpoints, simulates mutations, and outputs DNA sequence alignments for species tree reconstruction. Our extensive simulation inquiry into the effect of recombination and ILS includes both multi species coalescent and concatenation approaches, and genome-scale datasets corresponding to a 48 species bird phylogeny, and a 250 species phylogeny mimicking that of Lake Tanganyika cichlid fishes.

POA-412

Vomeronasal type 2-like receptors showed different evolutionary patterns in basal teleosts such as eels and in other teleosts

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

[1]Academia Sinica (Taiwan), [2]National Kaohsiung University of Science and Technology (Taiwan), [3]National Applied Research Laboratories (Taiwan)

Vomeronasal receptor family 2 (V2R) genes control odor sensitivity and pheromone specificity in fish species. As fish live in various natural habitats, their V2R genes may show high evolutionary divergences. However, there seems to be no study comparing the V2R genes in basal teleosts such as eels (Anguilliformes) with those in derived teleosts to infer the trend of evolution of V2R genes after the teleost-specific whole genome duplication (3R). Here we studied V2R genes in one Elops and 12 Anguilliformes species that live in habitats with various water depths. We used barcoding V2R amplicons and next generation sequencing to analyze V2R genes and compared them with those in other teleosts. A total of 15,962 complete amplicons with an average length of 344 bp were collected and 23 to 101 functional V2R genes were identified in each species. By comparing our data to published V2R genes in 25 derived teleosts, we found that many more V2R genes were retained in Anguilliformes genomes. However, a higher degree of pseudogenization was observed in Anguilliformes, especially in a few deep-sea eels, compared to derived teleosts. Using phylogenetic analysis, we found several Anguilliformes-specific V2R subfamilies. In conclusion, Anguilliformes species have retained a larger V2R repertoire than derived teleosts after the 3R event in the common ancestor of teleosts and V2R genes in Anguilliformes have diverged into subfamilies. Our study seems to be the most comprehensive investigation on the evolution of V2R genes in eel species and other teleosts.

POA-413

The Prediction of the Externally Visible Characteristics of Medieval Human Remains from Poulton, UK, using Next Generation Sequencing

Amanda June Skinner[1], Ashley Lynn May[1], Anders Gotherstrom[2], Linus Girdland Flink[1], Kyoko Moores Yamaguchi[1]

[1]Liverpool John Moores University (United Kingdom), [2]Stockholm University (Sweden)

This study examined the externally visible characteristics of medieval individuals excavated from Poulton, an archaeological site located in Cheshire, England. DNA was extracted from ten tooth samples in a dedicated ancient DNA laboratory at LJMU and constructed into double stranded sequencing libraries following Meyer and Kircher (2010). The libraries were sequenced on the HiSeq X platform at the National Genomics Infrastructure at SciLife, Stockholm, Sweden. Approximately 10 million sequences were obtained per individual/library. Sequences were mapped to the human reference genome using Burrows-Wheeler Aligner version 0.7.8. We used Integrative Genomics Viewer (IGV) to genotype the 24 SNPs that are required for use with the HIrisPlex system to predict eye and hair color. Four samples had partial allele information of the 24 SNPs, ranging from three to 13. The allele present was assumed to be either homozygous or heterozygous when the coverage is low (1 or 2x), so HIrisPlex was ran multiple times with different combinations of possible genotypes to obtain the range of probability for each phenotype. The sample with the highest number of SNPs present showed prospect of having blue eyes, with the probability range being 0.898 to 0.953. Additionally, our results showed possibility of dark blonde/dark brown hair assuming homozygosity of the observed allele when the coverage is low, as the homozygous state is shown with more frequency in Europeans. Future studies should focus on doing targeted next generation sequencing or using multiplex PCR to predict eye and hair color more precisely.

POA-414

Experimental Determination of the Rate of Muller's Ratchet in Escherichia coli

Joshua John Miranda[1], Mrudula Sane[1], Deepa Agashe[1]

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

Adaptation proceeds via spontaneous beneficial mutations and is retarded by deleterious mutations. However, spontaneous mutations are rarely beneficial, with the majority being deleterious or neutral. Deleterious mutations can thus constrain evolutionary trajectories, and are especially important in small populations where genetic drift dominates selection, potentially causing extinctions. Muller's landmark idea proposed that an asexual population, as a result of fixing random mutations, will decrease in fitness with increase in mutational load in an irreversible ratchet-like manner. Despite the importance of the ratchet in determining the fates of populations, few experimental measures of its rate exist. Moreover, theoretical models of the ratchet are limited by the lack of information on the fitness effect of spontaneous mutations and the extent of epistasis in mutational effects. Here, we experimentally quantify the effect of Muller's ratchet on the fitness of *Escherichia coli*. We estimate the deleterious mutation rate by quantifying the fitness effects of spontaneous mutations that accumulate under relaxed selection (mutation accumulation). As expected the average fitness across evolved lineages decreases and the among-lineage variation in fitness increases with increase in generations. We investigate the dependence of the rate of the ratchet on the basal mutation rate, the mutational spectrum and the environment. Additionally, we measure the contribution of epistasis to the observed fitness trajectories as compared to the prediction of a non-epistatic model. Our work is the first large-scale experimental determination of the rate of Muller's ratchet, with the deleterious mutation rate estimates of broad application to evolutionary theory.

POA-415

TBD