1

Testing the neutral hypothesis of phenotypic evolution

Jianzhi Zhang[1]

1. University of Michigan, Ann Arbor, MI, United States

It is generally accepted that a large fraction of genomic sequence variations within and between species are neutral or nearly so. Whether the same is true for phenotypic variations is a central question in biology. On the one hand, numerous phenotypic adaptations have been documented and even Kimura, the champion of the neutral theory of molecular evolution, believed in widespread adaptive phenotypic evolution. On the other hand, phenotypic studies are strongly biased toward traits that are likely to be adaptive, contrasting genomic studies that tend to be unbiased. It is thus desirable to test the neutral hypothesis of phenotypic evolution using traits irrespective of their potential involvement in adaptation. Here we present such a test for 210 morphological traits measured in multiple strains of the yeast Saccharomyces cerevisiae and two related species. Our test is based on the premise that, under neutrality, the rate of phenotypic evolution declines as the trait becomes more important to fitness, analogous to the neutral paradigm that functional genes evolve more slowly than functionless pseudogenes. Neutrality is rejected in favor of adaptation if important traits evolve faster than less important ones, parallel to the demonstration of molecular adaptation when a functional gene evolves faster than pseudogenes. After controlling the mutational size, we find faster evolution of more important morphological traits within and between species. By contrast, an analysis of 3466 yeast gene expression traits fails to reject neutrality. Further, intraspecific and interspecific variations in yeast gene expression conform to the phylogenetic relations of the strains rather than their ecological environments. Thus, yeast morphological evolution is largely adaptive, but the same does not apply to the transcriptome, suggesting that phenotypic variations at different levels are shaped by different evolutionary forces.

2

Branch-heterogeneous models of sequence evolution

Bastien Boussau[2, 1]

1. LBBE, CNRS, Lyon, France

2. LBBE, Université Lyon 1, Université de Lyon, Lyon, France

Early models of sequence evolution made several important regularity assumptions to describe the process of nucleotide or amino-acid evolution in a simple and manageable manner. Much of the methodological work in statistical phylogenetics over the past decades has been devoted to relaxing these assumptions. As a result, nowadays the most popular models of sequence evolution assume that not all transitions between nucleotides or amino-acids are equally likely, that different sites evolve independently, but that they evolve at different rates and even sometimes according to different transition matrices. These models however still make several patently unrealistic assumptions. In particular, they assume that sequence evolution has operated according to a unique transition matrix along the entire phylogeny, an assumption that is not realistic notably for all data sets in which nucleotide or amino-acid composition varies among sequences. In this talk I will review models of sequence evolution that relax this assumption, that I will globally call branch-heterogeneous models. I will explain why these models are particularly difficult to use in either the maximum likelihood or the Bayesian frameworks, and I will present our current efforts to improve their usability.

3

Methods to find similar sites in alignments

Rob Lanfear[1], **Brett Calcott**[2], **Paul Frandsen**[3]

1. Macquarie University, Sydney

2. University of Sydney, Sydney

3. Smithsonian Institution, Washington, D.C.

In this talk, I will give an overview of methods for choosing partitioning schemes for the phylogenetic analysis of genome-scale datasets. This involves moving from analyses in which users provide information on the heterogeneity of evolutionary processes among sites (e.g. by defining different loci or codon positions) to analyses in which the heterogeneity is determined automatically. To do this, we have to develop methods that are both fast and free of bias. Both of these things are challenging, and despite much progress in recent years, a number of open questions remain.

4

Among genes heterogeneity of the phylogenetic signal in genome data: causes, symptoms, and treatments

Omar Rota-Stabelli[1]

1. Fondazione Edmund Mach, San Michele All'Adige,, TN, Italy

Inference of phylogeny is currently based on the concatenation of many genes, a procedure that enables reducing the stochasticity associated with single gene phylogenies. All possible drawbacks of this approach are however not fully understood, particularly the among gene heterogeneity of the phylogenetic signal. I studied the distribution of phylogenetic signal in the model system *Drosophila* using two genome-scaled datasets. Although both datasets *apparently* resolve most of the relationships with high support when analysed at the nucleotide level, there are at least two types of among genes phylogenetic incongruences. First, the phylogenetic signal is not homogenously distributed among nuclear coding, mitochondrial coding, and non-coding genes, which robustly support competing topologies at some nodes, particularly close to tips. Second, the phylogenetic signal is not homogenously distributed among ontology classes, whereby nuclear genes involved with the metabolism tend to carry their own signal. Most, but not all of these incongruences, are due to substitutions at synonymous sites which I show being affected by different mutational pressures in different types of data. Counter intuitively, partitioning is not successful in alleviating these incongruences, which are instead revealed by using across-site heterogeneous model or by using a coalescent aware approach. These results advocate that extra care should be taken when interpreting high supports from the analysis of genome scaled phylogenies, and that signal associated with synonymous sites may be unreliable even at the genus level. Phylogenetic incongruences may be however extremely instructive in disentangling possible sources of systematic error, as well as in revealing peculiar aspects of species biology such as introgression or incomplete lineage sorting due to fast radiation.

5

Assessing Methods for Outlier Detection in Phylogenetic Inference

Daisy Shepherd[1], Steffen Klaere[1], Jale Basten, Ole Geldschlager

1. Department of Statistics, The University of Auckland, Auckland, New Zealand

Different sites in an alignment can support phylogenetic trees in various ways. Most notably, deep phylogenies tend to have issues with outlier or saturated sites, that can drastically change the favoured topology. In some situations, these sites can even drive the inference toward selecting the wrong topology.

To avoid systematic errors related to saturation issues, current approaches propose to remove these saturated sites to improve the stability of tree inference. These methods use statistics to determine which sites to remove, before performing the phylogenetic analysis.

A number of studies have applied these methods to alignment data prone to saturation issues. The results indicated a common breakthrough: the removal of saturated sites led to a change in the preferred phylogenetic tree.

However, these methods have a couple of issues: (1) It is not clear how the methods fare when the data have no systematic error, and (2) whether the amount of sites claimed as saturated is appropriate. In our study, we addressed the first issue by simulating "good" alignments and investigating the performance of the proposed statistics. Further, we addressed the second problem by comparing which sites were declared saturated under each statistic, whilst invoking the discrete nature of the OV distance as a means of visualisation of the scores.

6

Morphological and molecular convergences in phylogenetic inference

Zhengting Zou[1], Jianzhi Zhang[1]

1. University of Michigan, Ann Arbor, MICHIGAN, United States

Phylogenetic trees reconstructed from molecular sequences are often considered more reliable than those reconstructed from morphological characters, in part because convergent evolution, which confounds phylogenetic reconstruction, is widely believed to be rarer for molecular sequences than for morphologies. However, neither the validity of this belief nor its underlying cause is known. Comparing thousands of characters of each type that have been used for phylogenetic inference, we find that on average morphological characters indeed experience much more convergences than amino acid sites, but this disparity is explained by fewer states per character rather than an intrinsically higher susceptibility to convergence for morphologies than sequences. We show by computer simulation and actual data analysis that a simple method for identifying and removing convergence-prone characters improves phylogenetic accuracy, potentially enabling, when necessary, the inclusion of morphologies and hence fossils for reliable tree inference.

7

Unsupervised Detection and Quantification of Demographic Structure in mtDNA via Multiple Correspondence Analysis

Adam B Rohrlach[1], Nigel G Bean[1], Simon J Tuke[1], Julien Soubrier[2], Graham Gower[2]

1. School of Mathematical Sciences, The Univeristy of Adelaide, Adelaide

2. Australian Centre for Ancient DNA, The Univeristy of Adelaide, Adelaide

One aspect of interest when exploring genetic data can be detecting if the data comes from a structured population. For example, does the population come from a collection of geographically separated sub-populations, or does time better explain a shift in genetic signal? Principal Component Analysis (PCA) is a standard tool for exploring these types of relationships between genetic and demographic signal. However, when data includes ancient samples, sometimes only mtDNA can be successfully recovered. Classical unsupervised methods using PCA can not be applied to mtDNA, and so researchers are left without an efficient or well understood method for exploring data.

We suggest applying Multiple Correspondence Analysis (MCA) directly to mtDNA. MCA is an intuitive generalisation of PCA to categorical variables, and so can be used and presented in a similar way to standard nuclear DNA analyses. The result of applying MCA to mtDNA data is that we produce a quantitative representation of the categorical data, in fewer dimensions than the original alignment data. Importantly, MCA is an unsupervised method, and so no prior knowledge of the demographic structure of the population is required.

Using this method, we apply a medoid based clustering algorithm to explore genetic similarity and dissimilarity between individuals. We then show that the data, and the clustering structure, can be compared to any supplementary variables (such as latitude and longitude, time or culture) to test for significant correlation.

Finally, we apply our method to European human mtDNA spanning the Upper Paleolithic. Our method reveals evidence for a vast change in the human genetic landscape after the most recent ice age, consistent with other analyses.

8

Coordinating robust with plastic developmental responses in the developing fly wing

Christen Mirth[1], Marisa M Oliveira[2], Alexander W Shingleton[3]

1. Monash University, Clayton, VIC, Australia

2. Institute of Molecular and Life Sciences, University of Zurich, Zurich, Switzerland

3. Biology, Lake Forest College, Lake Forest, Illinois, USA

Growth throughout the body responds to a multitude of environmental conditions, generating variation in body and organ size in accordance with the developmental environment. For animals, this poses a particular challenge; animals need to coordinate the development of their organs to ensure that irrespective of plasticity in organ and body size, organ patterning remains robust to maintain correct organ function. We explore the genetic mechanisms that coordinate plastic versus robust developmental processes across a range of environmental and physiological conditions using the fruit fly,*Drosophila melanogaster.* To understand how the body coordinates its development with that of its organs, we devised a staging scheme for the developing wing disc. Using the changes in expression of three patterning genes in the wing, Achaete, Senseless and Dachshund, we correlated the progression of patterning in the wing with whole body developmental events in altered environmental and physiological conditions. We have found that the wing aligns its development at specific developmental milestones, at the moult to the third instar and at pupariation, but the progression of wing disc pattern varies significantly between these milestones. Next, we determined the role of the steroid hormone ecdysone, known to regulate moulting and metamorphosis, in regulating the progression of pattern and the growth of the wing disc. By

ablating the gland that produces ecdysone, we have found that the progression of Achaete, Senseless and Dachshund patterns differ in their requirement for ecdysone. Furthermore, we find that Senseless, but not Achaete or Dachshund, shows evidence of a disc size threshold for patterning to progress. Taken together, these data provide valuable insight into how developmental processes are coordinated across the body and how organs coordinate both plastic and robust processes throughout development.

9

The role of complex and young genes in the formation of organismal complexity during embryonic development

Dong Yang[1], Aishi Xu[1, 2], Jiayin Zang[1], Fuchu He[1]

1. State Key Laboratory of Proteomics, Beijing Proteome Research Center, National Center for Protein Sciences (Beijing), Beijing Institute of Radiation Medicine, Beijing, China

2. Animal Sciences College, Jilin University, Changchun, Jilin, China

During evolution, the increasing proportion of complex genes and young genes in the genome substantially contributes to the increasing of genome/organism complexity[1-4]. As for an individual organism, it is the process of embryonic development that makes the genome complexity be represented as the phenotypic complexity at organismal level. Our study focus on the basic question: how are the complex and young genes utilized during embryonic development to form the organismal complexity?

We found that complex genes, such as long genes, genes with multiple cis-regulatory modules[5], genes encoding long proteins or multi-domain proteins[1,3], tend to be utilized preferentially at each embryonic developmental stage, with the strongest strength of over-representation at mid-late stages. This trend is mainly due to complex genes with middle-age, which originated approximately during Cambrian Bio-radiation. On the other hand, young genes tend to be expressed at specific developmental stages. And it's obvious that young genes tend to be expressed in the early stage of the differentiation of embryonic stem cells (ESCs) into nerve cells. These main results are robust across protostomia and deuterostomia regardless of different technologies used to produce the data. These results indicate that complex genes and young genes contribute to the organismal complexity at two different levels: Complex genes contribute to the complexity of individual proteome at certain states, whereas young genes contribute to the diversity of proteomes at different spatial-temporal states.

This study also provided a new view for the relationship between evolutionary and developmental processes, which is a fundamental question in evolutionary developmental biology. Our results support "funnel-like model" in a new view, and can answer why there are different Evo-Devo-relation models, which gives new insights into Von-Bear theory.

1. Yang D, Zhong F, Li D, Liu Z, Wei H, Jiang Y, He F. General trends in the utilization of structural factors contributing to biological complexity. Mol Biol Evol 29.1957-1968 (2012).
2. Xu A, Li G, Yang D, Wu S, Ouyang H, Xu P, He F. Evolutionary Characteristics of Missing Proteins: Insights into the Evolution of Human Chromosomes Related to Missing-Protein-Encoding Genes. J Proteome Res14(12):4985-4994 (2015).
3. He X, Zhang J. Gene complexity and gene duplicability. Curr Biol 15, 1016-1021 (2005)
4. Lynch M, Conery JS. The origins of genome complexity. Science 302, 1401-1404 (2003)
5. Ferretti V, Poitras C, Bergeron D, Coulombe B, Robert F, Blanchette M. PReMod: a database of genome-wide mammalian cis-regulatory module predictions. Nucleic Acids Res 35, D122-126 (2007).

10

Transcriptomics of developmental evolution following long term selection

Josh Alpern[1], Adam Chippindale[1], Tomas Babak[1]

1. Queen's University, Kingston, ONTARIO, Canada

Identifying the developmental and molecular architecture of adaptation is a key goal of evolutionary geneticists. In this study, we focus on transcriptome differences in two selection treatments in *Drosophila melanogaster,* “Fast” and “Slow”, which have been selected for extreme developmental timing phenotypes for over 1000 generations with controlled replication. These selection treatments differ in life cycle length, with “Fast” flies having a shortened life-cycle of 9 days and “Slow” flies having an extended life-cycle of 28 days, from egg laying. We assessed the duration of developmental transitions by assessing larval and pupal morphology every two hours for 6 days. Remarkably, we observe highly concordant phenotypic adaptations among independently evolved populations, suggesting the existence of a limited set of developmental pathways involved, consistent with convergent molecular evolution. The developmental stage most impacted by the selection treatment is transition to 3rd instar larvae, which shows an average reduction of ~20 hours (~35%). To determine the molecular basis for these phenotypes we performed RNA-Seq on diverse stages of both Fast and Slow development. Enrichment for body morphogenesis-related pathways and strong conservation of gene expression changes among independently evolved biological replicates corroborated our observations. Together these findings suggest that divergent selection produces differences in gene regulation that are not necessarily mirrored by differences in the selected phenotype. We are further dissecting the contributions of cis and trans gene expression adaptations by mapping allele specific expression in Slow/Fast hybrids.

11

The transcriptional and developmental landscape of male and female mosquitoes

Luca Ferretti[1], Paolo Ribeca[1], Jaroslaw Krzywinski[1]

1. The Pirbright Institute, Pirbright, United Kingdom

Mosquitoes are vectors of devastating human diseases, including malaria, dengue, yellow fever, and Zika. *Anopheles gambiae* is the most efficient vector of malaria and a primary focus of wide-scale control programs based on use of insecticides. Rapid evolution and spread of insecticide resistance threaten the effectiveness of the current programs and call for novel more sustainable approaches, such as genetic control. Sex-specific molecular interactions occurring during mosquito development may offer a plethora of molecular targets for genetic control. However, the corresponding molecular circuits and their elements are very poorly known. Here we present a detailed high-throughput study of the sexed developmental transcriptome of A.gambiae. Comparisons between the male and female expression levels revealed several distinct sex-specific developmental processes. We found that in males, genes on the single X chromosome are significantly upregulated, consistent with operational dosage compensation in larvae and pupae. Differentially expressed genes across male and female development are related to sex-specific functions. Clustering analysis, followed by experimental validation, revealed that, among temporal developmental profiles, the most distinct corresponds to spermatogenesis, enabling prediction of function for a large number of yet unannotated genes related to sexual development. We

have characterized splicing patterns across development, with special focus on sex-specific splicing. We also identified candidate genes involved in sex determination and sex-dependent regulatory interactions. Finally, we have looked at the evolution of expression profiles across developmental stages in *A. gambiae* and other insects. Our analysis sheds new light on the transcriptional and developmental landscape of A. gambiae and lays the foundation for research into the components of sex determination, dosage compensation machinery, and sexual development in this important vector species.

12

Natural variation in expression and function at a *Drosophila melanogaster* odorant receptor locus alters olfactory responses

Coral G Warr[1], Katherine H Shaw[1], Marien de Bruyne[1]

1. School of Biological Sciences, Monash University, Clayton, VIC, Australia

In *Drosophila melanogaster* the responses of many olfactory neurons to volatile chemicals are determined by the odorant receptor (Or) protein family. *Or22a* and *Or22b* are two tandemly arranged members that have resulted from a relatively young gene duplication. In the Canton-S laboratory strain both genes are present, but only *Or22a* is functional and determines the response of the ab3A class of olfactory neurons[1]. A copy number variant at this locus, *Or22ab*, is present at different frequencies in natural populations and at latitudinally varying frequency in Australia, suggesting the locus-specific action of selection[2,3].

To investigate whether the *Or22ab* allele alters olfactory responses we generated individual isochromosomal lines from Australian natural populations and recorded responses from ab3A neurons. We found some lines had the same response profile as the Canton-S strain, and this correlated with two gene copies. However, in other lines we found a second phenotype with altered responses to many odorants, and this correlated with the *Or22ab* allele. Crucially, by expressing the *Or22ab* allele in empty neurons we showed that *Or22ab* mediates this phenotypic change in response profile. We further showed that behavioural responses to ecologically relevant odorants depend on whether *Or22ab* or *Or22a* is present, suggesting that changes to olfactory-driven behaviour may underly natural selection for the *Or22ab* allele.

We also identified a third naturally occurring ab3A phenotype, in which there are still two gene copies, but only *Or22b* is transcribed. We have thus identified at least three different molecular genetic changes at this locus that cause changes in the response profile of the ab3A neuron: (1) two genes are present and expressed but one encodes a non functional protein; (2) only one gene is present and expressed; and (3) two genes are present but only one is expressed.

1. Dobritsa et al., (2003). Neuron 37, 827-841.
2. Aguade (2009). Mol Biol Evol. 26, 61-70.
3. Turner et al., (2008). Genetics 179, 455-473.

13

Title: The molecular language of peptide communication in the marine sponge

Michael Hammond[1], Scott Cummins[1], Tianfang Wang[1], Bernie Degnan[2]

1. University of the Sunshine Coast, Maroochy River, QLD, Australia

2. University of Queensland, Brisbane, QLD, Australia

The development of multicellular organisms from unicellular ancestors constitutes one of the major evolutionary transitions. At the molecular level, a key to organising and coordinating cells is the release of signalling molecules from one cell to stimulate another. From the study of peptide cell communication in basal metazoans, we can gain a better understanding of the fundamental signalling requirements that all animals need to function, regardless of complexity. Our focus was to initially define the peptide secretome of the marine sponge *Amphimedon queenslandica*. Secretory vesicles were isolated from adults and their protein contents analysed through mass spectrometry. Amongst those peptides identified included those with similarity to vascular endothelial factor and beta thymosin. These peptides have also been identified from secretome analysis of sponge adults and larvae. Secreted ependymin-like peptides within sponge cell aggregations were further investigated through production of a recombinant in a bacterial expression system; this is currently being analysed for structure determination which should provide some insight into its precise function, possibly in chemotaxis, cell proliferation or adhesion. The presence of lysine-arginine motifs in numerous precursor proteins, commonly used as cleavage sites within other metazoans, raised the possibility that *A. queenslandica* may have a similar secretory protein machinery. We identified 3 signal peptidase complexes, 5 prohormone convertases and 4 carboxypeptidases that are known to process and export secreted proteins and peptides. Some of these match closely with higher metazoan enzymes, while others are novel. The information gained from this research is expected to expand our knowledge in the area of cell communication and how multicellular animals established peptide communication.

14

Early stages of diversification in the Rab GTPase gene family revealed by genomic and functional studies in *Paramecium* species

Lydia Bright[1], Jean-Francois Gout[1], Michael Lynch[1]

1. Indiana University, Bloomington, IN, United States

New gene functions arise within existing gene families as a result of gene duplication and subsequent diversification. However, the evolutionary forces that act on gene duplicates (paralogs) over time, as well as the intermediate steps on the path to functional change, are not well understood. In an effort to understand functional diversification in paralogs on both the genomic and subcellular level, we tracked duplicate retention patterns and subcellular markers of functional diversification in the Rab GTPase gene family in *Paramecium aurelia* species. Rab GTPases are on the whole more highly expressed and more highly retained than other genes in *Paramecium* genomes after whole genome duplication (WGD). Additionally, consistent with early steps in functional diversification, expression levels of these recent Rab paralogs appear to be diverging more rapidly from one another than other genes in the genome. We uncovered evidence of diversification at the subcellular level by localizing GFP-tagged paralogs from the Rab11 subfamily. Because the functionally diversifying paralogs are closely related to and derived from a clade of functionally conserved paralogs, we were able to pinpoint two specific amino acid residues that may be the drivers of the change in localization and, thus, function.

Interestingly, the functionally conserved proteins label compartments involved in both endocytic recycling, the conserved Rab11 function, and phagocytic recycling, revealing evolutionary links between the two pathways.

15

Transcriptional activation of Wnt and Tgfb signaling pathways in regeneration of sponges

Marcin Adamski[1, 2], Mary Laplante[2], Ilja Borisenko[2], Jon Bråte[3], Alexander Ereskovsky[4], Maja Adamska[1, 2]

1. Australian National University, Acton, ACT, Australia

2. Sars Centre, University of Bergen, Bergen, Norway

3. University of Oslo, Oslo, Norway

4. Mediterranean Institute of Biodiversity and Ecology, CNRS, Marseille, France

Regenerative capacity of sponges, allowing them to recover complete bodies from fragments of tissue or even dissociated cells, is truly remarkable, and not many animals can match this potential. While continuously growing amount of information is available on genes involved in regeneration of model species, including cnidarians, acoels, planarians and chordates, nothing is known about molecular aspects of sponge regeneration.

Using a combination of gene candidate approach and unbiased detection of differentially expressed genes, we have investigated molecular mechanisms of regeneration in sponges. We were particularly interested in the earliest steps of this process, the wound healing, and have analysed it in detail in the emerging calcisponge model system, *Sycon ciliatum*. In this species, complete body can be regenerated from narrow transverse sections of the body column within 5 days of dissection.

Within three hours after dissection, over one thousand genes are differentially regulated. While many of these genes encode for novel proteins or appear to be non-coding, components of two key metazoan developmental pathways, Wnt and TGF-beta, are among the significantly and dramatically upregulated transcripts. In particular, eight (out of 22 present in Sycon) TGF-beta ligands and six (out of 21) Wnt ligands, as well as effectors (Smads, Tcf) and modulators (SFRPs) of these pathways are differentially expressed either throughout the regeneration process or concomitantly with specific morphogenetic events. In situ hybridization of selected genes allowed us to localize the transcripts to specific cells involved in the regeneration.

We have previously shown that in intact sponges many of Wnt and Tgf-beta pathway components are specifically expressed or enriched in the apical (oscular) region, supporting homology between the osculum and the head organizer of cnidarians. Thus, in both sponges and eumetazoans, “organizer genes” are also involved in regeneration, suggesting a common genetic basis of regeneration processes across the animal kingdom.

16

Animal mtDNA as we don’t know it: unusual mt-genomes in non-bilaterian lineages

Dennis V. Lavrov[1], Walker Pett[1]

1. Iowa State University, Ames, IA, United States

Animal mitochondrial DNA (mtDNA) is commonly described as a small, circular molecule that is uniform in size, gene content, and organization. However, data collected in the last decade have challenged this view by revealing substantial diversity in animal mitochondrial genomes. Much of this diversity is found in non-bilaterian animals (phyla Cnidaria, Ctenophora, Placozoa, and Porifera), which, from a phylogenetic perspective, form the major branches of the animal tree along with Bilateria. Within these phyla, mtDNA shows substantial variation in size, organization, genetic code, gene content, presence/absence of introns, tRNA structures, mRNA processing, and rates of evolution. This newly discovered diversity allows a better understanding of the evolutionary plasticity and conservation of animal mtDNA and provides insights into the molecular and evolutionary mechanisms shaping mitochondrial genomes.

17

Mitochondrial variation and the evolution of the germline

Andrew Pomiankowski[1], Nick Lane[1], Arunas Radzvilavicius[1], Zena Hadjivasiliou[1]

1. University College London, London, LONDON, United Kingdom

A dedicated germline has been interpreted as necessary for somatic specialization or a device to suppress selfish conflict in multicellular organisms. These views are difficult to reconcile with the lack of a germline in plants and basal metazoans. Here we show that the need to control deleterious mutations amongst mitochondria could have driven the evolution of the germline. Unlike nuclear genes, which are clonally transmitted through mitosis, the mitochondrial population doubles and segregates at each cell division. In organisms with low mitochondrial mutation rates, random segregation over multiple cell divisions generates sufficient variation for selection to control mutation load. This explains the retention of somatic gametogenesis in many simpler organisms. But with higher mitochondrial mutation rates, selection favours the transition to early sequestration of non-dividing germ cells. We predict this occurred in active bilaterians and promoted the emergence of oogamy, atresia as well as complex developmental processes.

18

Mito-nuclear incompatibilities at early stages of species formation

Ricardo J Pereira[1], Thiago G Lima[2], Ronald S Burton[2]

1. Natural History Museum of Denmark, Copenhagen, NA, Denmark

2. Marine Biology Research Division, Scripps Institution of Oceanography, University of California San Diego, La Jolla, CA, USA

The mitochondria is essential for cell physiology, having a direct impact in organismal fitness. Because mitochondrial function requires a close interaction with nuclear encoded proteins, the fast evolving mitochondrial genes are likely to elicit coevolution at nucelar genes. Hybridization can disrupt such coadaptation, resulting in mito-nuclear incompatibilities and hybrid breakdown at early stages of population divergence. In the copepod *Tigriopus californicus*, extreme rates of mitochondrial evolution have resulted in hybrid breakdown between multiple populations. Yet, it is unclear how many genomic regions are involved in mito-nuclear incompatibilities, and whether those genes are the same in independently evolving populations. To answer these questions, we generate replicated hybrid swarms characterized by similar nuclear composition but alternative

mitochondrial backgrounds, and evolve these populations over 11 generations. Life history traits showed consistent fitness recovery to, or above, parental level, suggesting Darwinian evolution against mito-nuclear incompatibilities in both mitochondrial backgrounds. Whole genome re-sequencing has shown consistent allelic frequency changes across replicates evolving under the same mitochondrial background, identifying genomic regions likely affected by incompatibilities. Although some of these genomic regions are the same in alternative mitochondrial backgrounds, other regions only change in one background. Together, these results show that although mito-nuclear incompatibilities have a relatively simple genetic architecture, they are often asymmetric, as predicted by theoretical work.

19

The Doubly Uniparental Inheritance of Mitochondria: a useful model for mitochondrial biology and evolution.

Fabrizio Ghiselli[1], Liliana Milani[1], Marco Passamonti[1]

1. University of Bologna, Bologna, Italy

Mitochondria are a fundamental component of eukaryotic life, representing much more than just 'the power plants of the cell'. Nevertheless, the current knowledge about their biology (e.g. heredity, biogenesis) and function is largely incomplete, and mostly biased toward a few taxonomic groups. Focusing the research on a small and homogeneous subset of organisms entails the risk of losing a big part of the molecular and functional diversity of mitochondria.

Our work is focused on the Doubly Uniparental Inheritance (DUI) of mitochondria, observed, so far, only in several bivalve species. DUI organisms have two mitochondrial lineages, one transmitted through eggs (F-type), the other through sperm (M-type), whose mtDNAs show up to 50% of amino acid divergence. In DUI species, after amphimixis, the embryo is heteroplasmic for its mtDNA, a status that is eventually maintained only in males, where F-mtDNA is localized in somatic tissues, while M-mtDNA is localized in both germ line and soma. Conversely, in females M-mtDNA is degraded (or diluted below detection limits), restoring the homoplasmic condition. There is both molecular and phylogenetic evidence that DUI evolved as a modification of maternal inheritance, thus our research aims at expanding the knowledge about DUI to make it a model system for mitochondrial biology. Thanks to its unusual features, DUI can shed light on mitochondrial inheritance and biogenesis, and on the relationship between mitochondria and germ line components. Moreover, DUI represents a unique experimental system for studying mitochondrial heteroplasmy, and two processes that shape genome evolution: genomic conflicts and mito-nuclear coevolution. DUI males are naturally heteroplasmic, therefore the biological functions of mitochondria and their interactions with the nucleus are the unaltered result of evolution. All of this makes DUI a novel and extremely useful model to study mitochondrial biology and evolution.

20

Unexpectedly high and variable tempo of plastid DNA integration in mitochondrial genomes of angiosperms

Logan W. Cole[1], Jeffrey P. Mower[2], Jeffrey D. Palmer[1]

1. Department of Biology, Indiana University, Bloomington, Indiana, USA

2. Department of Agronomy and Horticulture, University of Nebraska - Lincoln, Lincoln, Nebraska, USA

All sequenced angiosperm mitochondrial genomes contain plastid-derived DNA, usually between 5 kb and 25 kb of these foreign sequences and sometimes over 100 kb. However, little is known about the tempo and pattern of gain and loss of plastid DNA in angiosperm mitochondrial genomes. To address this issue, we devised a computational approach to estimate rates and sizes of plastid insertions and deletions in a phylogenetic context and applied it to several groups of closely related mitochondrial genomes. Of particular interest are seven mitochondrial genomes in *Monsonia*(Geraniaceae). We estimate that across this group, some 400 plastid-DNA insertion events have occurred, many of them within the past million years. Rates of plastid DNA insertion and deletion each vary significantly over time in this group and independently of one another. Overall, the rate of gain of new nucleotides per site that can be attributed to plastid DNA insertion is more than an order of magnitude higher than the synonymous substitution rate in *Monsonia* mitochondrial DNA. Analysis of mitochondrial genomes in other angiosperm groups suggests similarly high rates of ongoing plastid DNA insertion. These findings decidedly reject the published claim that most plastid-derived sequences in angiosperm mitochondrial genomes are the result of relatively few and ancient transfer events.

21

A mitonuclear 'supergene' explains mitonuclear discordance and nuclear gene flow between two climate-associated forms of a bird species

Hernan E Morales[1], Sasha Pavlova[1], Kaspar Delhey[1], Leo Joseph[2], Paul Sunnucks[1]

1. Monash University, Clayton, VICTORIA, Australia

2. Australian National Wildlife Collection, National Facilities and Collections, CSIRO, Canberra, ACT, Australia

The selective neutrality of mitochondrial DNA variation is increasingly refuted based on data from natural populations and model organisms. Instead, mitochondrial and associated nuclear (mitonuclear) variation emerge as drivers of adaptive evolution and divergence. Species with discordant mitonuclear geographic patterns inexplicable by commonly advocated selectively-neutral causes are powerful systems for studying mitonuclear fitness, adaptation and speciation.

The Eastern Yellow Robin is a common, widespread Australian bird. Populations carrying mtDNA of either an inland or a coastal haplogroup occupy different climates, show interspecific-level mitochondrial DNA difference, yet exchange many neutral nuclear genes. To examine mitonuclear evolution in this species, we sequenced whole mitogenomes, ~1000 nuclear sequence markers, and screened >60,000 genomewide, short-read sequence markers. We tested for selection, derived coalescent estimates of divergence times, gene flow and effective population sizes, conducted genome-scale analyses of population differentiation and linkage disequilibrium, and mapped the short-read markers to a reference genome to ascribe genome position and putative gene functions.

Reconstruction of population history indicated allopatric separation ~2,000,000 years ago, then secondary contact ~100,000 years ago. Current mitonuclear discordance can be explained by two separate adaptive introgressions of mitochondrial DNA. Inland and coastal populations have generally similar nuclear genomes but substantial difference in mitochondrial DNA and a subset of nuclear genes. Remarkably, most of these nuclear genes map to two main locations in the reference genome: a 16 megabase region of autosome 1A, and a smaller region of the Z sex chromosome. The inferred genes in 1A disproportionately have mitonuclear functions and experience very low recombination. Thus 1A acts as a 'supergene' of co-inherited, functionally-related genes.

We hypothesize that variants of the 1A mitonuclear supergene are co-adapted to their local environments, possibly via metabolic adaptation. The supergene genomic architecture can explain how strong mitonuclear adaptation could survive substantial gene flow.

22

Evolution of gene expression: from mutation to polymorphism to divergence

Patricia J Wittkopp[1]

1. University of Michigan, Ann Arbor, MICHIGAN, United States

Gene expression is controlled by a complex network of molecular interactions. Genetic changes that alter this network contribute to phenotypic differences within and between species. To better understand how mutation and selection affect the evolution of gene expression, we have investigated properties of new mutations that create regulatory variation in Saccharomyces cerevisiae and compared their effects to those of regulatory variants segregating in the wild. Comparisons between these two datasets are providing are being used to make inferences about how regulatory networks evolve. Genetic variation segregating within a species reflects the combined activities of mutation, selection, and genetic drift. In the absence of selection, polymorphisms are expected to be a random subset of new mutations; thus, comparing the effects of polymorphisms and new mutations provides a test for selection. When evidence of selection exists, such comparisons can identify properties of mutations that are most likely to persist in natural populations. We have been investigating how mutation and selection contribute to variation in cis- and trans-regulatory sequences controlling gene expression by empirically determining the effects of new mutations and polymorphisms in Saccharomyces cerevisiae

23

Detection of human adaptation during the past 2,000 years

Jonathan Pritchard[1]

1. Stanford University, CA, United States

Detection of recent natural selection is a challenging problem in population genetics, as standard methods generally integrate over long timescales. In my talk I will describe a new statistic, the Singleton Density Score (SDS), a powerful measure to infer very recent changes in allele frequencies from contemporary genome sequences. When applied to data from the UK10K Project, SDS reflects allele frequency changes in the ancestors of modern Britons during the past 2,000 years. We see strong signals of selection at lactase and HLA, and in favor of blond hair and blue eyes. Turning to signals of polygenic adaptation we find, remarkably, that recent selection for increased height has driven allele frequency shifts across most of the genome. Moreover, we report suggestive new evidence for polygenic shifts affecting many other complex traits. Our results suggest that polygenic adaptation has played a pervasive role in shaping genotypic and phenotypic variation in modern humans. This project was led by Yair Field, Evan Boyle and Natalie Telis.

24

Yeast populations adapt at different rates, with a different DFE under different nutrient limitations

Jamie Blundell[1, 2]**, Katja Schwartz**[1]**, Daniel Fisher**[1]**, Gavin Sherlock**[1]**, Sasha Levy**[2]

1. Stanford University, Stanford, CA, United States

2. The Laufer Center for Physical and Quantitative Biology and the Department of Biochemistry & Cell Biology, Stony Brook University, Stony Brook, NY, USA

We have previously developed a DNA barcoding system, that allows us to perform lineage tracking of half a million yeast lineages as they undergo adaptive evolution (Levy et al., 2015). Using this system, we were able to determine that on the order of 20,000 lineages carrying beneficial mutations are able to establish within the first 200 generations of an experimental evolution performed by serial transfer in limiting glucose. We have whole genome sequenced several hundred of these beneficial lineages, and find that the Ras/PKA and Tor pathways are the major pathways for adaptation, and that autodiploidization confers a significant fitness effect (see Y. Li abstract). In this work, we have performed a new set of experimental evolutions, but this time in the presence of limiting nitrogen. We find that the rate of adaptation, as measured by population mean fitness, is much slower than in limiting glucose, and that the DFE consequently is made up of mutations of smaller effect, suggesting that laboratory yeast maybe better adapted to this condition than to limiting glucose. Furthermore, we find that the DFE is to some extent predictive of the population dynamics. We have also sequenced several hundred beneficial lineages, and find that the major targets of adaption in this growth condition are largely non-overlapping, despite both conditions involving serial batch growth with glucose as the carbon source.

25

Functional characterization of evolutionarily advantageous sequence divergence between MHC alleles

Federica Pierini[1]**, Tobias L. Lenz**[1]

1. Department of Evolutionary Ecology, Max Planck Institute for Evolutionary Biology, Ploen, Germany

The highly polymorphic genes of the major histocompatibility complex (MHC) encode for cell-surface glycoproteins with a key role in adaptive immunity. Divergent allele advantage, a mechanism of balancing selection, has been proposed to maintain the exceptional sequence divergence and ancient allelic lineages at these genes. Heterozygous individuals with more divergent MHC allele combinations (i.e. larger sequence difference along the antigen-binding domains) are thought to carry glycoproteins presenting a wider array of antigens to immune effector cells, conferring an advantage against pathogen infections. As quantification of the MHC-bound antigen repertoire is unfeasible in natural populations, different measures of MHC sequence divergence are commonly used as a proxy. However, the direct correlation between sequence divergence and the corresponding repertoire of bound peptides has not been studied systematically across the different MHC genes. Here, we investigated the relationship between sequence divergence and peptide binding properties for the five key classical human MHC genes (human leukocyte antigen; HLA): HLA-A, -B, -C, -DRB1, and -DQB1. Pairwise sequence divergence was correlated with allele-specific binding properties obtained by established computational HLA binding prediction of 115,752 pathogen-derived peptides. For all five HLA genes, the genetic distance between two alleles of a heterozygous genotype showed a significant positive correlation with the combined number of bound peptides. In accordance with the major biological function of MHC class I and class II molecules, we observed for HLA-B and HLA-DQB1 alleles particularly strong correlations for peptides derived from intracellular and extracellular pathogens, respectively. Finally, we observed significant correlations between an allele's population frequency and its average pairwise sequence divergence for four of the investigated HLA genes, suggesting still ongoing selection for divergent HLA genotypes in modern human populations. Overall, our results support the divergent allele advantage hypothesis as a meaningful scenario, contributing to the exceptional genetic diversity in classical MHC genes.

26

Cooperation between distinct viral variants promotes growth of H3N2 influenza in cell culture

Katherine S Xue[1, 2], **Kathryn A Hooper**[2, 3], **Anja R Ollodart**[2, 3], **Adam S Dingens**[2, 3], **Jesse D Bloom**[1, 2]

1. Department of Genome Sciences, University of Washington, Seattle, WA, United States

2. Division of Basic Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA, USA

3. Molecular and Cellular Biology Program, University of Washington, Seattle, WA, USA

RNA viruses rapidly diversify into quasispecies of related genotypes. This genetic diversity has long been known to facilitate adaptation, but recent studies have suggested that cooperation between variants might also increase population fitness. We demonstrate strong cooperation between two H3N2 influenza variants that differ by a single mutation at residue 151 in neuraminidase, which normally mediates viral exit from host cells. Residue 151 is often annotated as an ambiguous amino acid in sequences in Genbank, indicating mixed viral populations. We show that mixed populations grow better than either variant alone in cell culture. Pure populations of either variant generate the other through mutation and then stably maintain a mix of the two genotypes. We suggest that cooperation arises because mixed populations combine one variant's proficiency at cell entry with the other's proficiency at cell exit. Our work demonstrates a specific cooperative interaction between defined variants in a viral quasispecies.

27

Causes and consequences of recombination rate variation across the *Drosophila melanogaster* genome

Josep Comeron[1]

1. University of Iowa, Iowa City, IA, United States

Recombination rates vary within and between species and this variation is known to be influenced by genetic and epigenetic factors. Moreover, in most eukaryotic species examined so far, crossovers occur non-randomly along chromosomes. This variation is predicted to impact rates of evolution and levels of diversity across genomes. Here, I present an analysis of this dual property of variable recombination rates, investigated in terms of genetic and epigenetic causes and evolutionary consequences applied to the model system *Drosophila melanogaster*. In terms of evolutionary consequences, previous studies showed that background selection (BGS) plays a major role explaining levels of diversity across the genome, and thus BGS predictions are adequate as baseline to infer instances of balancing selection or recent selective sweeps. I now show that differences in recombination landscapes among populations of *D. melanogaster*, and the corresponding population-specific differences in predicted BGS, play a significant role explaining highly-localized population-specific differences in nucleotide diversity without requiring invoking local adaptation and recent selective sweeps. In terms of causes, I report analyses to investigate whether DNA motif distribution across the *D. melanogaster* genome can be used to predict some of the observed variation in crossover rates. This study exposes a combinatorial influence of motif presence able to account for more than 40% of the variance in crossover rates across the whole genome, an unprecedented result in any species. This high predictive power is maintained after removing sub-telomeric and -centromeric regions known to have strongly reduced crossover rates. The study also shows that transcriptional activity during early meiosis and differences in motif use between autosomes and the X chromosome add to the predictive power of the models.

28

Saving the Tasmanian devil from extinction

Katherine Belov[1]

1. University of Sydney, Sydney, NSW, Australia

The Tasmanian devil, Australia's largest remaining marsupial carnivore, faces extinction in the wild due to the emergence of a new infectious disease. Devil Facial Tumour Disease (DFTD) is a contagious cancer that is spread as an allograft during biting. I will discuss how the use of genomics and transcriptomics has helped us to understand the disease, its evolutionary trajectory and played a pivotal role in our quest to save the species from extinction.

29

Disentangling Kimura's 3-parameter model of nucleotide evolution.

Michael Hendy[1]

1. University of Otago, Dunedin, OTAGO, New Zealand

In his three-substitution types model (K3ST) of nucleotide evolution Kimura proposed substitutions at a site could be classified into transitions, and transversion of two types, with substitutions within the classes occurring at independent rates α;β and γ. Over time period t, the expected numbers of the substitutions are $q_{\alpha=}\alpha t$; $q_{\beta=}\beta t$ and $q_{\gamma=}\gamma t$ respectively. We have shown that time-dependency can be removed by specifying the expected numbers of substitutions of each type $q_{\alpha}(e)$,$q_{\beta}(e)$ and $q_{\gamma}(e)$ across each edge e of a phylogenetic tree T, as the parameters for the model. The probabilities of each possible pattern of nucleotides observed at the tips of T at that site can then be derived by 4-state Hadamard conjugation from these parameters. The invertibility of Hadamard conjugation means that T and all of its edge-length parameters can be easily derived from these probabilities which can be estimated from the frequencies of the nucleotide patterns in an alignment of homologous sequences.

Using the nucleotide pairings of R/Y (puRine, pYrimidine), W/S (Weak, Strong) or M/K (aMino, Ketone), a 4-state sequence can be projected to a 2-state sequences in three ways. In this presentation I will show that a further application of a Hadamard matrix allows us to parameterise the K3ST model by the expected numbers of substitutions of these 2-state sequences. Although the parameters are independent, the 2-state sequences evolved on the same tree T, giving the opportunity for 3 independent tree reconstructions (using your preferred tree-builder) from your sequence data, as a test of accuracy of the method.

30

GST: A mixture model for phylogenetic inference of heterogeneously evolved sequence data

Stephen M Crotty[1], **Bui Q Minh**[2], **Barbara R Holland**[3], **Lars S Jermiin**[4], **Jonathan Tuke**[1], **Nigel G Bean**[1], **Arndt von Haeseler**[2]

1. *University of Adelaide, Torrensville, SA, Australia*
2. *Centre for Integrative Bioinformatics Vienna, Vienna, Austria*
3. *University of Tasmania, Hobart, Tasmania, Australia*
4. *CSIRO Ecosystem Sciences, Canberra, ACT, Australia*

Heterogeneous evolutionary processes have cast a shadow over the reliability of phylogenetic inference for as long as it has been attempted. These processes bring with them the inevitable consequence of model misspecification, which one would obviously like to minimise. Much work has been done in this area and mixture models that account for rate heterogeneity amongst sites have been in widespread use for some time. These models however are too restrictive to truly represent heterotachous evolution. At the cost of complexity, we introduce a more general mixture model capable of recovering tree and model parameters from datasets generated under heterotachous conditions. We then apply our model to a real dataset, where it demonstrates evidence of convergent evolution in a sodium channel gene of two geographically distinct lineages of teleosts.

31

Incipient speciation of the grey shrike-thrush, *Colluricincla harmonica*, revealed by multi-locus phylogeography

Annika Lamb[1], **Paul Sunnucks**[1], **Leo Joseph**[2], **Sasha Pavlova**[1]

1. *Monash University, Melbourne, Clayton, Victoria, Australia*
2. *Australian National Wildlife Collection, National Facilities and Collections, CSIRO, Canberra, ACT, Australia*

Australia harbours a rich and unique array of avian species that vary in phenotype across their geographic ranges. By analysing species genes it is possible to identify the evolutionary processes that have shaped this diversity. The five currently-recognised subspecies of the grey shrike-thrush (*Colluricincla harmonica*) have been classified based on phenotypic distinctiveness. Whether these subspecies represent independently evolving units or just the product of selection on a few phenotypic traits is unclear. Here, multi-locus phylogeography was employed to address this knowledge gap and to identify the processes that influenced species evolution. This was achieved by (1) assessing the genetic distinctness of each subspecies (2) testing the genetic diversity within subspecies and the degree of differentiation between them (3) estimating divergence times among subspecies and (4) estimating gene flow among subspecies. The subspecies were found to be genetically distinct and to experience low rates of migration among them. The phylogenetic breaks and approximate divergence times among subspecies are concordant with the impacts of putative biogeographical barriers that formed during the Pleistocene. These findings indicate that the subspecies have unique evolutionary histories that have likely been influenced by Pleistocene environmental change. Furthermore, *C. harmonica* lineages are shown here to be distinct, suggesting that they are independently evolving units that may represent incipient species. A genome-wide analysis of genetic diversity would provide further insight into the historical processes and selective forces that have shaped the evolution of *C. harmonica*.

32

Resolving the Phylogenetic Position of Coelacanth: The Closest Relative Is Not Always the Most Appropriate Outgroup

Naoko Takezaki[1], **Hidenori Nishihara**[2]

1. *Kagawa University, Kitagun, KAGAWA, Japan*
2. *Graduate School of Bioscience and Biotechnology, Tokyo Institute of Technology, Yokohama, Kanagawa, Japan*

Determining the phylogenetic relationship of two extant lineages of lobe-finned fish, coelacanths and lungfishes, and tetrapods is important for understanding the origin of tetrapods. We analyzed datasets from two previous studies along with a newly collected dataset, each of which had varying numbers of species and genes and varying extent of missing sites. We found that in all the datasets the sister relationship of lungfish and tetrapods was constructed with the use of cartilaginous fish as the outgroup with a high degree of statistical support. In contrast, when ray-finned fish were used as the outgroup, which is taxonomically an immediate outgroup of lobe-finned fish and tetrapods, the sister relationship of coelacanth and tetrapods was supported most strongly, although the statistical support was weaker. Even though it is generally accepted that the closest relative is an appropriate outgroup, our analysis suggested that the large divergence of the ray-finned fish as indicated by their long branch lengths and different amino acid frequencies made them less suitable as an outgroup than cartilaginous fish.

33

The evolution of bioluminescence and light detection in deep-sea decapods

Heather Bracken-Grissom[1], **Jorge Pérez-Moreno**[1], **Hernán Vázquez Miranda**[1]

1. *Florida International University - Biscayne Bay Campus, North Miami, FLORIDA, United States*

Many organisms rely on bioluminescence for communication, feeding, and defense, especially in the deep sea where downwelling light is limited. This research combines phylogenetic and transcriptomic studies to test several hypotheses addressing the evolution of bioluminescence and light detection in a remarkable family of deep-sea shrimp. All shrimp within the family Oplophoridae use a luminescent secretion discharged from the mouth to deter predators, while only some possess a second mechanism of bioluminescence in the form of cuticular photophores. Photophores are light-emitting organs found across the body that are thought to function in counterillumination. These different mechanisms of bioluminescence emit light at slightly different wavelengths and spectral bandwidths. Past studies have shown shrimp with both the secretion and photophores possess unique visual systems to possibly distinguish between different types of emitted bioluminescence, however genomic approaches have never been applied to investigate this system. In addition, new preliminary evidence suggests that the photophores contain photopigment proteins (opsins) that allow for light detection. This is the first indication that autogenic light organs (those in which the animal itself makes the luciferases and/or luciferins for luminescence) may also have light detection capabilities. Here, a phylogenetic approach is used to investigate the evolutionary origins of the two bioluminescence modes (secretion and photophore) within oplophorid shrimp. Secondly, this project will characterize the visual sensitivity system in the eyes of deep-sea shrimp to better understand how shrimp distinguish between different wavelengths of emitted bioluminescence. Lastly, findings will be presented that investigate the photosensitivity in a non-bacterial (autogenic) light organ – the photophore.

34

The Impact of Migratory Flyways on the Spread of Avian Influenza Virus in North America

Mathieu Fourment[1], **Aaron Darling**[1], **Edward Holmes**[2]

1. University of Technology Sydney, Sydney, NSW, Australia

2. University of Sydney, Sydney, NSW, Australia

Wild birds are the major reservoir hosts for influenza A viruses (AIVs) and have been implicated in the emergence of pandemic events in livestock and the human population. Understanding how AIVs spread within and across continents is critical to develop successful strategies to manage and reduce the impact of influenza outbreaks. In North America, many bird species undergo seasonal movements along the North-South axis, hence fostering opportunities for viruses to spread over long distances. To assess the relative contribution of bird migration along flyways, we undertook a large-scale phylogeography analysis of AIVs sampled in the USA and Canada. We developed a genetic algorithm for the maximum likelihood estimation of highly dimensional models. Based on phylogenies reconstructed from nucleotide data sets, our results show that migration rates within flyways of AIVs are significantly higher than between flyways, suggesting that migratory birds are an important driver for the dispersal of avian influenza viruses. These findings provide valuable insight on the maintenance and transmission of AIVs, hence allowing the development of improved surveillance and risk assessment programs.

35

Demographic Inference Using the Large Sample Joint Frequency Spectrum

Hua Chen[1], **Kun Chen**[2]

1. Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing, BEIJING, China

2. Dana Farber Cancer Institute, Harvard Medical School, Boston, Massachusetts, United States

The joint frequency spectrum (JFS) is a fundamental summary for genetic polymorphism of samples from multiple populations, and is useful for population genetic inference of ancient demography. Analytical theory and methodology of the JFS have been developed using coalescent theory, but are only practicable for small samples. When the sample size is large (e.g., n>50), the computation of JAFS becomes numerically intractable. We present accurate approximation for the JFS for large samples from one or multiple populations. The novelties include: (1) the exact formulas of the JFS are in simple form and computationally efficient for large samples without the numerical intractability existing in former studies; (2) arbitrarily time-varying population size can be modeled. The accuracy of the results is demonstrated by comparing to coalescent simulations. The work provides an useful tool for JFS-based demography inference using large-sample genomic sequencing data.

36

Temporary pulses of accelerated mutagenesis in human and great ape population history

Kelley Harris[1]

1. Stanford University, Palo Alto, CA, United States

There is accumulating evidence that the germline mutation rate is not a clocklike constant, but is better regarded as an evolving physiological trait. This rate places a fundamental limit on how quickly populations can evolve, and it is not well understood how quickly this rate is itself altered by evolutionary forces like selection and genetic drift. Here, I look at dynamics of mutation rate change during great ape evolution by summarizing evidence of changes in the mutational spectrum: the relative mutability of different genomic regions as a function of local sequence context. These changes can be detected by looking at the mutational spectrum as a function of allele frequency, a proxy for allele age. I find evidence that several mutation spectrum changes occurred relatively early in primate history and fixed in certain lineages; for example, an increase in the rate of A→T transversions appears to have fixed in the common ancestor of humans and chimpanzees. More surprisingly, I find instances in multiple great ape species where the mutational spectrum appears to have changed and then reverted back to its original state, as though a mutational process acted for a short period of time and then subsided. Such a pulse appears to have affected Europeans between 10,000 and 20,000 years ago, producing a previously observed excess of transitions in the context TCC→TTC. The relative abundance of this mutation is greatest among SNPs that segregate around 1% frequency in Europeans; the mutational spectrum of rarer European variants is closer to that of Asian and African variants. A few other mutation types appear to be minor components of the same pulse, peaking in relative abundance around 1% frequency in Europeans, most interestingly the dinucleotide mutations GC→AA and TC→AA, which are putatively associated with error-prone DNA replication by polymerase ζ.

37

Sperm-mediated epigenetic effects in Zebrafish (*Danio rerio*)

Berrit Kiehl[1], **Simone Immler**[1]

1. Uppsala University, Uppsala, Sweden

A recent study in the zebrafish *Danio rerio* showed that changes in the social environment, and specifically in the intensity of sperm competition, do not only lead to modifications in sperm motility and sperm velocity, the paternally experienced environment of sperm competition also translates into differences in hatching timing and larval survival (Zajitschek et al, 2014). The aim of my PhD project is to assess where in the sperm the trigger for this mechanism lies (RNA vs. chromatin structure vs. methylation processes) and how this is transferred into the offspring. I am investigating the role of sperm in parental effects and aim to identify potential mechanisms causing paternal effects (i.e. sperm-mediated epigenetic mechanisms). This involves two experiments in particular - one to actually isolate RNA from sperm and create a transcriptome for the zebrafish ejaculate, which can then be used as a reference for the entire project. In a second step, I perform experiments to manipulate the paternal effects and look into the underlying molecular mechanisms. I assess the causes for variation in epigenetic effects in males by manipulating for example, sperm competition levels among males and comparing the epigenetic patterns in their sperm for which I use three different approaches, i.e. (A) comparing the RNA profile in sperm (B) investigating chromatin structures in sperm and (C) determining the methylation patterns in sperm. Experimental tools such as *in vitro* fertilisation and experimental manipulation are combined with transcriptomics and epigenomics to address these two questions using a model organism.

1. Zajitschek, S; Hotzy, C; Zajitschek, F; Immler, S (2014): Short-term variation in sperm competition causes sperm-mediated epigenetic effects on early offspring performance in the zebrafish. Proceedings of the Royal Society B-Biological Sciences 281(1785) DOI: 10.1098/rspb.2014.0422.

38

Establishing bipotentiality for gonadal differentiation.

Yisheng Yang[1], Summer Sorensen[1], Megan Wilson[1]

1. Department of Anatomy, University of Otago, Dunedin, Otago, New Zealand

In order for development of the reproductive system to occur, the formation of gonad anlagen is first required. Mammalian gonads arise from a bipotential progenitor gonadal tissue called the urogenital ridge (UGR). However, very little is known about how the molecular networks that shape its formation and the molecular preparations made to allow for two developmental trajectories. The LIM-homeobox gene, *Lhx9* is among only a handful of genes known to be required for UGR formation. In order to investigate the molecular underpinnings involved in UGR formation, we took a large-scale approach using chromatin immunoprecipitation followed by next generation sequencing (ChIP-seq) to identify genes involved in this process.

ChIP-seq on mouse UGRs (E11.5) was performed for Lhx9 and for two histone modification marks, H3K4me3 and H3K27me3, which highlight regions of active and repressive transcriptional states respectively. Lhx9 target genes were confirmed by RT-qPCR using a *Lhx9* knockout line.Genes required for processes such as sex determination, sexual differentiation, cell proliferation, angiogenesis and cell migration were identified as regulated by Lhx9. Several Lhx9 target genes, whose expression patterns in the UGR were not previously characterized, were analysed further by *in situ* hybridization.

Furthermore, looking at both histone mark ChIP-seq datasets, many genes were found to possess a 'bivalent' histone modification dynamic, whereby both H3K4me3 and H3K27me3 were found in the promoter or enhancer regions. This histone dynamic has been characterized as a feature that highlights certain lineage regulatory genes, holding them in a 'poised' transcriptional state. In particular, many genes involved in the *Wnt* signaling pathway were identified to possess bivalent histone marks. Bringing both ChIP-seq datasets together, we provide a wider scope of the transcriptional and epigenetic regulatory network that is necessary for UGR formation, but also the preparation for sexual development.

39

DNA methylation and sex chromosome dosage compensation

Shafagh A Waters[1], Alexandra Livernois[2], Hardip Patel[3], Paul D Waters[1]

1. The University Of New South Wales, Sydney, NSW, Australia

2. University of Canberra, Canberra

3. John Curtin School of Medical Research, Australian National University, Caberra, ACT, Australia

Cytosine methylation is an epigenetic modification that plays a role in regulation of transcription. Methylation, particularly at promoter CpG-islands, can lead to silencing of the associated gene. In mammals, DNA methylation has several well characterized regulatory functions, including the chromosome-wide epigenetic silencing of the X chromosome (called X-chromosome inactivation; XCI). XCI is part of a dosage compensation system in therian (eutherian and marsupial) mammals that results in almost equal average transcriptional output from the X chromosome between the sexes. DNA methylation is a late and stabilizing step in maintaining transcriptional silence of the X in eutherian mammals, but there are limited detailed data about DNA methylation in marsupials, monotreme and birds. Here we present a genome wide analysis DNA methylation in non-eutherian representatives from three amniote vertebrate lineages, each with independently evolved dosage compensation systems.

40

Epigenetic memory in vertebrates

Julian R Peat[1], Michael P Collins[1], Olga Kardialsky[1], Morgan T Jones[1], Timothy A Hore[1]

1. Department of Anatomy, University of Otago, Dunedin, New Zealand

Epigenetics provides a mechanism for cells to 'remember' early developmental decisions in the absence of the signals which first initiated them. Methylation of CG dinucleotides is amongst the most iconic of all epigenetic systems because there is a well defined mechanism by which it transmits molecular memory following cell division. Accumulation of methylation during development is essential for mammalian development, and its removal can help cells to attain the naive pluripotent state that defines the most developmentally potent cells of the body. Currently, the extent to which CG methylation exists in divergent vertebrate species is largely unknown, and the extent to which it can be erased and manipulated is poorly characterised. We report our recent results in this area and discuss its implications for the understanding of epigenetic memory systems.

41

Whole-transcriptome profiling in a model sex-changing fish identifies genes that maintain flexible sexual phenotypes.

Erica V Todd[1], Hui Liu[1], Neil J Gemmell[1], Melissa Slane-Lamm[2], John Godwin[2]

1. University of Otago, Dunedin, OTAGO, New Zealand

2. Department of Biological Sciences, North Carolina State University, Raleigh, NC, USA

Sex is increasingly seen as a continuous, rather than a dichotomous trait. Sex is phenotypically plastic in many marine fishes and results from environmentally-sensitive differential gene regulation. Bluehead wrasse (*Thalassoma bifasciatum*) are highly-social reef fish and well-studied models of sexual plasticity. These diandric, protogynous (female-first) hermaphrodites have three sexual morphs as adults whose development is plastic and socially cued. Bluehead wrasse mature as male (primary males) or female, but each have the capacity to become dominant (secondary) males later in life. Large, brightly coloured secondary males actively defend and court a harem of females, whereas primary males are female-mimics that employ a 'sneaker' mating strategy. Using whole-transcriptome RNA-sequencing (RNA-seq) we have explored the molecular basis of plastic sexual phenotypes in bluehead wrasse brain and gonad. Differential expression analysis identified thousands of genes important in the maintenance of the primary male, secondary male, and female phenotypes. Brain expression profiles of primary males reflect their female-like behaviour, not their male sex. Secondary male brains were most different. We find that isotocin (homologue of mammalian oxytocin) is overexpressed in secondary males, supporting recent evidence for a regulatory role in teleost social interactions, especially those related to dominance and rank. Gonadal expression profiles were strongly sex-biased, although secondary males upregulated genes involved in androgenesis and in the maintenance of secondary sexual characteristics (i.e., colouration and territoriality). Further investigations into the

molecular basis of sexual plasticity are now underway, including transitions between alternative phenotypes and comparisons of gene expression patterns in evolutionarily divergent systems.

42

Aging-associated DNA methylation changes through long term follow-up in African green monkey

Ja-Rang Lee[1], Se-Hee Choe[1, 2], Hyeon-Mu Cho[1, 2], Sang-Je Park[1], Young-Hyun Kim[1, 2], Sang-Rae Lee[1, 2], Kyu-Tae Chang[1, 2], Jae-Won Huh[1, 2]

1. National Primate Research Center, Korea Research Institute of Bioscience and Biotechnology (KRIBB), Cheongju, Chungbuk, Republic of Korea

2. University of Science & Technology (UST), National Primate Research Center, Korea Research Institute of Bioscience and Biotechnology (KRIBB), Cheongju , Chungbuk, Republic of Korea

Epigenetic changes plays an essential role in the phenotypic changes and biological processes associated with aging. DNA methylation is one of important epigenetic control mechanism that has been shown to be aging-related gene silencing. Age-related DNA methylation changes have been called 'epigenetic drift', but the defining features of this phenomenon remains unclear. To further explore and characterize the relationship between DNA methylation level and age, we performed the targeted bisulfite sequencing analysis over the next year in blood samples of 17 years old African green monkey. We observed that within the identical individual, global CpG, CHG methylation level trend to increase over time. In addition, we identified 59 aging-associated differential methylated regions (DMRs) in gene expression regulated or CpG island regions. Aging-associated hypermethylation regions were enriched regarding metabolic or biosynthetic process gene ontology (GO) terms, whereas hypomethylated regions showed no enrichment. According to aging-related gene function in DMRs, we selected 17 candidate genes, and performed the validation experiments through long term follow-up. Here we report that aging-associated DNA methylation changes in African green monkey identical individual through long term follow-up. These finding are invaluable resource to better understanding of epigenetic drift. Furthermore abundance accumulation of global methylation analysis of long-term follow-up in identical individuals might contribute to our understanding about genetic mechanisms of age-related phenotypes and diseases.

43

Deiodinase type 3 methylation increases in response to thyroid stimulating hormone in a fasting adapted mammal

Bridget Martinez[1, 2], Neil Gemmell[1], Daniel E Crocker[3], Rudy M Ortiz[2]

1. Anatomy , University of Otago, Dunedin, OTAGO, New Zealand

2. Molecular and Cellular Biology, University of California, Merced, Merced, California, United States

3. Department of Biology, Sonoma State University, Rohnert Park , California, United States

Epigenetic regulation through the methylation of genomic cytosine nucleotides is a crucial component in the regulation of gene expression. DNA methylation in mammalian genomes serves as an important modulator of when and where a gene should be expressed, by primarily targeting 5'-CpG-3' dinucleotides which have been associated with gene regulatory sites, including promoter sequences. Food deprivation in mammals is associated with increases in reverse triiodothyronine (rT3), an inactive form of the metabolism driving thyroid hormone triiodothyronine (T3), as a direct consequence of increases in the expression of the gene deiodinase type 3 (DI3). However, in prolong-fasted, metabolically active elephant seal pups, cellular thyroid hormone-mediated components are up-regulated with fasting duration, while rT3 levels remain low. To address our hypothesis that methylation status of DI3 contributes to the silencing of a gene that would otherwise be expressed under such physiological conditions, we infused early fasted pups with thyroid stimulating hormone (TSH), extracted and treated genomic DNA with sodium bisulfite in an effort to detect and quantify 5-methylcytosine (5mCs) at single base-pair resolution. In reference to the genomic sequence of DI3 in elephant seals, prior to infusion with TSH, there is 50% methylation of 15/30 CpG sites in adipose tissue. Twenty-four hours post infusion with TSH, the degree of methylation of DI3 increases to 93%, with methylation observed in 27/29 CpG sites. Moreover, in muscle, methylation reaches 90%, with 26/29 CpG sites being methylated at only 60 minutes post-infusion with TSH, suggesting a tissue specific sensitivity. Aside from providing an initial and novel assessment of methylation regulation and status in elephant seals, the data demonstrates a unique adaptation to a methylation pattern where the compacted chromatin structure of DI3 plays a crucial role in suppressing the normal physiological response to fasting otherwise present in most mammals.

44

The regulatory evolution of the human vocal tract

David Gokhman[1], Eran Meshorer[1], Liran Carmel[1]

1. The Hebrew University of Jerusalem, Jerusalem, Israel

It is broadly accepted that changes in gene regulation are a key driver of phenotypic differences between closely related species, and play an important role in short-term adaptations. However, the specific regulatory changes that occurred during the recent evolution of humans are largely unknown. To this end, we have produced and analyzed dozens of DNA methylation maps from chimpanzees and humans, including both modern and archaic human groups. We found ~1,400 differentially methylated regions that emerged on our lineage after the split from Neanderthals and Denisovans. Functional analyses of these genes revealed that our vocal tract has gone through a particularly rapid evolution that is not shared by other archaic human groups. Additionally, we detected substantial regulatory changes in the *NFIX* gene, and show that this gene might have been a major driver of changes in our craniofacial and vocal features. We also show that changes in isoform ratio of the *AUTS2* gene, which are unique to anatomically modern humans, might have played a key role in our cognitive abilities, and in our susceptibility to Autism. Altogether, this study establishes a comprehensive catalogue of regulatory changes in recent human history.

45

Avian genome evolution and the origins of species diversity

David W Burt[1]

1. The Roslin Institute and Royal (Dick) School of Veterinary Studies, University of Edinburgh , Easter Bush Campus, MIDLOTHIAN, United Kingdom

Our knowledge of avian genomes has increased rapidly, starting with the publication of the chicken genome in 2004, a milestone for avian and evolutionary biology. Advances in DNA sequencing now make it possible to produce draft sequences of any vertebrate genome, quickly and cheaply. We have seen the completion of draft genomes of more than 50 other birds, with plans to sequence all 10,000 by the B10K Consortium. With advances in long read sequencing, we are seeing genome assemblies with N50 contigs of more than 3Mb. The annotation of genomes has also been under continuous improvement, taking advantage of transcriptome data generated both by short and long read RNA sequencing technologies. These approaches provide improved gene models, with complex patterns of transcription with multiple RNA isoforms. This has worked well for both coding and non-coding transcripts. Recently, the analysis of 44 bird genomes by the Avian Phylogenomics Consortium opened up new opportunities. For individual species, the sequences coupled with the initial annotations, can serve as a vehicle for basic research. On the other hand generating a multiple genome sequence alignment can enable comparative studies, which benefits all these species. Such studies broaden our understanding of genome evolution and the evolution of traits or can help to disentangle phylogenetic relationships. Our main aim is to analyse the integrated data with a focus on creating a detailed functional map relevant to birds. Such a map can be used to drive the identification of novel protein-coding and non-coding genes, binding sites for transcription factors, enhancers or other functional elements. I will review the current status of avian genome annotation and open up the discussion on future possibilities using phylogenomics in the study of species diversity and the evolution of avian traits.

Supported by BBSRC, EC, Wellcome Trust and University of Edinburgh.

46

Speciation and domestication in the zebra finch, an avian model system for evolutionary genomics

Christopher Balakrishnan[1], Allison Lansverk[1], John Davidson[1], Simon Griffith[2]

1. East Carolina University, Greenville, NORTH CAROLINA, United States

2. Macquarie University, Sydney, NSW, Australia

Zebra finches, native to Australia, have long been an important model system across a diversity of fields. Due especially to their role as a model for the study of vocal learning, the zebra finch was the second bird species with a completely sequenced genome. Zebra finches, however, also have tremendous potential as a model system for evolutionary genomics. Such a model is important in part due to the unique features of birds including their unusually stable karyotype and recombination landscapes, and their apparent slow evolution of postzygotic reproductive isolation. I will highlight our progress in two areas: 1) describing patterns of genomic variation among wild and domesticated zebra finches and 2) improving our understanding of gene regulatory evolution in birds. Despite the role of domesticated zebra finches in research, little is known about how domestication has influenced patterns of morphological and genetic variation. Domesticated populations are of interest because they often differ from wild populations in predictable ways, and domestication itself may influence the behaviors that are under study. To quantify patterns of divergence, we sequenced whole genomes of wild and domesticated zebra finches at medium depth (~8x). As expected, we find significant reductions in genetic diversity in domesticated populations. We also use multiple lines evidence to identify putative regions of the genome that have experienced selection during domestication. In addition to surveying genetic variation, we have also used RNA-seq to characterize brain gene expression divergence between two zebra finch subspecies that have been geographically isolated for over one million years. In contrast to many previously examined systems, we find that gene misexpression, indicative of potential Dobzhansky-Muller incompatibilities, is relatively rare. Like previous studies in other systems, we also find that regulatory divergence has occurred predominantly in *cis* regulation. Thus, these findings highlight both shared and unique features of avian regulatory evolution.

47

Evolution of antimicrobial peptides in birds

Yuanyuan Cheng[1], Michael Dennis Prickett[2], Weronika Gutowska[3], Richard Kuo[3], Katherine Belov[1], David Burt[3]

1. School of Life and Environmental Sciences, University of Sydney, Sydney, Australia

2. University of Trieste, Trieste, Italy

3. The Roslin Institute, University of Edinburgh, Edinburgh, UK

Antimicrobial peptides (AMPs) are a diverse group of molecules with potent, broad-spectrum activities against microbes. AMPs are found in a wide range of invertebrate and vertebrate organisms, representing an ancient form of innate immunity. Defensins and cathelicidins are two major families of AMPs in birds. With a key role in host defense against rapidly evolving pathogens, defensins and cathelicidins provide an ideal system for studying adaptive molecular evolution. Recent release of whole-genome sequences for various bird species enabled us to perform a comprehensive evolutionary analysis of avian AMPs across multiple bird lineages.

We mined 53 avian genomes representing 32 orders and identified 758 AMP genes, including 714 β-defensins and 44 cathelicidins. Both gene families form a generally conserved gene cluster in avian genomes, with certain genes being more prone to duplication or pseudogenisation events. Intense negative selection was detected in most of examined gene domains, likely accounting for the conservation of certain amino acid residues that are essential for the functioning of β-defensins and cathelicidins in birds. Episodic positive selection also played an important role in driving the diversification of certain peptide residues, contributing to high variability of gene sequences and electrostatic property of the peptides. Our results also revealed that selection may have acted on cathelicidins to maintain a balanced charge between the propiece and mature peptide domains, so that the high cationicity of the mature peptide is neutralised by the negative charge of the propiece before peptide secretion.

These findings greatly improved our understanding of the molecular evolution of avian AMPs and will be useful for understanding their role in the avian immune system. Additionally, the large dataset of defensin and cathelicidin peptides also provides a valuable resource for translational research and development of novel antimicrobial agents in the future.

48

Localisation of recombination hotspots at gene promoters and retrotransposons in avian genomes

Takeshi Kawakami[1], Carina F Mugal[1], Alexander Suh[1], Reto Burri[1], Marty Kardos[1], Ludovic Dutoit[1], Hans Ellegren[1]

1. Uppsala University, Uppsala, UPPSALA LäN [SE-03], Sweden

Characterising regional variation in recombination rates across the genome is critical for understanding the forces that shape patterns of variation in genetic diversity and divergence. The Prdm9 plays a key role for determining positions of meiotic crossovers and demarcating the location of 'recombination hotspots' in several mammalian species. However, other vertebrate species without Prdm9, such as birds, also show highly variable recombination rate across the genome, but much less is known about the underlying mechanism.

Here we estimate fine-scale population recombination rate based on the patterns of linkage disequilibrium across the whole genomes of 89 collared flycatchers (*Ficedula albicollis*). We identified 2,669 recombination hotspots with average recombination rate being more than 10 times higher than the genomic background. These hotspots showed a strong association with gene promoter regions, first exons and CpG islands, although a substantial portion of hotspots was located in intergenic regions. Consistent with this pattern, overall recombination rate was higher at gene promoter regions, first exons and CpG islands than the genomic background. This is in line with the 'open chromatin hypothesis', where genomic regions with loosely packed chromatin, such as gene promoters and transcription start sites, are more easily accessible to the recombination machinery. Interestingly, recombination rate was also higher at regions containing LINE and LTR retrotransposons. This may indicate an indirect correlation between retrotransposons and high recombination rate such that highly recombinogenic regions characterised by open chromatin are also targets for the integration of retrotransposons, which also requires an accessible chromatin state. However, CT-rich motifs were significantly enriched in highly recombinogenic LTR retrotransposons, which may indicate a causal relationship between retrotransposons and recombination hotspots. Because similar CT-rich motifs are known to be associated with meiotic crossovers in plants and yeasts, these CT-rich retrotransposons may increase local recombination rate by modifying chromatin marks in bird genomes.

49

Genomic basis of tool manufacture and use in New Caledonian crows

Nicolas Dussex[1], R. Axel W. Wiberg[2], Verena E. Kutschera[3], Gavin Hunt[4], Russell D. Gray[4, 5], Robert Fleischer[6], Christian Rutz[2], Michael G. Ritchie[2], Jochen B. W. Wolf[3], Neil J. Gemmell[1]

1. University of Otago, Dunedin, New Zealand

2. University of St Andrews, St Andrews, UK

3. University of Uppsala, Uppsala, Sweden

4. University of Auckland, Auckland, New Zealand

5. Max Planck Institute for Science of Human History, Jena, Germany

6. Smithsonian Conservation Biology Institute, Washington, DC, USA

The discovery of avian cognitive abilities has revolutionised our understanding of the evolution of intelligence. However, the genetic basis of these abilities is unknown. One possibility is that numerous changes in genes across a wide range of functional domains are required for the evolution of complex intelligence. Alternatively, only a limited number of genetic tweaks might be required. Corvids are well known for their cognitive abilities such as episodic-like memory, problem solving, and tool use. The New Caledonian crow (*Corvus moneduloides*) is particularly intriguing as it is one of the few non-human species to manufacture foraging tools, making it an ideal model to study the genetic basis of cognition. Here we present genome-wide (~18,000 protein coding genes) phylogenetic comparisons among 12 crow species including *C. moneduloides* and scrutinize the genome for signatures of selection. In order to test whether similar genetic changes may have arisen earlier in the evolutionary history of the lineage, we perform the same analysis focusing on the closely-related tropical but non-tool using white-billed crow (*Corvus woodfordi*). We then use transcriptome data from *C. woodfordi*and avian protein databases to link candidate genes under selection to the species biology. We expect to detect signatures of positive selection (i.e. dN/dS >1) in genes associated with brain function and bill morphology allowing tool use and manufacture. Our results will help uncover the evolution and genetic basis of cognition in the wild and will reveal the nature of changes required to evolve cognitive abilities.

50

Avian ecological epigenetics: the role of DNA methylation in the evolution of avian personality

Kees van Oers[1], Martijn FL Derks[2], Kyle M Schachtschneider[2, 3], Ole Madsen[2], Veronika N Laine[1], Koen JF Verhoeven[4]

1. Department of Animal Ecology, Netherlands Institute of Ecology (NIOO-KNAW), Wageningen, Netherlands

2. Animal Breeding and Genomics Centre, Wageningen University, Wageningen, Netherlands

3. Department of Animal Sciences, University of Illinois, Urbana-Champaign, Illinois, USA

4. Department of Terrestrial Ecology, Netherlands Institute of Ecology (NIOO-KNAW), Wageningen, Netherlands

The world is changing rapidly and organisms need to adapt to those changes. Animal personality describes how individuals differ in how well they cope with such challenges and is an important factor for explaining fitness of individuals and viability of populations. The search for hereditary mechanisms underlying these animal personality traits has focussed on the identification of underlying genomic polymorphisms, but especially for natural populations this has not been fruitful. It has become clear that hereditary variation is more than genomic variation alone. Also epigenetic mechanisms, such as DNA methylation, can alter gene expression over multiple generations, while this epigenetic variation may originate from both genetic and environmental factors. This raises the question whether there is an epigenetic basis for hereditary variation in personality traits, and how genetic variation interacts with epigenetic variation?

Here we combine whole-genome sequencing with whole-methylome (bisulfite) sequencing to study the influence of DNA methylation on the variation and evolution of personality traits in both natural populations and lines selected for extreme personalities in the great tit (Parus major). We integrate long-term behavioural data, whole-genome sequence and RNAseq data with whole-methylome, RRBS and candidate gene methylation data. We show that gene expression is a complex interaction between genomic and epigenetic information. The epigenome, as a regulator of the expression of the genome, thereby is an important factor explaining the heritable-but-plastic character of behavioural traits. Moreover, since epigenetic changes are more likely to interact with environmental variation, epigenetic mechanisms may react faster to selection pressusres compared to genetic polymorphisms. This is important for estimating the potential for species to adapt to environmental change such as global warming and urbanization.

51

Modeling the biological cost of cousin alliance on human population genetics

Elsa G Guillot[1], Martin L Hazelton[2], Tatiana M Karafet[3], J Stephen Lansing[4], Herawari Sudoyo[5], Murray P Cox[2]

1. DEE, Université de Lausanne, Lausanne, Switzerland

2. Massey University, Palmerston North, New Zealand

3. University of Arizona, Tucson, USA

4. Complexity Institute, Nanyang Technological University, Singapore

5. Eijkman Institute, Jakarta, Indonesia

Population genetics has opened a new era of population history reconstruction, developing methods to infer past migration, settlement and admixture from DNA. Relying on simple models, these reconstructions always assumed random mating within population. However social-anthropology tells us that humans, modern and ancient, live in societies constrained by non random-mating, made of taboo and preferred alliance. Are these social constraints imposing a biological cost via inbreeding? Can we explain the emergence of particular marriage rules? Should population genetics methods be corrected for non random mating in humans?
We explore the effect of a specific cousin alliance system on human population genetics using a new simulation tool (SMARTPOP). This permits to simulate a vast range of social, genetic and demographic parameters to quantitatively model the effect on genetics. We also investigate the ability of our simulations to infer the social history of small population, using empirical genetic data, in an Approximate Bayesian Computation framework.

52

The genetic basis of female choice in Drosophila

Daniel Matute[1]

1. 250 Bell Tower Road, Chapel Hill, NC, United States

Speciation occurs when two populations split and accumulate genetic differences, which can lead to the evolution of reproductive isolating mechanisms. Species of the *Drosophila* genus remain some of the best models for the study of speciation. In spite of the power of *Drosophila*, little is known about the genetic basis of female mate preference. Previous work identified a pair of recently diverged races within *Drosophila melanogaster* that show strong behavioral isolation but little reproductive isolation of any other type. We leveraged the power of evolutionary genetics in *Drosophila* to dissect the genetic basis of the trait and identified to alleles involved in female choice.

53

Evolving mutation rate advances invasion speed of sexual species

Marleen Cobben[1], **Alexander Kubisch**[2]

1. Animal Ecology, NIOO-KNAW, Wageningen, Netherlands

2. Theoretical Evolutionary Ecology, University of Wuerzburg, Wuerzburg, Germany

Many species are shifting their ranges in response to global climate change. Range expansions have important genetic and evolutionary consequences, caused by non-random mating at the expanding range margins. One of the consequences is the evolution of dispersal, which increases colonisation speed, provided that a species can adapt sufficiently fast to novel local conditions. Mutation rates can evolve too, under conditions that favor an increased rate of adaptation, but the increase in frequency of a modifier increasing the mutation rate is dependent on genetic hitchhiking on a beneficial mutation at a gene under selection. In sexual populations recombination can then quickly erode the generated linkage disequilibrium. Here we use an individual-based model to show that non-random mating at the expanding range margin practically eliminates the possibility of recombination during invasion, allowing the maintenance of the established linkage disequilibrium. This causes the evolutionary increase of mutation rates, which clearly advances the range expansion both through its effect on the evolution of dispersal rate, and the evolution of local adaptation. This occurs both in a scenario of steadily increasing temperatures across the landscape as with variably increasing temperatures. In contrast, in a spatially stable population, strong directional selection causes the evolution of mutation rates as well as shown in previous theoretical studies, but not when we add variance to the mean selection pressure. By this we show that the evolution of mutation rates possibly plays a more important role in the adaptation of sexual species due to its enhancement by non-random mating during range expansions. This has possibly far-reaching consequences concerning species' invasiveness and the rate at which species can adapt to novel environmental conditions during range expansions under global climate change.

54

Disassortative mating maintains inversion polymorphism and changes the rules of speciation in a mimetic butterfly

Mathieu Chouteau[1], **Violaine Llaurens**[2], **Annabel Whibley**[3], **Florence Piron Prunier**[2], **Mathieu Joron**[1]

1. Centre d'Ecologie Fonctionnelle et Evolutive (CEFE), CNRS, Montpellier, France

2. Institut de Systématique, Evolution, et Biodiversité (ISYEB), CNRS, Paris, France

3. Department of Cell and Developmental Biology, John Innes Centre, Norwich, United Kingdom

Mate choice is a key component of the dynamics of adaptive alleles in a population. Emphasis has been given to assortative mating where mating occurs between individuals with similar phenotypes, facilitating local adaptation and speciation. Less is known about disassortative mating (or negative assortative mating) defined by preferential pairings between individuals differing in phenotype. We investigate wing pattern polymorphism displayed by toxic butterflies (*Heliconius*). Their patterns are signals of toxicity learned by local predators, and enhance survival through predator avoidance and mimicry with other local toxic prey. Warning signals are often used as mating cues, driving assortative mating and speciation. However, the Amazonian species *Heliconius numata* maintains a stable polymorphism with multiple colour morphs coexisting within populations, a situation which mimicry alone or assortative mating cannot explain. Morphs are controlled by a single locus (supergene) formed by multiple chromosomal inversions. Using experimental mating trials we found strong disassortative mating between morphs, and showed this was mediated through female choice. The wing pattern supergene showed clear excesses of heterozygotes in natural populations, contrasting with an otherwise freely panmictic genome. This may be expected if chromosomal inversions carry recessive deleterious mutations, as is known for other supergenes, causing heterozygous advantage. Between populations, genomes were atypically undifferentiated across the continental distribution. We modelled the consequences of disassortative mating on polymorphism in spatially structured populations. Disassortative mating produced negative frequency-dependence, which favours rare morphs, including those bringing poor survival benefits (absence of local co-mimics) so long as they are recessive, which fits our empirical observations. Effective gene flow between populations was enhanced through mating benefits to recessive wing-pattern alleles, acting against population differentiation, and enhancing effective population size. Through both local and global effects on the mixing of genomes, disassortative mating at the supergene maintains adaptive polymorphism, and acts against speciation.

55

Inbreeding avoidance and dispersal behaviours in humans from Inner Asia

Nina Marchi[1], Tatiana Hegay[2], Philippe Mennecier[1], Tchoudouraa Mikhaîlovna Dorjou[3], Boris Chichlo[1], Laure Ségurel[1], Evelyne Heyer[1]

1. MNHN, Paris, FRANCE, France

2. Institute of Immunology, Uzbek Academy of Sciences, Tashkent, Uzbekistan

3. State University of Tuva Republic, Kyzyl, Tuva, Russia

Inbreeding is the biological consequence of reproduction between closely related individuals. It results in an increase in the number of homozygous sites within genomes and a decrease in genetic diversity. This can reveal recessive deleterious alleles associated with genetic diseases, decrease fertility and impede the adaptive response of individuals. In humans, two strategies can limit inbreeding. First, individuals can migrate out of their native group and mate inside a new group, which corresponds to geographic exogamy. Second, in the absence of dispersal, individuals can mate within their groups according to specific matrimonial rules.

In Inner Asia, multiple human populations with contrasted social organisations and different levels of geographic exogamy cohabit. This area therefore represents an interesting opportunity to test for the presence of inbreeding avoidance strategies. In this study, we collected both ethnological and genomic data for 369 men and 177 women in 18 populations from Inner Asia (Uzbekistan, Tajikistan, Kyrgyzstan, Siberia and Mongolia). This allowed us to detect the presence of geographical exogamy for each couple and to estimate the genetic inbreeding of each individual.

First, based on genetic estimates, all populations are less inbred than under random mating, suggesting they all have some strategies to avoid inbreeding. Second, we found that the proportion of exogamous couples was highly variable between populations, from 0% to 72%. Furthermore, we found that the endogamous populations are less inbred than the exogamous ones. However, mostly or entirely endogamous populations are organized under a cognatic society while mainly exogamous populations are patrilineal. Social organization (patrilineal or cognatic), correlated to differences in dispersal behaviors, seems to lead to different patterns of genetic inbreeding.

56

Evolution of effective population size and sex ratio under inbreeding

Marcy K Uyenoyama[1], Naoki Takebayashi[2]

1. Duke University, Durham, NORTH CAROLINA, United States

2. Biology and Wildlife, University of Alaska, Fairbanks, Alaska, United States of America

Our Bayesian estimation procedure uses a modified Ewens Sampling Formula to generate posterior distributions of relative effective number and sex ratio among reproductives under gynodioecy, androdioecy, and other regular systems of inbreeding. Within a coalescence framework, those quantities depend upon the proportions of hermaphrodites and gonochores (males or females) and the rate of self-fertilization of hermaphrodites. Estimates obtained from the analysis of genetic data indicate that relative effective number is nearly but not quite maximized in three natural gynodioecious and androdioecious populations. Our evolutionary analysis indicates that the evolution of the sex ratio to the evolutionarily stable strategy (ESS) implies maximization of relative effective number if the viabilities of hermaphrodites and gonochores are identical. Accordingly, the departure of relative effective number from its maximum provides a means of inferring the relative viability of the sexes. We present posterior distribution for the relative viability of males and females relative to hermaphrodites in the populations studied.

57

Evolution through successive population isolation and reconnection events: process and signature

Severine Vuilleumier[1]

1. EPFL - Ecole Polytechnique Fédérale de Lausanne, Lausanne, VAUD, Switzerland

Many aspects of the environment that are critical for the species survival can vary through time, from day to day changes, to those across years or over geological periods. They not only modify the selective pressures acting on populations but also change the size and distribution of the populations themselves, isolating or connecting them. Such changes may result in extinction but alternatively can lead to diversification processes. In this context, our theoretical investigations have shown that rapid evolutionary processes may be triggered by population isolation and reconnection events occurring over a large time-scale, as these generate a transient excess of genetic diversity in populations providing the required raw material for rapid evolution. Accordingly, we have demonstrated that successive population isolation and reconnection events occurring during the quaternary climate cycles are associated with high diversification rates in animal clades and intricate patterns of genetic diversity in African hippopotamus. Importantly, these events leave distinguishable genomic signatures on SFS and as well as on common summary statistics of DNA variation in populations. Using these theoretical results and full genome polymorphism data we infer the recent evolutionary history of human immunodeficiency virus (HIV-1) in Asia and South America and successfully retrieve the successive HIV subtype colonization events in these regions.

58

Probabilistic macrosynteny model for inferring the structure of ancient pre-WGD genomes

Yoichiro Nakatani[1], Aoife McLysaght[1]

1. Trinity College Dublin, the University of Dublin, Dublin, Ireland

The human genome harbors ~7000 ohnolog genes, or duplicates that derive from whole-genome duplication (WGD) events at the origin of vertebrate. They are often associated with human diseases, and it is therefore important to make a comprehensive catalog of ohnologs. High-confidence identification of ohnologs hinges on synteny analysis and inference of pre- and post-WGD ancestral genome structures, but the ancient timing of teleost and vertebrate WGD events impedes high accuracy inference. Because of this difficulty, previous studies excluded a large part of the human genome with ambiguous synteny, resulting in low-coverage reconstructions

With the aim of explicitly dealing with reconstruction uncertainty, we developed a probabilistic model of macrosynteny conservation and devised variational Bayes algorithms for inferring the structure of pre-WGD genomes. We obtained high-coverage reconstructions of the ancestral vertebrate and teleost genomes by applying the method to the human, mouse, chicken, spotted gar, zebrafish, stickleback, *Tetraodon*, medaka, and amphioxus genomes. The results show that previously excluded regions in the modern vertebrate genomes tend to be comprised of multiple

smaller synteny blocks with varying degrees of reconstruction probability, which represents reconstruction uncertainty due to incomplete genome assembly, intensive local rearrangements, etc. Our reconstructions provide an improved picture of early vertebrate genome evolution, showing how ancestral vertebrate chromosomes are retained in the modern genomes, how inter-chromosomal rearrangements occurred in individual vertebrate lineages, and how specific regions in the human genome originated by the vertebrate WGDs.

59

Standard Codon Substitution Models Overestimate Purifying Selection for Non-Stationary Data

Benjamin D Kaehler[1], Von Bing Yap[2], Gavin A Huttley[1]

1. ANU College of Medicine, Biology and Environment, Acton, ACT, Australia

2. Statistics and Applied Probability, National University of Singapore, Singapore

Estimation of natural selection on protein-coding sequences is a key comparative genomics approach for de novo prediction of lineage specific adaptations. Selective pressure is measured on a per-gene basis by comparing the rate of non-synonymous substitutions to the rate of neutral evolution, typically assumed to be the rate of synonymous substitutions. All published codon substitution models have been time-reversible and thus assume that sequence composition does not change over time. We previously demonstrated [1] that if time-reversible DNA substitution models are applied blindly in the presence of changing sequence composition, the number of substitutions is systematically biased towards overestimation. We extend these findings to the case of codon substitution models and further demonstrate that the ratio of non-synonymous to synonymous rates of substitutions tends to be underestimated over three data sets of insects, mammals, and vertebrates. Our basis for comparison is a non-stationary codon substitution model that is heterogeneous across lineages and allows sequence composition to change. Model selection and model fit results demonstrate that our new model tends to fit the data better. Direct measurement of non-stationarity shows that bias in estimates of natural selection and genetic distance increases with the violation of the stationarity assumption. Additionally, inferences drawn under time-reversible models are systematically affected by compositional divergence. As genomic sequences accumulate at an accelerating rate, the importance of accurate de novo estimation of natural selection increases. Our results establish that our new model provides a more robust perspective on this fundamental quantity.

1. Kaehler, B. D., Yap, V. B., Zhang, R., and Huttley, G. A. 2015. Genetic distance for a general non-stationary markov substitution process. Syst Biol, 64(2): 281–293.

60

Multiple nucleotide mutations cause rampant false positive inferences of selection on the human lineage

Aarti Venkat[1], Joseph Thornton[1], Matthew Hahn[2]

1. University of Chicago, Chicago, IL, United States

2. Department of Biology, Indiana University, Bloomington, Indiana, United States

The branch-sites test has been the basis for thousands of inferences of genes under lineage-specific positive selection. The test's models assume that mutations occur independently. But DNA replication is known to produce multiple mutations within a codon more frequently than expected under independence, and such multi-nucleotide mutations (MNMs) are more likely than single mutations to be non-synonymous. We therefore hypothesized that MNMs produced by neutral processes might cause false inferences of positive selection in the branch-sites test and sought to determine the extent of this bias, if any, along the human lineage. We analyzed a mammalian genome-wide dataset and found that codons with MNMs provided all the support for the positive selection model. When multi-nucleotide mutational processes were incorporated into the branch-sites model, 93% of genes positively selected in the original test lost their signatures of selection. To determine if realistic rates of MNM generation cause false positive inferences, we simulated evolution under model parameters derived from the mammalian dataset with MNMs but without positive selection; we found that conditions associated with 96% of genes analyzed led to unacceptable false positive rates by the branch-sites test. Under typical genome-wide evolutionary conditions, a rate of MNM production considerably lower than that observed in experimental studies of mutational processes was sufficient to cause frequent false positive inferences. Our results indicate that many genes found to be under positive selection using the branch-sites test – including the majority of such genes on the human lineage -- may be artifacts of unincorporated neutral mutational processes.

61

Assessing structural awareness in models of protein evolution

Claudia C Weber[1], Peter B Chi[2], Stephen D Shank[1], David A Liberles[1]

1. Temple University, Philadelphia, PA, United States

2. Ursinus College, Collegeville, PA, United States

A key application of phylogenetic methods is to characterize natural selection on proteins. However, their success depends on how well they capture evolutionary constraints. The dominant method of selecting models for an analysis is to compare statistical fit to observed data. Unfortunately, this provides limited information on which properties of coding sequences selection acts on.

Codon models typically assume amino acid changes are, on average, selectively disfavoured without considering physiochemical amino acid properties. Empirical protein models that consider exchangeabilities from alignments seem comparatively appealing, and may incorporate between-site heterogeneity to reflect intramolecular variation in constraints. Nevertheless, they often fit data no better than codon models. This needn't suggest that biophysical amino acid properties are unimportant as they also disregard site specific preferences. Given these limitations, we approach developing more structurally aware models from two angles: First, we consider Mutation-Selection models, allowing estimation of selective coefficients associated with different classes of amino acid change and site-specific preferences. Second, discussed here, we examine how well available models capture structural constraint at phylogenetically relevant timescales.

We therefore perform forward simulations on the SH2 domain, assessing deviation of the evolved sequence from the native structure. For each time-point, we predict the evolved structure using Rosetta and determine the RMSD. Selection criteria include: a) Exchangeabilities from LG08+4dG; b) Physicochemical distances (Grantham 1974); c) Site specific amino acid preferences (Rodrigue et al., 2010); d) Fold stability based on contact affinities (Miyazawa and Jernigan, 1985), incorporating heterogeneity and epistasis; e) A coarse-grained biophysical model (Grahnen et al., 2012). Additionally, we examine how LG08 performs with and without heterogeneity, providing insight into how rate variation contributes to more

realistic models. While some of these methods have seen limited use in a phylogenetic context, comparing them with established approaches allows us to take a broad view on useful selection criteria for future models.

62

Patterns of adaptive evolution: a structural perspective

Greg Slodkowicz[1], **Nick Goldman**[1]

1. EMBL-European Bioinformatics Institute, Cambridge, UNITED KINGDOM, United Kingdom

Protein structure is a major cause of site-to-site evolutionary rate variation. Many structural features such as solvent accessibility, local packing density and proximity to active sites or interfaces have been shown to modulate the evolutionary rate. It is, however, not well understood how these features affect the prevalence of adaptive evolution. Most codon-based models, which are commonly applied for detecting sites under positive selection, do not incorporate any information about the protein structure.

In this study, we attempted to form a better view of adaptation on molecular level by asking whether residues under positive selection are close to each other on the protein structure. We generated a large dataset of trees and alignments for 39 mammalian species (covering over 80% of human genes) and calculated sitewise values of selective constraint (dN/dS). We then mapped positively-selected sites onto available crystal structures and analysed whether they tend to be co-located by statistically assessing the distribution of pairwise distances between them.

We find that positively-selected sites frequently form tight clusters on protein structures and that this conclusion is robust to low alignment quality and other technical issues. Identified clusters can be assigned into one of several categories: we find that groups of positively-selected residues can surround active sites, occur in binding regions, and form small, linear clusters in the N-termini of proteins. To our knowledge, the last of these findings has not been previously reported. Additionally, the prevalence of clustering varies in different enzyme classes, with oxidoreductases exhibiting the most evidence for clustering.

63

Disentangling Sources of Selection on Exonic Transcriptional Enhancers

Rachel Agoglia[1], **Hunter Fraser**[2]

1. Department of Genetics, Stanford University, Stanford, CA

2. Department of Biology, Stanford University, Stanford, CA

In addition to coding for proteins, exons can also impact transcription by encoding regulatory elements such as enhancers. It has been debated whether such features confer heightened selective constraint, or evolve neutrally. We have addressed this question by developing a new approach to disentangle the sources of selection acting on exonic enhancers, in which we model the evolutionary rates of every possible substitution as a function of their effects on both protein sequence and enhancer activity. In three exonic enhancers, we found no significant association between evolutionary rates and effects on enhancer activity. This suggests that despite having biochemical activity, these exonic enhancers have no detectable selective constraint, and thus are unlikely to play a major role in protein evolution.

64

Malaria parasites drive adaptation in mammalian genomes

Emily R Ebel[1], **Dmitri A Petrov**[1], **David Enard**[1]

1. Stanford University, Stanford, CA, United States

Human malaria parasites are closely related to dozens of other *Plasmodium* species that infect non-human primates, rodents, and other mammals. Although several examples of recent adaptation to malaria in human populations have been discovered, little is known about the deeper evolutionary impact of the parasite on the mammalian lineage. From a set of 9,338 proteins conserved in 24 mammal species, we manually identified 412 proteins linked to malaria phenotypes in the literature. Models of codon evolution demonstrate that these 'malaria-interacting-proteins', or MIPs, have been exceptional targets of positive selection throughout mammalian evolution.

We find that MIPs with immune functions have been the primary targets of adaptation. Interestingly, about half of MIPs are also known to interact with viruses and bacteria, and adaptive patterns do not appear limited to lineages known to contract *Plasmodium*. In many cases, the pervasiveness of pleiotropy can make it difficult to attribute genetic adaptation to a single selective pressure. However, we use comparative analyses to show that overlap with viruses and bacteria cannot explain the excess of adaptation in MIPs. Instead, we suggest that *Plasmodium*, along with other insect-vectored Apicomplexan parasites, have driven pervasive adaptation in a large set of proteins throughout mammalian evolution.

65

Evolution of dominance/recessivity interactions between self-incompatibility alleles in Arabidopsis.

Samson Simon[1], **Vincent Castric**[1]

1. CNRS, Villeneuve D'Ascq, France

Self-incompatibility in plants of the Brassicaceae family is controlled by a highly diversified molecular lock-and-key system consisting of a large set of specific haplotypic combinations of only two genes. This system has been a textbook example of natural (balancing) selection, in the form of a strong reproductive advantage for individuals expressing rare alleles. These haplotypes also form a striking linear dominance/recessivity hierarchy, whereby most heterozygote combinations express only one self-incompatibility specificity at the phenotypic level. In this seminar, I will detail how we recently identified the molecular determinants of this dominance hierarchy and showed that it is based on a complex regulatory network based on the interactions between a dedicted set of small non-coding RNAs produced by dominant alleles and their target sites in recessive alleles. I will review several key features of the topology of these interactions and combine theoretical modelling and functional approaches to discuss our current understanding of the functional and selective constraints on the evolution of the network.

66

Evolutionary changes in the regulatory behaviour of the lac operon

Olin Silander[1], Diana Blank[2]

1. *Massey University, Albany, Auckland, AUCKLAND, New Zealand*
2. *Biozentrum, University of Basel, Basel, Switzerland*

Many phenotypic differences between species are driven by changes in transcriptional regulation, frequently detected as a change in mRNA transcript levels. Here we use a plasmid-based system and flow cytometry to explore how the regulation of the lac operon differs among natural isolates of *E. coli*. We find both cis- and trans- changes that affect regulation. These changes not only affect transcript levels, but inducer sensitivity, the speed of transcriptional change, and the level of variation between individual cells (transcriptional noise). We then focus specifically on two SNPs within the lac operon that differ between two closely related natural isolates, and quantify the effects each of these polymorphisms on regulation. The substantial differences we find in the regulatory behaviour of the *E. coli* lac operon suggests that there is significant, on-going selection on this phenotype in nature.

67

Mechanisms underpinning the rapid functional evolution of the human brain

Guy Barry[1], James A Briggs[2], Boris Guennewig[3, 4], Frank Rigo[5], Ernst J Wolvetang[2], Kristen J Brennand[6], John S Mattick[3, 4]

1. *QIMR Berghofer Medical Research Institute, Brisbane, QLD, Australia*
2. *Australian Institute for Bioengineering and Nanotechnology, The University of Queensland, Brisbane, Qld, Australia*
3. *Garvan Institute of Medical Research, Darlinghurst, NSW, Australia*
4. *St Vincent's Clinical School, The University of New South Wales, Sydney, NSW, Australia*
5. *Isis Pharmaceuticals, Carlsbad, CA, USA*
6. *Departments of Psychiatry and Neuroscience, Icahn School of Medicine at Mount Sinai , New York, NY, USA*

Recent advances in next generation sequencing (NGS) and human induced pluripotent stem cell (iPSC) technologies have allowed an unprecedented view into the transcriptional dynamics of the human brain. Particularly revealing has been the discovery of long non-coding RNAs (lncRNAs) that have substantially expanded in recent evolution, with >50,000 lncRNA genes identified in the human genome, one-third being primate-specific and the majority expressed predominantly in the brain. We employed whole transcriptome deep sequencing of activated neurons derived from human iPSCs (control and schizophrenia-associated) to investigate gene expression. We found that the expression of distinct subsets of mRNAs and lncRNAs are altered in response to neuronal depolarisation and these changes are strongly associated with metabolic systems. Altered metabolism may have significantly played a role in changes underpinning the substantial expansion of the human brain over the last 2 million years in response to changing diets and high energy foods obtained through the invention of tools and the domestication of fire for cooking. Furthermore, we find that there are significant differences in the robustness of transcriptomic responses between iPSC-derived neurons from schizophrenia patients compared with unaffected controls, which are only evident upon activity, suggesting that psychiatric conditions may arise from fragilities in newly evolved mechanisms relating to dynamic neural pathways. These results present evidence that metabolic pathways may be intimately involved in the recent sophistication of the human brain and dysregulation thereof may contribute to psychiatric disease.

68

Evolutionary constraints reveal stress-triggered gene regulation by poly(A)-binding protein

D. Allan Drummond[1], Joshua A Riback[2], Chris Katanski[1], Tobin R Sosnick[1], Evgeny V Pilipenko[1], Jamie L Scott[1], Edward WJ Wallace[3]

1. *Biochemistry & Molecular Biology, The University of Chicago, Chicago, IL, 60637*
2. *The University of Chicago, Chicago, ILLINOIS, United States*
3. *University of Edinburgh, Edinburgh, Scotland*

In eukaryotic cells, heat shock and other stresses trigger the accumulation of proteins and RNA into cytosolic ribonucleoprotein (RNP) stress granules marked by poly(A)-binding protein. Formation of stress granules is thought to be an adaptive response involved in global regulation of translation, yet these ideas have been difficult to test given the lack of mutational perturbations linking protein behavior to phenotype. Stress granule formation has been reported to involve multivalent RNA/protein and protein-protein interactions such as those mediated by intrinsically disordered regions (IDRs) whose molecular evolution remains enigmatic. Here we report that poly(A)-binding protein itself, Pab1 in yeast, autonomously forms heat-induced RNP granules *in vitro*. Pab1's highly conserved IDR, the proline-rich "P domain", modulates but is not essential for the formation of heat-induced RNP granule assembly *in vivo* and *in vitro*. Evolutionary analysis of the P domain reveals previously unappreciated patterns of selection on its composition, particularly its aliphatic residues. We show experimentally that these residues tune the IDR's conformational and biological properties including collapse of the domain, heat-induced Pab1 granule formation *in vivo* and *in vitro,* and yeast thermotolerance. Although heat-induced protein aggregation is generally thought to be harmful, we discover that mutations that reduce Pab1's heat-triggered aggregation also reduce cells' ability to grow at elevated temperatures. Pab1's role as a translational regulator appears tightly linked to its stress-triggered self-assembly. Our results indicate that poly(A)-binding protein's heat-induced aggregation represents a largely autonomous, evolutionarily tuned, adaptive self-assembly response to stress.

69

Recessive selection in complex disease and implications for variant discovery

Daniel M Jordan[1], Daniel J Balick[2], Ivan A Adzhubei[2], Shamil R Sunyaev[2], Ron Do[1]

1. *Icahn School of Medicine at Mount Sinai, New York, NY, United States*
2. *Brigham and Women's Hospital, Boston, MA, United States*

Methods to identify genes and variants for complex diseases generally assume that effects of alleles are additive, meaning that a homozygous genotype confers twice the risk of a heterozygous genotype. Genes under recessive selection exhibit different population dynamics, especially in

populations that have undergone dramatic bottlenecks followed by re-expansion, such as the European population [1]. The majority of genome-wide association studies (GWAS) have assumed additive models; however, the few studies that have tested recessive models have discovered recessive associations [2]. We have developed a novel method to quantify the magnitude and mode of selection of all protein-coding genes across the human genome by comparing European population sequencing data from the Exome Aggregation Consortium (ExAC) (N=35,000) with simulated evolutionary histories for both additive and recessive alleles, This method could inform model choice by identifying genes and pathways likely to be under recessive selection. We find a variety of biologically meaningful categories enriched in the predicted recessive class, including glycoproteins (Benjamini P=6.3x10^{-13}), immunoglobulin domains (P=0.023), and inflammatory response (P=0.0052). The enrichment for inflammatory genes in particular suggests that many complex diseases with inflammatory components may be under recessive selection, such as Crohn's disease and rheumatoid arthritis (RA). In the case of Crohn's disease, we find that genes with large and well-validated effects are predicted to be under recessive selection, while genes implicated by GWAS studies have no enrichment for recessive selection. Similarly, RA loci discovered by GWAS show no enrichment for recessive selection, despite the fact that RA is known to involve pathways that are highly enriched for recessive selection according to our method. This highlights the need for using recessive models in GWAS, and the potential usefulness of our catalog of mode of selection as a tool for gene prioritization.

1. Balick DJ, Do R, Cassa CA, Reich D, Sunyaev SR (2015) Dominance of Deleterious Alleles Controls the Response to a Population Bottleneck. PLoS Genet 11(8): e1005436. doi:10.1371/journal.pgen.1005436
2. Nikpay M, Goel A, Hall LM, Willenborg C, et al. for the CARDIoGRAMplusC4D Consortium (2015) A comprehensive 1000 Genomes–based genome-wide association meta-analysis of coronary artery disease. Nature Genet 47: 1121-1130. doi:10.1038/ng.3396

70

Genomic signatures of soft selective sweeps at human complex traits

Emily Wong[1], Matthew Robinson[2], Joseph Powell[3]

1. School of Biological Sciences, University of Queensland, St Lucia, Australia

2. The Centre for Neurogenetics and Statistical Genomics, Queensland Brain Institute, University of Queensland, St Lucia, QLD, Australia

3. Institute for Molecular Bioscience, University of Queensland, St Lucia, QLD, Australia

A central goal in population genomics is to detect targets of adaptive natural selection. Yet, elucidating the functional consequences of those variants under selection and understanding their role in driving phenotypic diversity remains a significant challenge. We leverage the availability of human population genomics data where, through genome-wide association studies, genetic variants have been associated with numerous disease and non-disease traits. We use this information to uncover selective forces underlying phenotypic evolution. We measure the extent of positive selection on standing genetic variants using principle component analysis to identify those segregating across subpopulations at a rate higher than genetic drift. We find evidence for genetic adaption among human populations at SNPs associated with complex diseases.

71

Molecular Evolution of Human Breasts

Reed A Cartwright[1]

1. Arizona State University, Tempe, AZ, United States

The development of mammaries in humans is unique among mammals. During puberty, the breasts of human females enlarge, and no other mammal exhibits similar development before pregnancy and lactation. This phenotype is fixed among humans, evolving some time after our lineage separated from chimpanzees and bonobos, and anthropologists have offered many competing hypotheses about the selective advantage of human breasts. In female mammals, mammary development occurs across several stages: fetal development, puberty, pregnancy, lactation, and post-lactation. We hypothesize that in humans the development of the mammary ductal system during puberty also causes the surrounding tissue to permanently enlarge, producing the unique human phenotype. We have examined the molecular evolution of genes related to mammary development in order to identify candidate genes related to the pubertal enlargement of human breasts. Here we report our progress and discuss the difficulties of studying a non-pathogenic complex phenotype in humans.

72

A dominant TRPV4 variant underlies osteochondrodysplasia in Scottish fold cats

Barbara Gandolfi[1], Sultan Alamri[2], Bill G Darby[3], Jim C Lattimer[1], Richard Malik[4], Claire M Wade[5], Leslie A Lyons[1], Jianlin Cheng[6], John F Bateman[2], Peter McIntyre[3], Shireen R Lamande[2], Bianca Haase[5]

1. Veterinary Medicine and Surgery, University of Missouri-Columbia, Columbia, USA

2. Murdoch Childrens Research Institute, Melbourne, VIC, Australia

3. School of Medical Sciences, RMIT University, Bundoora, VIC, Australia

4. Centre for Veterinary Education, University of Sydney, Sydney, NSW, Australia

5. Veterinary Science, University of Sydney, Sydney, NSW, Australia

6. Computer Science Department, Informatics Institute, C. Bond Life Science Center, University of Missouri, Columbia, MO, USA

Inherited osteoarthropathy in human is characterized by a progressive arthropathy of fingers and toes and while affected individuals appear normal at birth, they develop severe arthropathy by adulthood. The distinctive and defining physical trait of the Scottish Fold cat breed is the characteristic ear phenotype. Scottish fold cats have ears which fold forward, this presumably reflecting lack of resilience of the pinna and auricular cartilages. There is convincing evidence that Scottish Fold cats have an underlying congenital defect which affects the structure and function of cartilage, resulting in progressive bone, joint and cartilage abnormalities that subsequently lead to progressive dysfunction. Cats develop a variable osteochondrodystrophy causing abnormal bone development likely through defective endochondral ossification, progressive osteoarthritis and lameness. Pedigree analyses and breeding experiments have shown that the trait is inherited as an autosomal monogenic dominant trait with variable expression. We applied a whole-genome SNP association mapping approach using a total of 78 cats (53 Scottish fold cats and 25 Scottish shorthairs). DNA samples were genotyped with the feline Illumina 63kSNP genotyping microarray. A genome-wide significant association on chromosome D3 has been identified and confirmed with fine structure mapping. A candidate gene analysis revealed a missense mutation in a calcium channel associated with skeletal dysplasia in humans.

73

The Evolution of Venoms: Lessons from Cone Snails

Baldomero Olivera[1], **Helena Safavi-Hemami**[2, 1]

1. Department of Biology, University of Utah, Utah

2. University of Copenhagen, Copenhagen

The cone snails are a biodiverse lineage of ca 700 species of venomous marine gastropods. Several species have been intensively studied; each cone snail venom expresses 50-150 venom components, mostly small, disulfide-rich peptides. It is clear that the venom is organized into groups of bioactive peptides that act together; these groups have been called “cabals”. The cone snails use a combination drug strategy to achieve each specific physiological endpoint. Most remarkably, the genes that encode these venom peptides are subject to an unprecedented rate of accelerated evolution, so that venom peptides are different in each species.

Recently, a particularly insightful example of venom evolution was gained by the discovery that cone snails use insulin to make their prey hypoglycemic (and therefore easier to capture). Early in the evolution of the genus, there was a duplication of the insulin gene; the endogenous copy has remained essentially conserved; in contrast, the copy expressed in venom has greatly diversified. A comparison between the endogenous gene, and the insulin “exogene” is instructive in reflecting the different selective pressures that each of the duplicated copies has been subject to over evolutionary time. An overview of these differences will be presented.

74

Evolution and Development of Venom in Snakes

Michael K Richardson[1]

1. University of Leiden, UK, United Kingdom

Some snakes lineages have evolved specialised venoms. These represent a fascinating model of protein evolution under strong selection. We have sequenced and analysed the genomes of a number of snake species. Major toxic components of the venom, such as three finger toxins, have undergone massive duplication after recrutiment to the venom gland through a variety of mechanisms. Interestingly, we also find that venomous snakes are not only engaged in an arms race against their prey by evolving new toxins, but they may also be in an arms race against bacteria in the venom gland. Future prospects include the study of developmental changes in venom protein composition during the lifetime of individual snake as they switch from small arthropod prey to larger rodent prey. This type of study requires a new generation of bioassays for studying venom bioactivity with low volumes and without using mammalian models. We have developed specialised assays based on zebrafish embryos.

75

The rise and fall of an evolutionary innovation

Kartik Sunagar[1], **Yehu Moran**[1]

1. The Hebrew University of Jerusalem, Jerusalem, JERUSALEM, Israel

Animal venoms are theorized to evolve under the significant influence of positive Darwinian selection in a chemical arms race scenario, where the evolution of venom resistance in prey and the invention of potent venom in the secreting animal exert reciprocal selection pressures. Venom research to date has mainly focused on evolutionarily younger lineages, such as snakes and cone snails, while mostly neglecting ancient clades (e.g., cnidarians, coleoids, spiders and centipedes). Molecular evolutionary assessment of over 3500 nucleotide sequences from 85 toxin families spanning the breadth of the animal kingdom has unravelled a contrasting evolutionary strategy employed by ancient and evolutionarily young clades. We show that the venoms of ancient lineages remarkably evolve under the heavy constraints of negative selection, while toxin families in lineages that originated relatively recently rapidly diversify under the influence of positive selection. We propose that animal venoms mostly employ a ‘two-speed’ mode of evolution, where the major influence of diversifying selection accompanies the earlier stages of ecological specialization (e.g., diet and range expansion) in the evolutionary history of the species – the period of expansion, resulting in the rapid diversification of the venom arsenal, followed by longer periods of purifying selection that preserve the potent toxin pharmacopeia – the period of purification and fixation. Thus, for the first time, we highlight the significant role of purifying selection in shaping venoms, and we propose a new model to explain the evolution of venom in the animal kingdom (Sunagar and Moran, 2015).

1. Sunagar K, Moran Y. 2015. The Rise and Fall of an Evolutionary Innovation: Contrasting Strategies of Venom Evolution in Ancient and Young Animals. PLoS Genetics. 11:e1005596.

76

Toxin composition in the tentacles of the Australian cold temperate sea anemone, *Oulactis* sp.

Michela L. Mitchell[1], **Rodrigo A. V. Morales**[1], **Gerry Q. Tonkin-Hill**[2], **Tony T. Papenfuss**[2], **Anthony W. Purcell**[3], **Raymond S. Norton**[1]

1. Medicinal Chemistry, The Monash Institute of Pharmaceutical Sciences, Monash University, Parkville, Victoria, Australia

2. Bioinformatics Division, Walter & Eliza Hall Institute of Research, Parkville, Victoria, Australia

3. Department of Biochemistry and Molecular Biology, Monash University, Clayton, Victoria, Australia

Cnidarians are one of the oldest known animal lineages (approx. 700 million years) with a specialised envenomation apparatus to deliver toxins. These toxins have proven potential as a basis for new pharmaceutical therapeutics. Knowledge is lacking, however, about the diversity of sea anemone toxins or their distribution in relation to morphological regions used in prey capture and defence/aggression. The toxins are delivered through specialised cells, the cnidae, which are broadly classified as spirocysts, holotrichs, basistrichs, mastigophores and atrichs, each performing a different biological function. In addition to the limited knowledge of toxin source and distribution, endemic Australian sea anemone fauna have not been examined for their toxins. We use a venomic strategy to examine the toxin content and body distribution of these molecules in the endemic Australian species *Oulactis*sp. The tentacle transcriptomes of three individuals were assembled from Illumina sequencing data. Identified toxins were further validated by mass spectrometry (MS/MS) and their distribution in the animal tissue determined via MALDI-IMS (Matrix-assisted laser desorption/ionization imaging). Consistent with previous studies of cnidarians, we have identified and located a number of toxins, including cytolysins, phospholipases, peptidases and ion channel modulators. The strategy employed here will be used to examine other body regions

of *Oulactis* sp., significant for their unique cnidae profile, the actinopharnyx (throat), acrorhagi, mesenterial filaments and column. These data will aid our understanding of the functional evolution of sea anemone toxins and identify novel peptides that may be useful pharmaceutically.

77

When sex matters: Dramatic sexual dimorphism in the venom and venom system of the centipede *Scolopendra hardwickei*

Eivind Undheim[1, 2], D Richards[2], N Ngum[2], D Tooth[2], DH Kim[2], I Mellor[2]

1. Centre for Advanced Imaging, University of Queensland, St Lucia, QLD, Australia

2. School of Life Sciences, University of Nottingham, University Park, Nottingham, United Kingdom

Although sexual dimorphism has been well documented in spiders and certain snake species, it remains poorly documented among most venomous animals. This is particularly the case for species where there are no obvious differences in the appearances of males and females. Using a combination of gas-chromatography mass spectrometry, proteomics, transcriptomics, electrophysiology, and magnetic resonance imaging we describe the venom and venom system of the striking, aposematic centipede *Scolopendra hardwickei*. We also provide the first insight into the venom arsenal of any *singl*ecentipede specimen, as well as the first detailed characterisation of the low-molecular weight non-peptidic components of any centipede venom. Despite no obvious differences in non-reproductive behaviour or morphology between males and females, our results demonstrate dramatic sexual dimorphisms in venom composition, pharmacology, and venom gland morphology. We show that there are substantial differences in the relative abundance and expression levels of high *versus* low molecular weight components, and that males and females appear to employ very different venom strategies. Although we can only speculate as to the differences in function of male and female *S. hardwickei* venom, our results highlight the important role that sex-specific natural selection can play in the evolution of centipede venoms.

78

Going with the flow? The genetic basis for snake venom evolution

Giulia Zancolli[1], John Mulley[1], Wolfgang Wüster[1]

1. School of Biological Sciences, Bangor University, Bangor, United Kingdom

Snake venoms are complex cocktails of toxins acting synergistically to subdue and kill prey or predators. Venom composition is extremely variable not only between different species, but also within the same taxon causing grave problems to snakebite treatment. The underlying drivers and mechanisms of this variation remain poorly understood, particularly the relative importance of gene flow between populations. The Mohave rattlesnake, *Crotalus scutulatus*, displays extreme venom variation across a continuous distribution in Southwestern USA. Hence it represents an ideal model system to investigate both the genetic mechanisms underlying geographic venom variation and the extrinsic forces shaping it. Here we test whether venom variation in *C. scutulatus* (i) is the result of genetic differences at loci coding for toxin proteins, or (ii) is caused by differences in gene expression, and (iii) whether it reflects patterns of neutral genetic variation. We used proteomic analysis to characterize the venom phenotypes, a PCR assay to test for presence/absence of the most highly expressed toxin-encoding genes, and 16 microsatellite markers to infer population structure of *C. scutulatus*. Presence/absence of major venom proteins was strictly linked with presence/absence of the corresponding coding genes. Surprisingly, we found weak population structure and low genetic differentiation (Fst = 0.062), with no significant correlation between venom phenotypes and neutral genetic variation. These results suggest that forces other than neutral genetic drift are able to maintain marked differences of adaptive genes in the presence of gene flow. Furthermore, we revealed a genetic basis for snake venom variation across multiple toxin families.

79

Green eggs and sperm: gamete chemoattraction and sexual isolation in wild *Solanum* species

Leonie C Moyle[1], Cathleen P Jewell[1]

1. Indiana University, Blommington, INDIANA, United States

Sexual signals are critical for normal reproductive function within species and, when divergent between lineages, can also contribute to the expression of species reproductive isolation. We are examining the evolution and genetics of sexual signaling in several systems, including chemoattraction between male and female gametes in wild tomato species. Using a *semi in vivo* assay to directly observe interactions between pollen tubes and ovules, we show that pollen is guided to ovules via an ovule-secreted chemoattractant, and that some species pairs show reductions in pollen tube attraction to heterospecific ovules consistent with the evolution of partial isolation. We identify two cysteine-rich peptides as candidate ovule-secreted chemoattractants, and show that one candidate has fixed non-synonymous differences between species that show reduced chemoattraction. A genetic analysis using ovules from hybrid introgression lines shows that this locus induces pollen tube behavior that recapitulates con- and heterospecific patterns of gamete chemoattraction, supporting it as a strong candidate for the ovule-secreted chemoattractant. These findings address the phenotypic expression and molecular basis divergent heterospecific gamete signaling across multiple species a clade, and demonstrate how sexual signaling that facilitates intraspecific mating events can contribute to reduced sexual compatibility between lineages.

80

The evolution of sperm size and post-mating pre-zygotic reproductive isolation

Asher D Cutter[1]

1. University of Toronto, Toronto, ON, Canada

Sperm cells provide crucial, if usually diminutive, ingredients to successful sexual reproduction as the source of centrioles and half the diploid genome. Despite this essential conserved function, sperm competition and coevolution with female traits has driven spectacular morphological change across the tree of life in these discrete cells. Here I characterize repeated instances of convergent evolution of sperm gigantism across the phylogeny of *Caenorhabditis* nematodes. Species at the extreme end of the 50-fold range of sperm-cell volumes have sperm capable of comprising ~5% of egg-cell volume, representing severe attenuation of the magnitude of anisogamy. We demonstrate that sperm size variation establishes early in spermatogenic development, by meiosis I during the formation of primary spermatocytes. We hypothesize that life history differences

among species favored the evolution of alternative sperm competition strategies toward many small vs. fewer large sperm. A byproduct of the outcome of within-species sperm competition manifests during between-species mating errors, in which sperm from the 'wrong' species can invade distal gonad and somatic tissue of the mated female, leading to sterility and premature death. These gametic mismatches represent a form of post-mating pre-zygotic reproductive isolation, suggesting a potentially cryptic role of sexual selection by sperm competition in the speciation process.

81

A "sexy" co-evolutionary arms race: insights from structural and biochemical studies of egg-sperm interactions in Pacific abalone

Damien B Wilburn[1], Willie J Swanson[1]

1. University of Washington, Seattle, WASHINGTON, United States

A hallmark of reproductive proteins that mediate egg-sperm interactions is their rapid evolution. The strong selective pressure to maintain successful fertilization coupled to differences in male/female reproductive strategies often promote arms race dynamics that drive accelerated evolution to maintain high affinity protein-protein interactions. While reproductive isolation often occurs through changes in timing or location of reproduction, the continual co-evolution and molecular refinement of gamete recognition proteins can create boundaries to hybrdization. For the marine gastropod abalone, seven sympatric species with similar breeding seasons live off the coast of California, yet hybrids are rarely observed. During fertilization, abalone sperm secrete a 16 kDa acrosomal protein, lysin, which specifically binds to repeat domains in the egg coat protein VERL. Lysin-VERL interactions are species specific, and molecular evolutionary studies demonstrate strong signatures of positive selection and rapid co-evolution between the two proteins. However, lysin acquires approximately five times as many non-synonymous substitutions as VERL, and the molecular mechanism of how these mutational effects contribute to high-affinity, species specific interactions remains unclear. Using multidimensional NMR, we are characterizing the 3D structures of lysin and VERL from red abalone (*Haliotis rufescens*). Mutagenesis of key residues that mediate oligomerization state for these proteins strongly affect their functions without altering the tertiary structures. Homology modeling of lysin and VERL from additional abalone species will provide a structural framework to help understand the evolutionary forces driving high affinity conspecific interactions.

82

Sexual selection for genetic compatibility: new insights into the genetic basis of cryptic female choice in Chinook salmon

Cornelia Geßner [1], Patrice Rosengrave[1], Monika Zavodna[1], Shinichi Nakagawa[2], Robert Montgomerie[3], Neil Gemmell[1]

1. Department of Anatomy, University of Otago, Dunedin, New Zealand

2. School of Biological, Earth and Environmental Sciences, University of New South Wales, Sydney, Australia

3. Biology Department, Queen's University, Kingston, Canada

Cryptic female choice (CFC), a post-mating version of sexual selection, enables females to bias fertilization outcomes towards certain males in a non-random fashion. Here we report our investigations on several potentially important intersexual postcopulatory gametic interactions in a population of chinook salmon (*Oncorhynchus tshawytscha)*: the effect of female ovarian fluid (OF) on the behaviour of spermatozoa during fertilisation, and the effects of MHC variability, multilocus heterozygosity (as an index of male quality) and female-male genetic relatedness on sperm behaviour and and male fertilisation success when there is sperm competition in the presence of that ovarian fluid. To do this, we conducted a series of *in vitro* competitive fertilisation experiments and found that, when ejaculates from two males are competing for access to a single female's unfertilised eggs, fertilisation success was significantly biased toward the male whose sperm swam fastest in the female's ovarian fluid. Fertilization success was higher for males more similar to the focal female at both microsatellites, SNPs and the MHC class II than their competitor, perhaps indicating that the MHC class II and other genetic factors mediate sperm-egg interactions. Embryo survival—a measure of fitness—was also positively correlated with both sperm swimming speed in ovarian fluid and male multilocus heterozygosity, providing novel evidence that cryptic female choice is adaptive for the female, enhancing the early survival of her offspring and potentially influencing her fitness.

83

The case for selection at coevolving human gamete-recognition genes (*ZP2*, *ZP3*, and *C4BPA*)

Michael W Hart[1], Daryn A Stover[2], Sahar V Mozaffari[3], Carole Ober[3], Carina F Mugal[4], Ingemar Kaj[4]

1. Simon Fraser University, Burnaby, British Columbia, Canada

2. Arizona State University, Lake Havasu City, AZ, United States

3. University of Chicago, Chicago, IL, United States

4. Uppsala University, Uppsala, Sweden

Previous research has suggested that selection due to sexual conflict between mates over the optimal rate of sperm-egg binding has led to coevolution between male adaptations and female countermeasures at gamete recognition genes. Coevolution of pairs of gamete recognition genes expressed on sperm and eggs may be associated with both natural reproductive variation and clinical infertility in humans. We attempt to identify likely targets of selection in two genes (*ZP2*, *ZP3*) that contribute to the thick extracellular egg coat (or zona pellucida), and one gene (*C4BPA*) that encodes the associated sperm-head receptor. Specifically, branch-site codon models were applied to the 1000 Genomes Phase 1 dataset to identify candidate protein-coding mutations evolving under positive selection. These candidate mutations differentiated two haplogroups of intermediate frequency at each gene and may represent balanced polymorphisms. While strong linkage disequilibrium (LD) consistent with coevolving loci was not detected among candidate mutations in the 1000 Genomes dataset, significant LD was detected between *ZP2* and *C4BPA* in a well-studied founder population of Hutterites. Candidate mutations in *ZP2* and *C4BPA* were also significantly correlated with variation in family size and birth rate among Hutterite couples. Thus, the identified mutations may represent targets of positive selection impacting reproductive compatibility between human mates.

84

"Fishing" for vertebrate fertilization genes: proteomic and biochemical characterization of rapidly evolving threespine stickleback egg proteins

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

1. Genome Sciences, University of Washington, Seattle, Washington, United States

2. Divisions of Basic Sciences and Human Biology, Fred Hutchinson Cancer Research Center, Seattle, Washington, United States

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 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.

85

Down-regulation of EPAS1 transcription explains genetic adaptation of Tibetans to high-altitude hypoxia

Yi Peng[1], Chaoying Cui[2], Yaoxi He[1], Hui Zhang[1], Deying Yang[1], Qu Zhang[1], Bianbazhuoma NA[3], Lixin Yang[1], Yibo He[1], Kun Xiang[1], Xiaoming Zhang[1], Sushil Bhandari[1], Peng Shi[1], Yangla NA[2], Lan Liu[2], Ouzhuluobu NA[2], Baimakangzhuo NA[2], Gonggalanzi NA[2], Caijuan Bai[2], Bianba NA[2], Basang NA[4], Ciwangsangbu NA[4], Tianyi Wu[5], Shuhua Xu[6], Hua Chen[7], Xuebin Qi[1], Bing Su[1]

1. Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, YUNNAN, China

2. School of Medicine, Tibetan University, Lhasa, China

3. The Municipal People's Hospital of Lhasa, Lhasa, China

4. People's Hospital of Dangxiong County, Dangxiong, China

5. High Altitude Medical Research Institute, Xining, China

6. CAS-MPG Partner Institute for Computational Biology, Shanghai, China

7. Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing, China

Tibetans are well adapted to the hypoxic environments at high altitude. Although previous genome-wide scans have identified a large number of candidate genes harboring Tibetan-enriched sequence variants, the molecular mechanism of how these variants lead to the adaptive physiological changes in Tibetans remains largely unclear. Here we report systematic genetic and functional dissections of EPAS1, a gene encoding hypoxia inducible factor 2a (HIF-2a) with the strongest signal of selective sweep in Tibetans. We show that in Tibetan umbilical endothelial cells and placenta tissues, the Tibetan-enriched EPAS1 variants cause a down-regulation of its expression. In the mouse model, the heterozygous EPAS1 knockout mice (~50% expression reduction) perform better than the wild-type controls during prolonged hypoxic treatment, and display blunted physiological responses to chronic hypoxia, mirroring the situation in Tibetans. Furthermore, we conducted a survey of multiple physiological traits and genetic association analysis among 508 Tibetans living at high altitude (4,700m), and we found that the EPAS1 adaptive variants account for their relatively low hemoglobin levels as a protection from polycythemia, and these variants also contribute to the low pulmonary vasoconstriction response in Tibetans. Collectively, we demonstrate that the Tibetan-enriched EPAS1 variants down-regulate its expression, serving as the molecular basis of adaption to high altitude hypoxia in Tibetans.

86

Extreme distribution of deleterious variation in a historically small and isolated population – insights from the Greenlandic Inuit

Casper-Emil T. Pedersen[1], Kirk E. Lohmueller[2], Niels Grarup[3], Peter Bjerregaard[4], Torben Hansen[3], Hans R. Siegismund[1], Ida Moltke[1], Anders Albrechtsen[1]

1. Section for Computation and RNA biology (SCARB), Department of Biology, University of Copenhagen, Copenhagen, Denmark

2. Department of Ecology and Evolutionary Biology, University of California, Los Angeles, United States

3. The Novo Nordisk Foundation Center for Basic Metabolic Research, Section of Metabolic Genetics, University of Copenhagen, Copenhagen, Denmark

4. National Institute of Public Health, University of Southern Denmark, Copenhagen, Denmark

All non-African human populations have experienced bottlenecks following the expansion of modern humans out of Africa some 60,000 years ago. Recently, several studies of human populations have investigated whether smaller population sizes for out-of-Africa populations lead to less efficient purifying selection and thereby higher genetic load. Given that these studies mainly considered large continental human populations we examine these questions using exome data from 18 Greenlandic Inuit. The Greenlandic Inuit population is particularly interesting to study in this context because it has experienced a ~20,000 year long bottleneck during the last ~25,000 years, making it more extreme than most previously studied populations, such as native Americans, in terms of population size.

When comparing it to a European population, we do not observe a difference in the overall number of deleterious alleles per individual, implying a similar genetic load assuming an additive model. However, we observe a marked difference in the distribution of this load; the Greenlandic Inuit population has fewer variable sites, and thus on average each variable site has a higher load. Also, each variable site has a higher average derived allele frequency. Consequently, the Greenlandic Inuit carry more homozygous derived genotypes and a higher genetic load assuming a recessive model. Despite the long recent bottleneck, we find that selection has still been acting however, it has acted less efficiently.

Our analyses show that the Greenlandic Inuit population has great potential for mapping of disease-causing variants that are rare, and thus difficult to map, in Europeans and other large populations – for both Mendelian and complex diseases. To a certain degree, this characteristic has also been documented for other small populations, yet comparative results for several small populations establish the Greenlandic Inuit as the population with the highest potential for finding novel disease-causing variants.

87

The African roots of Mexico: Genetic structure and health of Mexicans of African descent

María C Ávila-Arcos[1, 2], Andres Moreno-Estrada[3], Karla Sandoval-Mendoza[3], Marcus W Feldman[4], Carlos D Bustamante[2, 5], Christopher Gignoux[6]

1. International Laboratory for Human Genome Research, National Autonomous University of Mexico, Querétaro, Mexico

2. Department of Genetics, Stanford University, Stanford, CA, USA

3. Laboratorio Nacional de Genómica para la Biodiversidad, Centro de Investigación y Estudios Avanzados, Instiuto Politécnico Nacional, Irapuato, Guanajuato, Mexico

4. Department of Biological Sciences, Stanford University, Stanford, CA, USA

5. Department of Biomedical Data Science, Stanford University, Stanford, CA, USA

6. Stanford University, Stanford

Between the 16th and 19th centuries, over 12 million people were kidnapped mainly in West and West Central Africa and transported to the Americas during the trans Atlantic slave trade. Despite having received 200,000 Africans during the slave trade, no study in Mexico has focused on the characterization of the African genetic ancestry of its Afro-descendant population. In this study we worked together with Afro-Mexican communities to characterize their genetic ancestry using dense genome-wide genotyping. The dataset consists of 380 self-identified Afro-descendants, indigenous and mestizo population from three Mexican states. Additionally, we collected genealogical, self-identification and phenotype data such as skin pigmentation, height, weight, hip to waist ratio and hemoglobin. Genotype data revealed that up to 46% of the genetic ancestry of some individuals is of African origin, with the remaining being mainly of indigenous origin. We are currently exploring local ancestry and admixture patterns in this population as well as correlations between the genetic ancestry, self-identification and phenotypes; and have identified trends of public health relevance. For example, we observed significantly higher prevalence of overweight and obesity among women from Afro-Mexican communities, when compared to the national mean and to neighboring indigenous and mestizo populations. Interestingly, we have also noticed that individuals who self-identify using labels that reflect the color of their skin (e.g. negro, moreno) tend to have higher levels of African ancestry than individuals who use the term “Afro” (e.g. Afro-Mexican, Afro-descendant, etc) for self-identification. Lastly, since Afro-Mexicans currently suffer from poverty, discrimination, lack of recognition as a vulnerable minority, and limited access to health services, this study contributes to their appreciation as part of Mexico’s mosaic of diversity and will hopefully set the stage for health interventions to abate the obesity epidemic in the sampled regions.

88

Dynamic Genetic Control of Gene Expression and DNA Methylation in Human Aging

Trevor Martin[1], Hunter Fraser[1]

1. Stanford University, Stanford, CA, United States

Gene mRNA expression, DNA methylation, and microRNA expression vary dramatically during the human lifespan, but the vast majority of studies mapping genetic variants affecting these traits have either ignored age, or treated it as a confounding variable. Given that age is the main risk factor for many diseases, the identification of genetic loci that affect how these traits change with age could shed light on mechanisms for disease pathogenesis. Performing a genome-wide analysis of age-by-genotype interactions across over 300 individuals, we identify and cross validate over 100 genomic loci that control gene expression, DNA methylation, and microRNA expression in the aging human brain. Furthermore, we applied our analysis across 44 human tissues and found striking enrichment of temporal genetic regulation in the brain. Overlapping our temporal control loci with genome-wide association study results we also see a striking enrichment for SNPs involved in neurological diseases including Alzheimer’s disease, schizophrenia, and Parkinson’s disease. Together, these results indicate that genetic loci that exhibit temporal control of mRNA, DNA methylation, and microRNA over human aging are an important class of genetic variants that may shed light on the pathogenesis of age related human disease. More broadly, this study emphasizes the importance of gene-by-environment interactions in the understanding of how natural genetic variation affects gene expression in the human brain.

89

Diagnostic sequence motifs distinguish all point mutations in humans

Yicheng ZHU[1], Von Bing Yap[2], Gavin Huttley[1]

1. Australian National University, Acton, ACT, Australia

2. Department of Statistics and Applied Probability, National University of Singapore, Singapore

Understanding the factors influencing mutation can improve mutation detection techniques, identify diagnostic signatures of disease-causing mutagens, and facilitate the development of more accurate models of genetic divergence. Hypermutability of CpG demonstrates the existence of mutation motifs, sequences of flanking bases that influence point mutation processes. These motifs can thus be indicative of specific mutation mechanisms. Here, we report novel log-linear models for identifying mutation motifs that further allows comparisons of these, and of the complete mutation spectra, between samples. Mutation motifs are visualised using a sequence logo type method.

We applied the methods to examination of each of the possible 12 point mutations in ~13.6 million human germline mutations (inferred from SNPs recorded in ENSEMBL) and ~181 thousand melanoma mutations from the COSMIC database.

Our method recovered the well known CpG effect which a conventional motif detection method failed to do. We establish that all point mutations have significant and distinct mutation motifs. While the major effects of flanking bases lie within 2bp of the mutated position, we refute previous reports that the effect magnitude decays monotonically with distance. Comparison between autosomes and X-chromosome supported a reduced contribution from methylation induced C→T mutation on the X-chromosome, consistent with a previous prediction.

Analyses of malignant melanoma confirmed reported characteristic features of this cancer. This included strand asymmetry of mutation processes and that neighbouring influences in malignant melanoma differ significantly from those affecting germline mutations. Interestingly, the CpG effect was largely subsumed by different neighbouring mechanisms.

The statistical methods we report can be used to examine the role of flanking sequence on mutation processes from polymorphism data. They further enable identifying differences in the operation of mechanisms of mutation between genomic regions, cell types or species. Our results have important implications for modelling context-dependent effects on sequence evolution.

90

A novel approach for assessing genetic burden and constraint in protein structures leveraging large sequencing cohorts: insights from MYH7 and hypertrophic cardiomyopathy

Julian R Homburger[1], Eric M Green[2], Colleen Caleshu[1], Margaret S Sunitha[3], Rebecca Taylor[1], Kathleen M Ruppel[1], Raghu Metpally[4], Steven D Colan[5], Michelle Michels[6], Sharlene Day[7], Iacopo Olivotto[8], Carlos D Bustamante[1], Frederick Dewey[9], Carolyn Y Ho[10], James A Spudich[1], Euan A Ashley[1]

1. Stanford University, Stanford, CA, United States
2. Myokardia, Inc, South San Francisco, CA, USA
3. Institute for Stem Cell Biology and Regenerative Medicine, Bangalore, India
4. Geisinger Health System, Danville, PA, USA
5. Boston Children's Hospital, Boston
6. Erasmus MC, Rotterdam
7. University of Michigan, Ann Arbor
8. Careggi University Hospital, Florence, Italy
9. Regeneron, Inc., Tarrytown, NY, USA
10. Brigham and Women's Hospital, Boston, MA, USA

Myosin motors are the fundamental force-generating elements of muscle contraction. Variation in the β-cardiac myosin gene (*MYH7*) can lead to hypertrophic cardiomyopathy (HCM), a heritable disease characterized by cardiac hypertrophy, heart failure, and sudden cardiac death. A key debate is whether there exist hotspots of pathogenic variation within the myosin structure. Previous studies have reported conflicting results and suffered from small sample sizes and lack of reference cohorts. Furthermore, how specific myosin variants alter motor function or clinical expression of disease remains incompletely understood. To address these questions, we developed a statistical method for analyzing disease burden and constraint in three-dimensional protein structures and surfaces that we apply to β-cardiac myosin. We combine structural models of myosin from multiple stages of its chemomechanical cycle, exome sequencing data from two population cohorts of 60,706 and 42,930 individuals, and genetic and phenotypic data from 2,913 HCM patients to identify regions of disease-variant enrichment within β-cardiac myosin. We develop computational models of the human β-cardiac myosin protein structure before and after the myosin power stroke. Then, using a spatial scan statistic modified to analyze genetic variation in protein three-dimensional space, we show a significant enrichment of disease-associated variants in the converter (p=0.002), a kinetic domain that transduces force from the catalytic domain to the lever arm during the power stroke. Focusing our analysis on surface-exposed residues, we identified a larger region significantly enriched for disease-associated variants that contains both the converter domain and residues on a single flat surface on the myosin head described as the myosin mesa (p=0.002). Notably, HCM patients with variants in the enriched regions have earlier disease onset than those with variants elsewhere. Our study provides a model for integrating protein structure, large-scale genetic sequencing and detailed phenotypic data to reveal insight into time-shifted protein structures and genetic disease.

91

Gene ORGANizer: Linking Genes to the Organs They Affect

David Gokhman[1], Guy Kelman[1], Adir Amartely[1], Guy Gershon[1], Shira Tsur[1], Liran Carmel[1]

1. The Hebrew University, Jerusalem, Israel

One of the biggest challenges in the era of genomics is linking genes to phenotypes and specifically to the body parts they affect. Many tools have tried to address this through indirect approaches using features such as expression levels, biochemical pathways and molecular function. Here, we present Gene ORGANizer (geneorganizer.huji.ac.il), a tool we have developed to directly link genes to organs. Gene ORGANizer allows researchers to analyze the anatomical effects of genes, and to understand the shared impact of groups of genes. We used more than 100,000 gene-phenotype associations to build the Gene ORGANizer database, which contains links of more than 7,000 genes to ~150 organs, body systems and anatomical regions. Using the tool, we showed that the most substantial regulatory changes in recent human evolution happened in the vocal tract. This trend is unique to modern humans and arose after the split from the Neanderthal and the Denisovan. We also confirmed previous hypotheses that chromosome X is significantly enriched with genes affecting the brain and the reproductive system. Surprisingly, we found an even more pronounced, yet previously unreported, pattern: chromosome X is significantly enriched with genes that affect facial features. These results suggest that chromosome X experiences a unique selection regime, where genes that affect some body parts are preferentially represented, while others are selected against. We expect Gene ORGANizer to be useful in a broad range of evolutionary, medical, and molecular studies, with applications in any analysis aimed at characterizing phenotypic consequences of genetic changes.

92

Probabilistic inference of positive and negative selection in cancer

Donate Weghorn[1], Matteo D'Antonio[2], Kelly Frazer[2], Shamil Sunyaev[1]

1. Division of Genetics, Brigham and Women's Hospital, Harvard Medical School, Cambridge, MA, United States
2. Division of Genome Information Sciences, University of California, San Diego, La Jolla, CA, United States

Seen from an evolutionary perspective, cancer is a highly complex system that evolves asexually under high mutation rates and strong selective pressures. Recurrence of mutations or their marked absence attest the selection acting on a given sequence, but knowing the proper mutational null model of how many mutations to expect is highly nontrivial. Here we present a probabilistic approach to addressing this question, which we apply to 17 cancer types. Using an empirical Bayes framework, we infer the distribution of mutation rate across genes that underlies the observed distribution of the synonymous mutation count within a given cancer type. This enables an inference of the posterior probability of nonsynonymous mutations under neutrality without additional parameters, however explicitly taking into account cancer type-specific mutational signatures, which are known to be highly distinct. We find substantial overlap of our predicted set of significantly positively selected genes with known cancer genes. In addition, we use our model and the large patient cohort to quantify negative selection. While the genome-wide average signal of negative selection is largely weak, we find marginally significant cancer type-specific sets of candidate genes. Moving from coding to non-coding sequence, we applied a similar approach to the detection of hypermutation in breast cancer DNAse hypersensitivity sites, an indicator for positive regulatory selection. Here, the background mutation rate is inferred from clustering according to known mutation rate covariates, and it again informs the expected number of mutations under neutrality. We find 22 putative breast cancer driver DHSs, three of which are significantly hypermutated across 19 cancer types. We further validated one of these DHSs experimentally and one based on expression data. Taken together, these applications show the impact of probabilistic modeling of mutation events to unveil the various signals of selection in cancer, which may inform targets of cancer therapy.

93

David Penny & Mike Hendy

David Penny[1], **Michael Hendy**[2]

1. Massey University, Palmerston North, MANAWATUE, New Zealand

2. Department of Mathematics and Statistics, University of Otago, NZ

David Penny and Mike Hendy will discuss their collaboration, particularly how combining mathematical rigor with biology has been able to provide new insights into evolutionary patterns and processes, and in turn solve long-standing questions. In a short video, Mike Steel will provide his unique perspective on Mike and David's contribution to evolutionary thought.

94

RNA evolution is dominated by rapid turnover, not molecular fossils.

Anthony Poole[1]

1. Biomolecular Interaction Centre, School of Biological Sciences, University of Canterbury, New Zealand

RNA has been argued to provide a window back to the earliest stages in the evolution of life on Earth. However, fewer than 1% of known RNA families are conserved across all three domains of life. For RNAs found only in a single domain, the picture is not much better: only 3.5% of RNA families show a distribution consistent with tracing back to the ancestor of the domain in which they are found. I will discuss two possible explanations for this result:

1. That most RNAs are evolutionarily young, and the result of ongoing de novo emergence of small RNAs from genomic noise.

2. That many more RNAs are old, but better data are needed to find these.

We have found that current data are insufficient to distinguish these two possibilities. Sequence conservation of RNA genes drops off precipitously quickly compared to protein-coding genes, and existing covariance-based search strategies therefore perform poorly on the skewed distribution of public genomic data - which is dominated by genomes of humans and their pathogens. We find there is a 'Goldilocks Zone' for comparative analysis of RNAs, where, for optimal identification of RNA genes, comparisons between genomes that are not too similar and not too distant yield rich information on noncoding RNAs. Unfortunately almost no transcriptomics data collected to date sit within the Goldilocks Zone, meaning we cannot gauge the age of most RNA genes. Moreover, we cannot detect these for the most studied lineages, as sampling is too narrow.

While we now know data are poor, I will nevertheless put a stake in the ground, and present our current thoughts on how RNA genes originate and evolve.

95

"The data are good but the models are bad!": examples from recent phylogenomic studies.

Frederic Delsuc[1]

1. Institut des Sciences de l'Evolution, CNRS - Université de Montpellier, Montpellier, France

As post-doc with David Penny back in 2003, "The data are good but the models are bad!" is one of the exclamations I most frequently heard from him. At the time, we were focusing on reconstructing mammalian phylogenetic relationships from complete mitochondrial genomes. The results obtained from mitogenomes were at odds with the ones obtained from nuclear genes. We showed that this incongruence mainly stemmed from heterogeneities in the mitogenomic data that were not accounted for by the simple models of sequence evolution available at the time. This led us to propose simple data reduction procedures such as RY-coding of nucleotide data to alleviate both substitutional and compositional biases. Far from being ideal, such an approach had the advantage to avoid phylogenetic reconstruction artifacts due to model misspecifications. With the advance in statistical modeling, more complex models now allow making sense of mitogenomic data. However, history repeated itself with the development of phylogenomics, as we are far from disposing of adequate models accounting for the full complexity of genomic data. Using examples from recent genome-scale studies, I will illustrate how David's statement still applies to phylogenomic inference.

96

Hadamard conjugation and phylogenetic inference.

Michael Charleston[1]

1. School of Physical Sciences, University of Tasmania, Hobart, TAS, Australia

Hendy, Penny and Steel made a startling discovery in the early nineties, when considering the relationship between branch lengths of a phylogenetic tree and the site patterns that could arise from sequences evolving on that tree.

Each site in an alignment can suggest a split or "bipartition" of the extant taxa of a phylogenetic tree into two sets, and the complete alignment constitutes a spectrum of such splits.
Correspondingly, each phylogenetic tree can be thought of as a collection of compatible splits.

So what is the relationship between the two?

Hendy et al. found an elegant way of converting between the relative frequencies of splits supported by an alignment, and branch lengths of the underlying tree.
This conversion uses a specialised family of Hadamard matrices (such matrices are all square and with entries that are all ±1). Moreover, this conversion, known as the Hadamard Conjugation, was possible even without choosing a particular tree – rather a remarkable feat – and so enabled researchers to get estimates of branch lengths without conducting a tree search.

Although the computational complexity of the conjugation is exponential in the number of taxa in the tree, this conjugation has led to several significant advances in phylogenetic inference.

This talk will sketch the Hadamard conjugation, and describe how it has influenced phylogenetic estimation over the last two decades.

97

Geomolecular Dating and the Origin of Placental Mammals

Matthew J Phillips[1]

1. School of Earth, Environmental and Biological Sciences, QUT, Brisbane, QLD, Australia

In modern evolutionary divergence analysis the role of geological information extends beyond providing a timescale, to informing molecular rate variation across the tree. Continuing research I began during my PhD with David Penny, I will use fossil calibrations to test the accuracy of models of molecular rate evolution for placental mammals, and reveal substantial misspecification associated with life history rate correlates. Adding further calibrations to reduce dating errors at specific nodes unfortunately transfers underlying rate errors to adjacent branches. Thus, to buffer against rate model errors we require tight calibration across the tree, including allowing maximum bounds to be tight when good fossil records permit. Otherwise, divergences deep in the tree tend to be inflated by the interaction of rate errors and asymmetric confidence in minimum and maximum bounds. For placental mammals the potential for transferring calibration and rate model errors across the tree is reduced by focusing on well-supported calibrations with appropriately conservative maximum bounds. The resulting divergence estimates are younger than others published recently, and provide the long-anticipated molecular signature for the placental mammal radiation observed in the fossil record near the 66 Ma Cretaceous-Paleogene extinction event.

98

From the genome to biodiversity: molecular evolutionary insights into macroevolution

Lindell Bromham[1]

1. Australian National University, ACT, Australia

[For the David Penny and Mike Hendy Symposium]

David Penny and Mike Hendy have made a remarkable contribution to the development of evolutionary biology by helping to develop not only the tools but also the insight needed to connect biomolecular change at the genomic level to microevolutionary change at the population level to the generation of biodiversity. These different levels of evolutionary change have commonly been studied by biologists from different disciplines, with geneticists and biochemists investigating mutation, population biologists and ecologists studying population level processes, and palaeontologists and systematists revealing patterns of biodiversity. But Penny and Hendy have contributed to the emergence of a new field whereby we can consider all of these levels of evolutionary change within a single framework, through the analysis of DNA sequences. Through molecular phylogenetic analysis, we can link change at the biochemical level to species characteristics and biodiversity generation, highlighting the continuity of processes of mutation (generation of variation), microevolution (population divergence) and macroevolution (lineage diversification).

99

Synergistic roles of rapid warming events and human impacts in the megafaunal extinctions

Alan Cooper[1], **Christian Turney**[2]

1. Australian Centre for Ancient DNA, University of Adelaide, Adelaide, SA, Australia

2. Climate Change Research Centre, University of New South Wales, Sydney, NSW, Australia

Improved resolution records from radiocarbon and ancient DNA studies of megafauna and human populations have allowed us to disentangle the roles of climate change and human impact in the Late Pleistocene megafaunal extinctions. We find that Holarctic megafaunal populations underwent repeated local or global extinctions in association with rapid warming events, known as interstadials, on a millennial scale. The extinction events took place both before and after the presence of modern humans on the landscape, and the metapopulation processes which appear to stabilize the ecosystem may have evolved to provide resilience to rapid and frequent climate shifts in the past.

In the Americas, we find that the rapid movement of the first Native Americans throughout both continents creates a powerful and unique model system due to the opposing climate trends in each hemisphere at the time. While megafaunal extinctions were associated with warming trends in both cases, the out of phase climate patterns caused the sequence and timing of events to be mirrored, providing an illuminating high-resolution view of the interactions of human colonization and climate change on megafaunal ecosystems.

100

The problems with distance-based phylogenetic relationships

David Bryant[1]

1. University of Otago, Dunedin, OTAGO, New Zealand

In a scientific correspondence to Nature in 1988, Steel, Hendy and Penny quantified the loss of information inherent in phylogenetic analyses based on genetic distances. In spite of that loss, distance-based methods for inferring both trees and networks have proven both useful and popular. Some even argue that they should work equally well for large data sets. Here we engage with the original debate, and explain how the loss of information in both distance and character data has stymied best efforts to construct effective recombination tests with NeighborNet or SpectroNet.

101

The archaeal ancestry of eukaryotes

Thijs J.G. Ettema[1]

1. Department of Cell and Molecular Biology, Uppsala University, Uppsala, Sweden

The origin of the eukaryotic cell represents an enigmatic evolutionary puzzle. Ever since the discovery of the archaeal domain of life by Carl Woese and co-workers, Archaea have featured prominently in hypotheses for the origin of eukaryotes. According to Woese's 'universal tree', eukaryotes and Archaea represent sister lineages, suggesting that Archaea and Eukarya emerged from a common ancestor. Whereas this classical 'Three domains' scenario has received considerable support in the past decades, recent studies have provided a growing support for cellular fusion scenarios in which eukaryotes emerged from within the archaeal domain of life. More specifically, the latest advanced phylogenomic analyses have indicated that eukaryotes form a monophyletic clade with the Lokiarchaeota, a group of archaea that was recently discovered in deep marine

hydrothermal sediments. Analysis of the first reconstructed Lokiarchaeota genome revealed that it encoded numerous 'eukaryotic signature proteins' (ESPs), several of which are indicative of sophisticated membrane remodeling capabilities (actin, ESCRT proteins, small GTPases).

In our ongoing effort to gain insight in the archaeal origin of eukaryotes, we have now obtained metagenome data from numerous sites worldwide in order to obtain genomic data of uncultivated archaea. We have reconstructed the genomes of several novel archaeal lineages that are related to the Lokiarchaeota, some of which only distantly. Phylogenomic analyses indicate that these new archaeal lineages are united in a putative new superphylum that also includes eukaryotes. Detailed analysis of these new archaeal genomes revealed, in addition to ESPs previously identified in Lokiarchaeota, the presence of many new eukaryotic features. Altogether, the newly acquired data provide new, important insights in the enigmatic origin of eukaryotes.

102

Natural history of eukaryotic viruses and transposons: What do they tell us about eukaryogenesis?

Eugene V Koonin[1]

1. National Institutes of Health, USA, Bethesda, MARYLAND, United States

Viruses are the most common, abundant and diverse biological entities. Recent efforts in comparative genomics and in particular joint analysis of the genomes of viruses and non-viral selfish elements, such as transposons, have led to breakthroughs in our understanding of the evolution of these elements, in particular, in eukaryotes. A combination of phylogenomics with the quantitative analysis of bipartite gene-genome networks enables us to elucidate, albeit with different degrees of confidence, the origins of all major groups of viruses of eukaryotes from prokaryotic ancestors. In particular, a major class of dsDNA viruses centered around Polintoviruses (Polintons), self-replicating elements that lead a dual lifestyle as viruses and transposons. This class encompasses an extreme diversity of eukaryotic viruses including the giant viruses and their virophages as well as several groups of bacterial and archaeal viruses with icosahedral capsids and protein-primed replication. The Polintoviruses appear to be direct descendants of prokaryotic tectiviruses, thus tracing the origin of the majority of eukaryotic dsDNA viruses and transposons to a specific prokaryotic root. Comparison of the representation of different classes of viruses among prokaryotes and eukaryotes reveals an apparent paradox. Eukaryotes are hosts to a huge diversity of RNA viruses which are quite rare in prokaryotes. In contrast, the prokaryotic virosphere is dominated by dsDNA viruses. At first glance, this unexpected distribution of viruses might imply the emergence of the eukaryotic virosphere directly from the primordial RNA world. However, a detailed dissection of eukaryotic RNA viral genomes reveals a more complex picture that is best compatible with the assembly of these viruses from different prokaryotic building blocks including reverse transriptases of retroelements that could give rise to viral RN-dependent RNA polymerases. Thus, there seems to be a discontinuity between the primordial and modern RNA worlds. This scenario is compatible with the synbiogenetic model of eukaryogenesis.

103

Endosymbiotic origin and differential loss of eukaryotic genes

Chuan Ku[1], **Shijulal Nelson-Sathi**[1], **Mayo Roettger**[1], **Filipa Sousa**[1], **Peter Lockhart**[2], **David Bryant**[3], **Einat Hazkani-Covo**[4], **James McInerney**[5], **Giddy Landan**[6], **William Martin**[1]

1. Heinrich-Heine University, Düsseldorf, Germany

2. Massey University, Palmerston North, New Zealand

3. University of Otago, Dunedin, New Zealand

4. The Open University of Israel, Ra'anana, Israel

5. The University of Manchester, Manchester, UK

6. Christian-Albrechts-University of Kiel, Kiel, Germany

Chloroplasts arose from cyanobacteria, mitochondria arose from proteobacteria. Both organelles have conserved their prokaryotic biochemistry, but their genomes are reduced, and most organelle proteins are encoded in the nucleus. Endosymbiotic theory posits that bacterial genes in eukaryotic genomes entered the eukaryotic lineage via organelle ancestors. It predicts episodic influx of prokaryotic genes into the eukaryotic lineage, with acquisition corresponding to endosymbiotic events. Eukaryotic genome sequences, however, increasingly implicate lateral gene transfer, both from prokaryotes to eukaryotes and among eukaryotes, as a source of gene content variation in eukaryotic genomes, which predicts continuous, lineage-specific acquisition of prokaryotic genes in divergent eukaryotic groups. Here we discriminate between these two alternatives by sampling protein sequences from a wide range of phylogenetically diverse eukaryotes and by clustering and phylogenetic analysis of eukaryotic gene families having prokaryotic homologues. Our results indicate (1) that gene transfer from bacteria to eukaryotes is episodic, as revealed by gene distributions, and coincides with major evolutionary transitions at the origin of chloroplasts and mitochondria; (2) that gene inheritance in eukaryotes is vertical, as revealed by extensive topological comparison, sparse gene distributions stemming from differential loss; and (3) that continuous, lineage-specific lateral gene transfer, although it sometimes occurs, does not contribute to long-term gene content evolution in eukaryotic genomes.

104

Mitochondrial origin of eukaryotic membrane complexity

Sven B Gould[1], **Sriram G Garg**[1], **William F Martin**[1]

1. Heinrich Heine University, Düsseldorf, NRW, Germany

It is evident that in the 4 billion years of evolution no prokaryote evolved a cellular complexity that even vaguely resembles that of a eukaryote, other than the archaeal cell that acquired the mitochondrion through endosymbiosis. Thoughts on the origin of the endomembrane system are linked to thoughts on the origin of eukaryotes themselves. Current models traditionally derive the origin of the endomembrane system from inward invaginations of the plasma membrane, such that the endoplasmic reticulum (ER) lumen is topologically homologous to the environment. But prokaryotic vesicle flux is outward not inward and such models also fail in explaining the transition from an archaeal plasma membrane based on isoprene ethers to a bacterial-type membrane based on fatty acid esters. We propose that the endomembrane system stems from outer membrane vesicles (OMVs) the mitochondrial endosymbiont secreted. In terms of the number and nature of evolutionary innovations required to evolve a basic endomembrane system our minimal premises can hardly be underbid, and account for: (i) the transitional mechanism that converted the composition of eukaryotic membranes from archaeal to bacterial lipids, (ii) the finding that eukaryotic lipid synthesis occurs predominantly at the ER and mitochondria, (iii) the circumstance that eukaryotes store Ca^{2+} in the ER, (iv) the formation of the nucleus from the ER, not vice versa, (v) the archaeal ancestry, localisation, and orientation of the eukaryotic V-ATPase in food vacuoles and other things we will discuss. From our proposal, a natural evolutionary order in the origin of several key characters of eukaryotic cells unfolds in that, during eukaryogenesis, the ER represented the

first autogenous (nonendosymbiotic) cell compartment, formed from OMVs secreted by the mitochondrion. Endosymbiotic theory directly accounts for the origin of eukaryote complexity and it required a 'garden variety' archaeal host to do so, as some theories for eukaryote origin predicted.

105

The early origin of the eukaryotic cell according to the Viral Eukaryogenesis hypothesis: an update in light of recent discoveries

Philip J L Bell[1]

1. Microbiogen, Lane Cove, NSW, Australia

Acknowledged as one of the most difficult challenges in evolutionary biology, the origin of the eukaryotic cell and the related origin of sex remain unresolved. The Viral Eukaryogenesis hypothesis proposes the eukaryotes and sex arose because the 'eukaryotic cell' is a symbiogenic consortium of three separate organisms: an archaeal ancestor of the eukaryotic cytoplasm, a bacterial ancestor of the mitochondrion, and a viral ancestor of the nucleus. One radical aspect of this theory is the descent of the nucleus from a virus. When first published 15 years ago, a pox-like virus was chosen as a nuclear ancestor because they possessed large linear chromosomes, they could replicate in their host's cytoplasm, and they possessed endogenous mRNA capping genes, a classic eukaryotic signature gene. Since 2001 science has progressed significantly and giant viruses such as the Mimiviruses and other members of the Megavirales (aka NCLDV viruses) have been discovered. Like the pox virus and the eukaryotic nucleus, a signature gene of this group are the mRNA capping genes. The discovery of giant viruses has radically altered the view of viruses as simple small 'filterable' entities and with other discoveries has strengthened the possibility that viruses play critical roles in evolution including involvement in eukaryotic origins. Furthermore, several new phyla of the Archaea have been discovered since 2001, including the Lokiarchaeota which form a monophyletic group with the eukaryotes. In this talk the Viral Eukaryogenesis hypothesis is reviewed in light of these recent discoveries, and it is argued that the profound scientific discoveries over the last 15 years have lent increased plausibility to the Viral Eukaryogenesis hypothesis.

106

Origin of the *Andalucia godoyi* mitochondrial cox15 gene by horizontal transfer from bacteria

Ding He[1], **Chengie Fu**[1], **Sandra L Baldauf**[1]

1. Uppsala University, Uppsala, Sweden

Members of the phylum Jakobida (superkingom Excavata) are known for their strikingly bacterial-like mitochondrial genomes (mtDNA). These mtDNAs have as many as 67 protein-coding genes, some of which are organised into operons with nearly canonical alpha-proteobacterial gene order. Jakobid mtDNAs also encode a bacterial-type RNA polymerase, which they use for mtDNA transcription rather than the viral type (T7-like) RNA polymerase used by all other eukaryote mtDNAs. This has led to the suggestion that jakobids are unique among eukaryotes in retaining primitive mtDNAs that trace to the mitochondrial progenitor. However, molecular phylogeny fails to place jakobids as sister to all other eukaryotes. Furthermore, we show that at least one of the uniquely bacterial-like mtDNA genes of the jakobid *Andalucia godoyi* is not in fact primitive, but rather appears to have been derived relatively recently by horizontal transfer from bacteria. These data suggest an alternative explanation for the bacterial-like nature of jakobid mtDNAs.

1. Burger G, Gray MW, Forget L, Lang BF. (2013) Strikingly bacteria-like and gene-rich mitochondrial genomes throughout jakobid protists. Genome Biol. Evol. 5:418-38.
2. He D, Fu CJ, Baldauf SL. (2016) Multiple Origins of Eukaryotic cox15 Suggest Horizontal Gene Transfer from Bacteria to Jakobid Mitochondrial DNA. Mol. Biol. Evol. 33:122-33.

107

Do retroCNVs resolve intralocus sexually antagonistic conflicts in *Drosophila*?

Esther Betran[1], **Anna Williford**[1]

1. University of Texas at Arlington, Arlington, TEXAS, United States

Retroposed copies of new genes (also known as retrogenes) are copies of genes that are produced from an mRNA intermediate. It has been observed that retrogenes often acquire testis-specific expression in a variety of lineages including mammals and dipterans, possibly as a consequence of new genes being inserted close to testis-expressed genes. In addition, in *Drosophila*, it has been shown that testis-expressed retrogenes often evolve under positive selection, show pattern of recurrent duplication in different lineages and exhibit an overrepresentation of particular gene ontologies, including nuclear-encoded mitochondria genes, nuclear transport and proteasome functions. We have proposed a model that postulates the existence of intralocus sexually antagonistic variation (i.e., allelic variation conferring opposite fitness effects for males and females due to the effects in male germline) in the parental gene, followed by gene duplication that resolves this antagonism. While DNA-mediated duplications could also resolve this antagonism and have the same effect (e.g., we do see a high number of testis-specific nuclear-encoded mitochondria duplications that are DNA mediated), the high frequency of retrogenes expressed in testis should facilitate this outcome. Because of this, we are undertaking an extensive search for retroCNVs using available genomic data, including genomic localization of retroCNVs, analysis of parent-daughter expression patterns and examination of the allelic variation in the parental genes. Candidate retroCNVs that show features consistent with the model will be used to study sex-specific fitness effects to explore their role in resolution of sexual antagonism. We will provide an update on those efforts.

108

Accumulation of structural mutations in an invasive fish of hybrid origin

Stefan Dennenmoser[1], **Fritz Sedlazek**[2], **Elzbieta Iwaszkiewicz**[1], **Xiang-Yi Li**[3], **Janine Altmüller**[4], **Arne Nolte**[1]

1. Max-Planck Institute for Evolutionary Biology, Ploen, SH, Germany

2. Simons Center for Quantitative Biology, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, USA

3. Department of Evolutionary Biology and Environmental Studies, , University of Zurich, Zurich, Switzerland

4. Institute of Human Genetics, University of Cologne, Cologne, Germany

Mixing of genetic material from previously separated gene pools may cause evolutionary change and hybrid speciation. This has been mostly explained through interactions of existing genetic traits. In contrast, evolutionary novelty through new mutations after admixture is less explored. Herein, we tested the hypothesis that structural mutations accumulate in invasive *Cottus*, an evolutionary young fish of hybrid origin, relative to both parental species *Cottus rhenanus* and *Cottus perifretum*. Among 10,979 genes screened for CNVs using comparative hybridization arrays, 25 genes showed significant changes in copy number in a natural population of invasive *Cottus*. Three genes with the most pronounced increase in copy number were previously found to be upregulated in invasive *Cottus*, suggesting potentially adaptive gene dosage effects. Transposable element copy number was estimated by mapping whole-genome sequencing reads against a de novo genome assembly of transposable elements. We found a significant copy number increase in 20.7% of all putative transposable elements in invasive *Cottus*, compared to a very rare decrease (0.01%) of TE copies relative to the parent species. The possibility of rapid evolution of de novo copy number variants was tested in laboratory crosses between the parental species. F2 individuals of two families (*C. rhenanus* x *C. perifretum*) were screened for de novo copy number changes of three candidate genes and two transposons using digital droplet PCR. Although our results do not support high rates new structural mutations they nonetheless show that structural mutations accumulate after admixture and that copy number increases of protein coding genes can manifest phenotypically in an evolutionary young hybrid species. This supports the idea that structural variants contribute to rapid evolution of admixed lineages.

109

Non-allelic gene conversion is ten times faster than the rate of point mutations in humans

Arbel Harpak[1], Xun Lan[2], Jonathan K Pritchard[2, 1, 3]

1. Department of Biology, Stanford University, Stanford, CA, United States

2. Department of Genetics, Stanford University, Stanford, CA, United States

3. Howard Hughes Medical Institute,

Gene conversion is the unidirectional transfer of genetic sequence from a “donor” region to an “acceptor”. In one of its modes, non-allelic gene conversion (NAGC, also known as interlocus gene conversion), the donor and the acceptor are homologous sequences on the same chromatid. Despite the implication of NAGC as the cause of various genetic diseases, and its role in the concerted evolution of many human gene families, the rates and contributing factors of NAGC are not well-characterized. Recent gene duplications are of focal interest in studying NAGC as NAGC is contingent on high sequence similarity between donor and acceptor. Notably, NAGC events are expected to distort the genealogy of a gene family at an affected region. Here, we develop a machine learning tools to survey duplicate gene families across primates in search of such local genealogy distortions, and identify converted regions in 44% of duplicate gene families surveyed. In addition, we estimate the parameters governing NAGC in humans. We estimate a tenfold higher rate of NAGC than point mutations in humans, with a median NAGC tract length of 525bp. Finally, we quantify the effects of genomic features which determine NAGC rates, including GC content, methylation levels and homology between donor and acceptor sequences. This work improves our understanding of the mechanisms behind NAGC and of the role NAGC plays in shaping sequence evolution in humans.

110

Characterizing complex structural variation at the human glycophorin receptors for Plasmodium falciparum

Ellen M Leffler[1, 2], Gavin Band[1], Kirk A Rockett[1, 2], Quang Si Le[1], Dominic P Kwiatkowski[1, 2], Chris CA Spencer[1], Malaria Genomic Epidemiology Network[1]

1. University of Oxford, Oxford, Oxfordshire, United Kingdom

2. Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire, United Kingdom

The human glycophorin gene cluster on chromosome 4 encodes two red blood cell surface receptors for *Plasmodium falciparum,* and variation in these genes also underlies the diverse MNS blood group system. Nearby signals have recently been identified in genome-wide scans for both severe malaria susceptibility and ancient balancing selection, but further insight into underlying functional mechanisms is hindered by the homology between the glycophorin genes. The high level of sequence similarity makes read mapping and variant calling problematic, yet also predisposes the region to frequent non-allelic homologous recombination that leads to structural variation. To identify such variants, we developed a hidden Markov model that is robust to this homology to infer copy number changes from sequence coverage data. We apply the method to worldwide sequence data from 2596 individuals, including 1046 from sub-Saharan Africa. This analysis reveals multiple large deletions and duplications, corresponding both to known blood groups and novel variation. In addition to whole gene duplications and deletions, several are predicted to carry hybrid glycophorin genes. We further consider the possible functional consequences at the red blood cell surface and potential for genetic interactions with variation in the corresponding parasite binding partners.

111

Structural variation at the TEP anti-pathogen locus in African malaria mosquitoes

Eric Smith[1], Bradley White[1]

1. University of California - Riverside, Riverside, CALIFORNIA, United States

Several species of mosquito in the *Anopheles gambiae* species complex are highly efficient vectors of malaria in sub-Saharan Africa. Two of these species, *An. coluzii* and *An. gambiae*, have only recently been elevated to the level of separate species, and were previously thought to be separate forms of the same species. Genome scans have revealed a region of high divergence on the left arm of the third chromosome (3L), containing at least 6 genes of the thioester-containing protein (TEP) family, which have been implicated in the innate immune response. Interestingly, *An. coluzzii* populations from Mali and Burkina Faso are fixed for an allele at *TEP1* – r^B, not present in sympatric *An. gambiae* populations – that confers additional resistance to the malaria parasite. To annotate the TEP Anti-Pathogen Locus (TAPL) and identify structural variation among the three known haplotypes – r^B, r^A, and S – in the TAPL region, we have sequenced BAC clones and generated consensus sequences of each of the three haplotypes. Sequence alignments indicate several significant structural differences between the three haplotypes, including a duplication in one haplotype that has created a novel chimeric *TEP* gene. *TEP1* had previously been shown to be the product of a similar chimerization, suggesting that this may be an under-appreciated mechanism of gene creation and genome evolution in *Anopheles* mosquitoes.

112

A second look at the chimpanzee genome

Lukas FK Kuderna[1], Tomas Marques-Bonet[1, 2, 3]

1. IBE (CSIC-UPF), Barcelona, CATALUNYA, Spain

2. Centro Nacional de Analisis Genomico (CNAG), Barcelona, Spain

3. Institucio Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain

The initial sequencing of the chimpanzee genome and its comparison to our own constituted a major leap in the understanding of the evolutionary trajectories of our species. Nevertheless, the proper characterization of some complex regions are far from complete as even the most current iteration of the chimpanzee assembly (Pantro_2.1.4) is still a draft a decade after its initial release. In this work, we will present a complete de-novo assembly for the common Chimpanzee, incorporating several different novel technologies such as Illumina 2x250 bp overlapping paired end reads, PacBio single molecule sequencing long reads, Illumina Moleculo and Dovetail HiC-chromosomal contact maps for scaffolding. We produce a high quality assembly derived from a single male individual (Clint), boosting contiguity by over 300% on the scaffold level, and by over 750% on the contig level, additionally resolving around 230Mbps of the genome. We are able to characterize more complexity in this version of the chimpanzee genome and explore the genome wide diversity of tandem repeats and copy number variation at population level

114

Identifying genetic and environmental factors causing developmental defects in humans and mice

Sally L Dunwoodie[1]

1. Victor Chang Cardiac Research Institute, Darlinghirst, NSW, Australia

Birth defects are present in 3-6% of live born humans and in greater numbers in those that die during gestation. Birth defects are the leading cause of infant death and are more prevalent than most chronic childhood diseases. The causes of birth defects are largely unknown with genetic and environmental factors, and a combination of these, proven and suspected to be the cause. We are identifying genetic and environmental factor that cause vertebral column and heart defects in human and mouse. We have evidence of gene-environment interaction in mouse causing developmental defects in mice.

115

An environmental DNA (eDNA) based method for monitoring spawning activity: a case study using the endangered Macquarie perch (*Macquaria australasica*)

Jonas Bylemans[1], Elise Furlan[1], Chris Hardy[2], Mark Lintermans[1], Dianne Gleeson[1]

1. University of Canberra, Canberra, ACT, Australia

2. CSIRO, Canberra

Determining the timing and location of reproductive events for threatened species can be critical for the design and evaluation of appropriate environmental management actions. Most methods currently available for monitoring reproduction in aquatic organisms are biased, costly, time intensive and often require lethal sampling. Here, we present an environmental DNA (eDNA) based methodology, which can overcome these constraints and has the potential to accurately determine spawning time and location. During spawning, the mass release of spermatozoa, which contain few mitochondria and highly protected nuclear DNA, is a major source of eDNA. Consequently, we hypothesized that the relative abundance of mitochondrial and nuclear eDNA will change during reproductive events. Using the nationally endangered Macquarie perch, we simulated spawning in experimental tanks and monitored the changes in mitochondrial and nuclear eDNA over time. Additionally, a small scale field survey was conducted in the upper Murrumbidgee River to validate the experimental results. The results of the experimental and field based studies revealed that both target fragments are equally abundant outside of the reproductive period. In contrast, samples collected after spawning was simulated/observed contained significantly more nuclear eDNA while mitochondrial eDNA concentrations increased more moderately. Hence, the ratio between both target fragments can be used to detect recent spawning activity.

116

From faeces to foxes: using genetics to manage an invasive predator for wildlife conservation

Anna J MacDonald[1], Aaron Aadamack[1], Catriona D Campbell[1], Elise F Dewar[2], Bernd Gruber[1], Elodie Modave[1], Sumaiya Quasim[1], Stephen D Sarre[1]

1. University of Canberra, Canberra, ACT, Australia

2. Department of Primary Industries, Parks, Water and Environment, Hobart, Tasmania, Australia

Genetic analysis of non-invasive samples is an important wildlife management tool, and is especially useful to inform management of rare or cryptic species. Here, we describe how genetic tools for species detection and individual identification can be applied within a management context, to find and monitor an invasive predator and understand its impacts on native wildlife.

The introduced red fox (*Vulpes vulpes*) has been implicated in extinctions and declines of many Australian vertebrates, and fox management remains a challenge to conservation across much of the continent. The island of Tasmania was considered to be fox-free, but from the late 1990s increasing evidence pointed to a fox incursion. Tasmania is home to species that have declined or become extinct elsewhere in Australia and are at particular risk of fox predation.

Since 2007, we have used genetics to study Australian foxes. We have conducted landscape-scale surveys for predator scats in Tasmania, and used a PCR and sequencing test to identify scats containing fox DNA. We have also used blind trials and bioinformatic tools to evaluate the sensitivity and specificity of this fox DNA test. Microsatellite and SNP genotyping have improved our understanding of Australian fox population structure, and we are now developing better genotyping tools for non-invasive mark-recapture studies.

Finally, DNA metabarcoding analysis of scats reveals the diets of introduced and native predators, including foxes, cats, Tasmanian devils and quolls. Scat metabarcoding also has great potential as a survey method for rare native species. To improve data interpretation we conducted captive feeding trials, analysing scats from predators fed controlled diets. We have also developed a reference DNA sequence database, which includes mitochondrial sequences from a wide range of Australian vertebrates, including almost all Tasmanian terrestrial mammals.

117

Landscape Conservation Genomics: from invasive to foundation species

Justin O Borevitz[1], Jared Streich[1], Megan Supple[1]

1. Australian National University, Acton, ACT, Australia

Genomics approaches allow the empirical evaluation of standing genetic diversity across geography and environment in several species. The patterns of neutral population genetic structure reveal the history of migration and gene flow defining the scale of isolation by distance and environment, which has fundamental implications for tracking invasion and managing ecosystems. We have surveyed thousands of individuals of a model invasive species *Brachypodium hybridum* and foundation species *Eucalyptus melliodora,* at scores of collection sites across the southeastern range of Australia. Multiple inbred colonizing lines of *B. hybridum* can be traced back to their native Mediterranean and West Asian range while a few locations have novel genetic diversity available to local selection, suggesting that limiting introduction is a higher priority than limiting expansion of naturalized populations. Although Eucalyptus forests are highly fragmented across their range, we found very little evidence of historical structure and isolation by distance explained only 4% of genetic variation. This suggests that a focus on local provenancing of a few remnant trees could be maladaptive given current land use and accelerating climate change causing a reduction of genetic diversity in this foundation species. These tools, when deployed for conservation, can lead to better management strategies than current reliance on null models of clonal invasion or local adaptation.

118

Adaptive and neutral genomic diversity of the global invader *Ambrosia artemisiifolia*

Lotte A van Boheemen[1], Kristin Nurkowski[1], Loren H Rieseberg[2], Kathryn A Hodgins[1]

1. School of Biological Sciences, Monash University, Clayton, Vic, Australia

2. Department of Botany, University of British Columbia, Vancouver, BC, Canada

To elucidate aspects of the genomic basis of rapid local adaptation, we have examined neutral and adaptive genetic variation of the widespread *Ambrosia artemisiifolia* along parallel environmental clines. Additionally, our goal is to enhance knowledge on the invasion history of this highly invasive weed, native to North America and recently introduced to many places worldwide, including Europe and Australia. Besides re-establishment along latitudinal clines we have found this species occurring outside of its native niche, making it an excellent study species to gain insight into local adaptation. Together with the use of molecular, genetic and advanced statistical tools we are able to reconstruct invasion history and population genetic structure at high resolution. This is feasible because of thorough sampling in native and two of the introduced ranges. Furthermore, we will associate genotypes to climatic data to shed light on recent adaptive change and identify divergent and parallel patterns at the genetic level. Comparison of adaptive and neutral genomic diversity between the native and introduced ranges will enhance our understanding of the mechanism at play during contemporary evolution.

119

Source dynamics of the naturally re-established carnivore, *Canis lupus*

Sarah Hendricks[1], Robert Wayne[2], Paul Hohenlohe[1]

1. University of Idaho, Moscow, IDAHO, United States

2. Ecology and Evolutionary Biology, University of California, Los Angeles, Los Angeles, CA, USA

Given the increase in anthropogenic influences, decreased genetic connectivity of wildlife populations is particularly important for management planning as it can threaten genetic diversity by reducing effective population size. Increased population genomics resolution with the advances of high throughput sequencing has allowed conservation biologist to assess previously unresolvable questions. For example, wolves in the Pacific Northwest USA (PNW) have recently re-established through colonization either from reintroduced inland US populations or from a unique ecotype found in coastal British Columbia (cBC; Canada). If the PNW wolves are contiguous with the inland populations, less protection may be deemed necessary; however, if genetic connectivity between PNW and cBC populations exists, different protection measures may be recommended. To assess the genetic source of PNW wolf populations, we generated data from a capture array and found that PNW wolves have multiple genetic sources and have varying degrees of admixture across geographic space. Additionally, we used ecological niche modeling to assign each pack's probability of presence in inland or coastal habitat. We found that two western PNW packs are more likely to exist in coastal habitat. Continued migration from cBC to PNW will contribute genetic diversity and possibly decrease inbreeding in the re-established PNW population. Our sequence capture array also targeted putative loci under selection, giving us insight into gene flow patterns for functional genetic variation potentially linked to local adaptation in these populations. These findings need to be considered when designing accurate conservation and management plans for wolves naturally re-colonizing the PNW.

120

Discovery of new RNA viruses in healthy honeybee populations with *Varroa* infection

Emily J Remnant[1], Mang Shi[1], Gabriele Buchmann[1], Tjeerd Blacquière[2], Edward C Holmes[1], Madeleline Beekman[1], Alyson Ashe[1]

1. University of Sydney, Sydney, NSW, Australia

2. Wageningen University, Wageningen, The Netherlands

Since the emergence of the parasitic mite *Varroa destructor*, the viral landscape of honeybees (*Apis mellifera*) has changed dramatically due to the increased spread of virulent viral strains, resulting in widespread loss of wild and managed honeybees. However, in various pockets of the world, populations of honeybees have naturally evolved or been selected for resistance to *Varroa* mites. These *Varroa*-tolerant populations provide a unique opportunity to capture the process of host-parasite adaptation as it develops in multiple locations, and to investigate the interrelationship between viruses, mite vectors and their honeybee hosts.

Using whole transcriptome sequencing, we examined the viral landscape of *Varroa*-tolerant honeybees from three locations in Europe, Africa and the Pacific. Almost all of the known honeybee viruses characterised thus far are positive-sense, single stranded RNA viruses of the Picornavirales order. Along with many of these common honeybee viruses, we found genomic evidence of seven previously undetected viruses, including four novel, negative sense RNA viruses. Two of these viruses belong to a common class of arthropod viruses, the Rhabdoviridae. Both Rhabdoviruses were found in all three of our geographically diverse locations and were also present in *Varroa* mites parasitising the bees. Small RNA profiles in infected bees indicated active Dicer-mediated degradation of both Rhabdoviruses. *Varroa* mites showed a different small RNA degradation profile suggesting that an alternate form of viral processing that occurs in mites.

These discoveries signify new classes of negative sense RNA viruses in honeybees, and raise the possibility that there are a multitude of undescribed viruses found in both *Varroa* and *A. mellifera* that may be adequately suppressed by innate immune pathways to prevent pathology.

121

Seeds of destruction? The relationship between inbreeding and male fertility in two threatened bird species

Helen R Taylor[1], James V Briskie[2], Patricia Brekke[3], Neil J Gemmell[1]

1. University of Otago, Dunedin, OTAGO, New Zealand

2. School of Biology, University of Canterbury, Christchurch, New Zealand

3. Institute of Zoology, Zoological Society of London, London, UK

Male fertility is negatively affected by inbreeding across a wide variety of taxa, but the cause of this apparently universal fragility remains mysterious. Sperm quality (sperm count, motility and morphology) has frequently been recorded as having a negative relationship with inbreeding, but cross-species comparisons are required to establish whether this universal fragility is attributable to the same genes in all species. The relationship between inbreeding and sperm quality has rarely been assessed in wild populations, has almost exclusively focused on mammals. In birds, males are the homogametic (ZZ) rather than the heterogametic (ZW) sex and, according to Haldane's rule, should be less susceptible to inbreeding depression than females. Here, we couple sperm quality data with extensive microsatellite genotyping to test the relationship between genetic diversity, inbreeding and sperm quality in bird populations that have experienced a range of bottleneck scenarios. Here we present data from South Island robin (*Petroica australis australis*) and hihi (*Notiomystis cincta*), both of which have experienced multiple translocation-induced bottlenecks in the past few decades. Inbred individuals of these species are known to exhibit increased nest failure compared to less inbred individuals, but the cause of this failure remains unknown. Our data suggest that at least some of these failures are due to male infertility as a result of reduced genetic diversity and increased inbreeding. We are currently assembling the South Island robin genome in an effort to identify particular genes that might be responsible for this trend to aid conservation management of these and other species.

122

Pedigree analysis reveals a generational decline in reproductive success of captive Tasmanian devil (*Sarcophilus harrisii*)

Katherine A Farquharson[1], Carolyn J Hogg[1, 2], Catherine E Grueber[1, 3]

1. School of Life and Environmental Sciences, Faculty of Veterinary Science, University of Sydney, Camperdown, NSW, Australia

2. Zoo and Aquarium Association Australasia, Mosman, NSW, Australia

3. San Diego Zoo Global, San Diego, CA, USA

Programs that manage threatened species in captivity are carefully designed to minimise genetic diversity loss, inbreeding and adaptation to captivity in order to maximise evolutionary potential. But, changes in productivity over generations in captivity can threaten the ability of captive breeding programs to reach their genetic targets and impair the potential of reintroduced captive animals to contribute to wild populations. These long-term changes have not yet been investigated in a zoo species. Marsupials are characterised by short generation times, putting them at particular risk of adaptation to captivity over short time periods. We tested for changes over time in captivity by analysing pedigree data from a nine-year Tasmanian devil insurance population studbook, including 338 animals across 17 breeding sites, and examined seven genetic, biological and management factors affecting reproductive success. We observed a substantial decline in reproductive success with increasing generations in captivity: captive-born females were less likely to produce a litter than wild-born females and when they did, they also produced fewer joeys. Reproductive success also declined as dam age at first breeding increased. The latter result implies a conflict with recommended conservation genetic strategy that aims to limit opportunity for selection and minimise adaptation to captivity by delaying reproduction. We suggest possible reasons for changes in productivity in relation to pedigree-based strategies for the management of captive populations. Our results have broad practical applications for the genetic management of threatened species worldwide; ongoing work will investigate the underlying molecular basis for these observations.

123

Revisiting an old evolutionary question: did the S mutation of the β-globin gene result from a single or multiple mutations?

Sandra Oliveira[1, 2], Giovanni Destro-Bisol[3, 4], António M. Santos[1, 2], Jorge Rocha[1, 2]

1. CIBIO, Vairão, VILA DO CONDE, Portugal

2. Faculdade de Ciências da Universidade do Porto, Porto, Portugal

3. Dipartimento di Biologia Ambientale, Sapienza Università di Roma, Rome, Italy

4. Istituto Italiano di Antropologia, Rome, Italy

The S mutation of the β-globin gene (*HBB*S*) in humans is a well-studied example of an advantageous allele with a protective role against malaria in the heterozygous carrier state. More than 50 years ago, Livingstone proposed that the emergence of tropical agriculture provided ideal habitats for the spread of malaria-transmitting mosquitoes allowing the rapid diffusion of a single, relatively recent *HBB*S* mutation. However, the concept of a single mutation was challenged by the finding that different *HBB*S*-linked haplotypes predominated in various non-overlapping geographical regions of Africa, India and the Arabian Peninsula. Currently, the most favored hypothesis explaining the geographic segregation of *HBB*S*-linked haplotypes is that *HBB*S* variants originated independently by recurrent mutation in each region where a single haplotype predominates. However, little work has been done to explicitly examine the effects of the spatial diffusion of the *HBB*S* allele on linked haplotype variation. Here, we explored a computer simulation framework to assess the evolution of *HBB*S*-linked haplotype variation in time and space, using a *stepping stone* model for the dispersal of an advantageous allele under different demographic scenarios. Moreover, we compared the simulated scenarios with an empirical dataset, consisting of 330 high resolution HBB*S-linked haplotypes defined by 11 microsatellites distributed across a 525 kb region. We show that the wave of advance of the*HBB*S* allele can originate patterns that mimic the spatial distribution of S-haplotypes, by creating several patches (or sectors), each formed by contiguous populations that share unique S-linked modal haplotypes exhibiting levels of haplotype diversity that are compatible with those currently observed in Africa. These findings bring back the hypothesis of a single origin as a plausible explanation for the evolution of the S mutation.

124

Trans-domain horizontal gene transfer from bacterium to animal: the post-transfer evolution of the aspzincins in the sponge *Amphimedon queenslandica*

Simone Higgie[1], Sandie Degnan[1]

1. The University of Queensland, St Lucia, QLD, Australia

The non-sexual horizontal transfer of genetic material across species boundaries is no longer considered a phenomenon occurring in bacteria only; increasingly it is reported in eukaryotes, including in animals. The extent and impact of horizontal gene transfer on animal evolution is not well understood; however, horizontal gene transfer is an evolutionary force with innovation potential – it can increase genomic variation, can directly transmit novel genes, and can provide sequence from which new genes can arise. Based on sequence similarities and genome location data, 576 putative horizontally transferred genes have been detected in a basal animal, the demosponge *Amphimedon queenslandica.* Within these genes, we found a large gene family of 83 bacterial-like aspzincins, a member of the metalloendopeptidase superfamily. We show that these genes are of bacterial origin, but result from one or few transfer events followed by extensive duplication. We suggest the transfer event(s) to be ancient since we have found bacterial-like aspzincins in seven additional sponge transcriptomes, which together represent both of the two major sponge lineages. Many of the aspzincins in *A. queenslandica* have striking transcriptional activity and have conserved the catalytically important residues; thus some may still have proteolytic activity. Our data support at least one case of transferred sequences fusing into one gene, creating a novel combination of an aspzincin and hemopexin domain containing protein. Together, these data contribute to our growing understanding of both the versatility of genomes and of the creative evolutionary power of horizontal gene transfer not only in bacteria, but across the web of life.

125

Tiny changes, big effects: the impact of microexons on neuronal differentiation, function and evolution

Manuel Irimia[1]

1. Centre for Genomic Regulation (CRG), Barcelona, Spain

One of the major challenges for the emergence of complex multicellular organisms is to generate an enormous diversity of cell types from a single genomic sequence. In the simplest scenario, the different cells would have the exact same protein complement available during embryo development to achieve their distinct functions and morphologies, often as divergent as those of a neuron or an erythrocyte. Therefore, how could neurons, for instance, tweak the structures and properties of this common set of proteins to optimize their specific and distinct neuronal functions without jeopardizing those in other cell types? A well-known evolutionary mechanism to overcome this challenge is gene duplication and functional subfunctionalization of the copies. However, a less well established – and yet probably more flexible and widespread – mechanism with similar consequences is alternative splicing (AS). Through differential processing of introns and exons, AS can produce cell type-specific protein isoforms that allow optimization of their specific cellular functions. One of the most striking examples of this is provided by microexons in neurons. These tiny exons, which can encode as little as one or two aminoacids, are switched on during neuronal differentiation and show the highest evolutionary conservation of all AS types. They are often located in structured domains of proteins, where they subtly sculpt their interaction surfaces thereby modulating protein-protein interactions in a neuronal-specific manner. Although we are still beginning to unveil their biological functions, we already know they crucial for proper neuritogenesis, axon guidance, and neuronal function. The remarkable example of microexons illustrates how a co-regulated program of cell type-specific AS can diversify proteins sequences to generate novel molecular functions as well as optimize ancestral ones for complex cell type-specific tasks.

126

Genome evolution of a parthenogenesis-inducing *Wolbachia* symbiont

Amelia Lindsey[1], John Werren[2], Stephen Richards[3], Richard Stouthamer[1]

1. University of California Riverside, Riverside, CA, USA

2. University of Rochester, Rochester, NY, USA

3. Baylor College of Medicine, Houston, TX, USA

Wolbachia is an intracellular symbiont of invertebrates responsible for inducing a wide variety of phenotypes in its host. These host-*Wolbachia* relationships span the continuum from reproductive parasitism to obligate mutualism, and provide a unique system to study genomic changes associated with the evolution of symbiosis. We present the genome sequence from a parthenogenesis-inducing *Wolbachia* strain (*w*Tpre) infecting the minute parasitoid wasp *Trichogramma pretiosum.* The *w*Tpre genome is the most complete parthenogenesis-

inducing *Wolbachia* genome available to date. We use comparative genomics across 16 *Wolbachia* strains, representing five supergroups, to identify a core *Wolbachia* genome of 496 sets of orthologous genes. Only 14 of these sets are unique to *Wolbachia* when compared to other bacteria from the Rickettsiales. We show that the B-supergroup of*Wolbachia*, of which *w*Tpre is a member, contains a significantly higher number of ankyrin repeat-containing genes than other supergroups. In the *w*Tpre genome, there is evidence for truncation of the protein coding sequencing in 20% of open reading frames, mostly as a result of frameshift mutations. The *w*Tpre strain represents a conversion from cytoplasmic incompatibility to a parthenogenesis-inducing lifestyle, and is required for reproduction in the *Trichogramma* host it infects. We hypothesize that the large number of coding frame truncations has accompanied the change in reproductive mode and potentially the evolution of mutualistic associations with its host.

127

Patterns and mechanisms of diminishing returns of advantageous mutations

Xinzhu Wei[1], Jianzhi Zhang[1]

1. University of michigan, Ann Arbor, MI - MICHIGAN, United States

Diminishing returns refers to the phenomenon that the same advantageous mutation is less beneficial when occurring in fitter genotypes than when occurring in less fit genotypes. Although diminishing returns has been frequently observed in experimental evolution, its prevalence in natural polymorphisms is unknown. Further, the cause of diminishing returns remains elusive. Here we address these questions using the genome sequences of 1005 haploid segregants produced from a cross between two budding yeast strains and the mitotic growth rates, a proxy of fitness (*F*), of these segregants in 47 different environments. Under each environment, we estimated the fitness effect size (*s*) of each single nucleotide polymorphism (SNP) from 100 segregants. Depending on the environment considered, 60-95% of SNPs exhibit diminishing returns (i.e., *s* decreases as the mean *F* rises from the 100 least fit segregants to the 100 fittest segregants). Surprisingly, given a median *F*, we found *s* to decrease as the environment quality (*Q*), measured by the mean fitness of all segregants in the environment, rises, revealing a previously unrecognized property of diminishing returns. The fraction of SNPs showing diminishing returns in an environment increases with *Q*, with or without the control for the median *F* of the segregants compared across environments. Our results are explainable by neither idiosyncratic nor global epistasis, but may arise under a modular epistasis model where the fitness of a genotype can be improved in a limited number of ways. Together, our findings reveal that diminishing returns is a general phenomenon that depends on both genotype quality and environment quality and that it arises from modular epistasis. Our findings explain why the poorer the environment the faster the adaptation.

128

Divergent patterns of marsupial-eutherian genomic imprinting revealed from RNA-seq analysis in the opossum, *Monodelphis domestica*

Xu Wang[1], Kory C Douglas[2], John L VandeBerg[3], Paul B Samollow[2], Andrew G Clark[1]

1. Cornell University, Ithaca, NY, United States

2. Texas A&M University, College Station, TX

3. Southwest National Primate Research Center, San Antonio, TX

Among the ~150-200 imprinted genes identified in mouse and human, only 20 marsupial orthologs have been examined to date, and eight of these were found to be imprinted. Here we ask, what is the marsupial imprinting status for the remaining 130 eutherian imprinted genes, and are there any marsupial-specific imprinted genes? We profiled genome-wide allele-specific expression (RNA-seq), histone modifications (ChIP-seq) and DNA methylation (PyroMark) in fetal brain and extra-embryonic membranes from reciprocal crosses of two opossum lines, providing an unbiased survey of parent-of-origin effects. Among 68 genes known to be imprinted in eutherians (and having an opossum ortholog), 52 were covered with sufficient informative SNPs to score allelic expression. Only three (<6%) were found to be imprinted in opossum, and 48 display biallelic expression, reflecting a striking lack of conservation of imprinting status. We also discovered and validated eight marsupial-specific imprinted genes that are not known to be imprinted in any other species. Surprisingly, three of these are non-coding lincRNA genes with no homology to any eutherian sequences, but they are present and are highly conserved in other marsupial species and non-mammalian vertebrates including chicken. Three of the rest five protein-coding imprinted genes were paralogous to eutherian genes, resulting from recent gene family expansions in opossum. Mechanistically, our epigenetic profiles confirmed that opossum-specific imprinted genes are regulated in the same way as eutherians by differential promoter methylation, despite the evolutionary fluidity of the imprinting profile. We estimate that opossums imprint only 30-40 genes, or about one-fifth the number imprinted by eutherian mammals. The smaller number and non-overlapping nature of imprinted genes could be due to the primitive placentation and shorter gestation time in marsupials compared to eutherians. Our study provides the first imprinting profile in a marsupial and sheds light on the regulation and evolution of genomic imprinting in mammals.

129

Transcriptional changes and the genomics basis of adaptation to an extreme environment

Joanna L Kelley[1], Anthony P Brown[1], Courtney Passow[2], Lenin Arias-Rodriguez[3], Michael Tobler[2]

1. Biological Sciences, Washington State University, Pullman, WA, United States

2. Division of Biology, Kansas State University, Manhattan, KS, United States

3. División Académica de Ciencias Biológicas, Universidad Juárez Autónoma de Tabasco, Villahermosa, Tabasco, México

Multiple closely related populations of *Poecilia* fish have adapted to hydrogen sulfide-rich springs from adjacent nonsulfidic streams, making them an ideal system to study the role of expression variation and structural changes in adaptation. Hydrogen sulfide (H_2S) is a potent toxicant that interrupts the mitochondrial respiratory chain and is also associated with severe hypoxia in aquatic environments. Constitutively high levels of dissolved H_2S place strong selective pressures on populations. RNA-sequencing of gills from natural and common-garden experiments provide evidence for the genomic and transcriptional basis of adaptation to this extreme environment. Genome-wide gene expression patterns across wild-caught replicated pairs of sulfidic and non-sulfidic populations clustered individuals by habitat type instead of by geographic and phylogenetic similarity. Though most differential gene expression between ecotypes was drainage specific, a small number of genes were consistently differentially expressed in the same direction in all sulfidic and nonsulfidic population pairs. Those shared differentially upregulated genes were associated with enzymatic H_2S detoxification and transport of oxidized sulfur species, oxidative phosphorylation, and pathways involved in responses to oxidative stress. In a common garden exposure study we recovered similar patterns of differential expression in wild-caught and laboratory populations that correspond to adaptation to H_2S. We found evidence that evolution, and less so ancestral plasticity, is responsible for generating variation in gene expression across replicated pairs of populations. We coupled these findings with a scan for highly differentiated and fixed loci between nonsulfidic and sulfidic populations and found evidence for selection on sulfide detoxification genes in all three drainages.

Moreover, the set of loci under positive selection is enriched for differentially expressed loci. Modification of processes associated with H_2S detoxification and toxicity through expression-level and sequence-level changes complement each other to mediate elevated H2S tolerance in sulfide spring fishes.

130

Using genomics to study evolutionary biology of the trichomonads

Jane Carlton[1]

1. Center for Genomics and Systems Biology, Department of Biology, , New York University, New York, United States

Trichomonads are ubiquitous anaerobic flagellated protists belonging to phylum Parabasalia. They infect many vertebrate and invertebrate species, with four species being classically recognized as human parasites: *Trichomonas vaginalis*, which causes the most common non-viral sexually transmitted infection; *Trichomonas tenax,* which is associated with periodontal disease; and *Dientamoeba fragilis* and *Pentatrichomonas hominis*, which are associated with intestinal upsets and may have a zoonotic origin. We are investigating the genomes of this "understudied, undervalued, but amazingly interesting" microbial eukaryotic lineage. The first trichomonad genome sequencing project was devoted to *T. vaginalis* (Carlton *et al.,* Science 2007), and we have leveraged this data trove in several ways. One standout finding of the project was the massive infiltration of the *T. vaginalis* genome by transposable elements (TEs). In this talk we will describe effects of these genomic "space invaders" on expression of the host protist's genes, as well as the regulation of TE genes themselves by small RNA molecules. We have also generated partial genome sequence data for ~100 *T. vaginalis* isolates to identify potential new biomarkers of antiparasite drug resistance, which has become a problem for treatment of trichomoniasis. Finally, we have developed 18S rRNA amplicon sequencing methods to survey *T. vaginalis* and other trichomonads in the urban environment of New York City.

131

On the quest of elucidating polyketide synthases responsible for ciguatoxin production

Anna L Kretzschmar[1], **Shauna Murray**[1], **Gurjeet Kohli**[1]

1. University of Technology Sydney, Ultimo, NSW, Australia

Ciguatera fish poisoning is a food borne disease that is increasing its prevalence in tropical areas. The causative organisms are marine dinoflagellates of the genus *Gambierdiscus*, some species of which produce the neurotoxins ciguatoxins, which bioaccumulate in fish. Determining the genetic basis for the production of these polyketide toxins will be key to understanding their ecology, evolution and pharmacology. However, due to their large genomes (0.5-40 x the size of the human haploid genome) and complex genetic processes, it is currently very difficult to sequence dinoflagellate genomes.

Here, we focused on sequencing the transcriptomes of two highly toxic strains of *Gambierdiscus polynesiensis* (as verified by LC-MS), as well as a non-toxic strain of *G. carpenteri*. The transcriptomes were assembled and de novo annotated. Polyketide related genes were investigated using hmmr searches and compared to known dinoflagellate polyketide and fatty acid synthase genes. Phylogenetic analysis of the polyketide active domains shows the complex evolutionary relationship of polyketide synthases involved in cell function and toxin production.

This work is instrumental on the road of pin pointing which polykrtide synthases are responsible for ciguatoxins production and designing an early detection system for ciguatera outbreaks.

1. Kohli, G.S., John, U., Figueroa, R.I., Rhodes, L.L., Harwood, D.T., Groth, M., Bolch, C.J. and Murray, S.A., 2015. Polyketide synthesis genes associated with toxin production in two species of Gambierdiscus (Dinophyceae). BMC genomics, 16(1), p.410.
2. Murray, S.A., Suggett, D.J., Doblin, M.A., Kohli, G.S., Seymour, J.R., Fabris, M. and Ralph, P.J., 2016. Unravelling the functional genetics of dinoflagellates: a review of approaches and opportunities. Perspectives in Phycology.

132

Fragmented evolution of an extremophilic red alga: Massive influx of bacterial DNAs revealed by the segmental landscape of the *Galdieria sulphuraria* genome

Ravi S. Pandey[1], **Garima Saxena**[1], **Debashish Bhattacharya**[2], **Rajeev K. Azad**[1]

1. University of North Texas, Denton, Texas, United States

2. Rutgers University, New Brunswick, New Jersey, United States

Galdieria sulphuraria stands out among the unicellular alga owing to its versatile characteristics, such as, survival on different carbon media and adaptation to extreme or hostile environments, namely, high heat, heavy-metal-toxic, and high acidity. The environmental adaptation has been attributed to the bacterial and archaeal genes that were acquired by *G. sulphuraria* during the course of its evolution. Previous studies reported ~5% of the *G. sulphuraria* genes to have been acquired through horizontal gene transfer. This was considered a conservative estimate, and the contribution of gene transfer to the extremophily of this organism is not yet well-understood, primarily because of the lack of the sequenced genomes of its close relatives. To circumvent the limitations of the current comparative genomics methods that require a rich sampling of the related genomes for a reliable inference of "alien" genes, we adapted a "comparison-free" method that had earlier been applied to assess the impact of horizontal gene transfer on the evolution of prokaryotes. We first analyzed the genome of *G. sulphuraria* for the presence of compositionally atypical genes using an integrated recursive segmentation and agglomerative clustering method. The clusters of compositionally homogeneous segments generated by this method revealed that ~30% of *G. sulphuraria* genome harbors atypical genes. The genes found in these regions were further assessed using pairwise sequence alignment of their protein products with protein sequences in the databases. A subset of predicted atypical genes (~15% of the *G. sulphuraria* genes) were found to have closest similarities with bacterial and/or archaeal genes but not the algal genes. The potential donor organisms, mostly proteobacteria and archaea, are known to thrive in various hostile conditions, and possess genes for the metabolism of various toxic elements. Our study reveals many, yet unreported, alien genes, documenting frequent transfer of genes from different prokaryotes to an extremophile eukaryote.

133

Evolution of the eukaryotic redox sensitive proteome via endosymbiosis

Christian Woehle[1], Tal Dagan[1], Giddy Landan[1], Assaf Vardi[2], Shilo Rosenwasser[2]

1. Institute of Microbiology, Kiel University, Kiel, Germany

2. Department of Plant Sciences, Weizmann Institute of Science, Rehovot, Israel

The redox sensitive proteome in eukaryotes comprises proteins that can be modified in the presence of reactive oxygen species (ROS). Reversible oxidation of proteins by ROS constitutes an important redox signaling mechanism. Anecdotal evidence suggested that several redox sensitive proteins originated from the plastid ancestor. However, the major evolutionary transitions of redox signaling in eukaryotes remain unknown. Here we perform a large-scale phylogenomic reconstruction of the eukaryotic redox sensitive proteome. In our study we analyzed redox sensitive Cys residues in protein sequences of the diatom *Phaeodactylum tricornutum*. Our comparative analysis includes 132 genomes from which 7,326 phylogenetic trees are reconstructed. We find that the majority of redox sensitive Cys residues (52%) are encoded in genes with prokaryotic nearest neighbors in eubacteria. The remaining reactive Cys are encoded in eukaryotic specific genes (39%), genes with prokaryotic nearest neighbors in archaea (5%) or genes that could not be classified (4%). Evolutionary reconstruction of amino acid ancestral states reveals that redox sensitive Cys residues are 2-fold enriched in genes whose origin coincides with the primary plastid endosymbiosis event in comparison to the background Cys residues. Furthermore, we find a similar enrichment regarding amino acid replacements into redox sensitive Cys that coincides with the secondary endosymbiosis event. Our results reveal direct inheritance of redox sensitive proteins from the plastid ancestor by primary endosymbiosis as one major expansion of the redox proteome. This was followed by adaptation of existing genes to extended redox signaling linked to the secondary endosymbiosis event. Thus, the primary and secondary endosymbiosis events played a key role in the evolution of the redox sensitive proteome in eukaryotes, either via endosymbiotic gene transfer or by linking already existing proteins into the redox signaling network.

134

Genetics and evolution of early branching *Saccharomyces.*

EmilyClare P Baker[1], William G Alexander[1], David Peris[1], Quinn K Langdon[1], Chris Todd Hittinger[1]

1. Laboratory of Genetics, University of Wisconsin - Madison, Madison, WI, United States

Yeast of the *Saccharomyces* genus, including the classical model *Saccharomyces cerevisiae* and less well known members of the genus, are important for many fermentative processes. Among these species are two early branching members of the genus, the sister species *Saccharomyces uvarum*(*Saccharomyces bayanus var. uvarum*) and *Saccharomyces eubayanus*. Important contributors to the brewing industry, they are strongly allied with different beverages, almost always with a *S. cerevisiae* genetic component. *S. uvarum* is associated with wines and ciders, while *S. eubayanus* is the non-*cerevisiae* parent of lager yeasts. Despite their differences in brewing associations and ~6.5% nucleotide divergence these sister species are physiologically very similar and it is not immediately clear why they show different brewing associations. Using genome sequencing and traditional genetic approaches we aim to understand what separates these species from each other and what contributes to their importance in the brewing industry. Comparing the *S. eubayanus* genome in lager yeasts to its parent genome we found that lager yeasts' *S. eubayanus* genome has experienced relaxation of selective pressure. Analyzing the *S. eubayanus* parent genome we identified fourteen genes that likely contribute to maltose utilization (essential for effective brewing). We are further exploring the function of these genes using gene-knockout and replacement studies. Moving from how and why a single species adapted to brewing to determining how both species have diverged genetically and phenotypically we are directly comparing their genome sequences and analyzing hybrids. Initial analyses indicate possible differences in RNA metabolism potentially leading to new insights into the role of RNA metabolism throughout *Saccharomyces.* As early branching members of *Saccharomyces* studying *S. eubayanus* and *S. uvarum* can not only provide insight into the genetic and physical characteristics of these economically important yeast, but further our understanding of the evolution of this scientifically and industrially important genus.

135

Systems Biology Studies of Yeast Gene Evolution at the Genomic Level Reveal Drivers of Evolution

Xianfa Xie[1], Noela Muma[1], Peter Lipke[2]

1. Virginia State University, Petersburg, VIRGINIA, United States

2. City University of New York, Brooklyn, New York, U.S.A.

What drives gene evolution remains to be answered. We try to address this question by studying the gene evolution across closely related yeast genomes. The nucleotide substitution and indel rates have been quantified for all the genes with good data, from which genes with high or low evolutionary rates are identified. Detailed analyses using a systems biology approach reveal the relationship of evolutionary rates to the location and function of the coded proteins, the gene expression levels, the essentiality of the gene, as well as the network positions of the protein products in protein-protein interaction, and the structural characteristics of the coded protein. Such results may help address the general question on the drivers of gene evolution at genomic levels.

136

Domestication, demography, and deleterious alleles in maize.

Jeffrey Ross-Ibarra[1]

1. University of California, CA, United States

Diversity across plant and animal genomes is patterned by the combined effects of drift and natural selection. To better understand the role of these processes in domesticated organisms, we investigated diversity across the genomes of domesticated maize and its wild relative teosinte. We first show that there is little evidence of selection on beneficial amino acid substitutions, and that the domestication bottleneck led to a decline in the efficiency of purifying selection in maize. We then show that rapid expansion post-domestication dramatically changed this relationship, with stronger purifying selection in maize, reflecting the much larger effective size of present day populations. Finally, we resequenced a number of genomes of landrace maize to show the impacts of colonization and demography on the distribution of deleterious alleles in the maize genome.

137

Domestication of the dog

Claire M Wade[1]

1. University of Sydney, University Of Sydney, NSW, Australia

The geographical origin of *Canis familiaris* has been an issue of some contention in the canine genomics community. There are conflicting opinions as to whether domestication occurred in the Middle East, South East Asia or Africa. Subtle interpretation of the genomic clues provides evidence for all stances. What is clear is that the domestic dog has a fascinating genomic architecture that has arisen from its unique population history in collaboration with humans. This unique genomic architecture and a suite of state of the art genomic tools means that the canine is one of the most successful organisms to use for Mendelian trait mapping. Through genomics we can see the effects of recent human-made selective sweeps and can discover how relatively few genomic changes may result in the multitude of diverse phenotypes that we observe in dogs today.

138

What makes the fungi cheesy: the adaptive genomic toolbox of *Penicillium* molds revealed by between- and within-species comparisons

Ricardo C Rodríguez de la Vega[1], **Emilie Dumas**[1], **Jeanne Ropars**[2], **Manuela Lopez-Villavicencio**[3], **Sandrine Lacoste**[3], **Stephanie Le Prieur**[1], **Robert Debuchy**[4], **Sophie Siguenza**[5], **Jerome Gouzy**[5], **Joelle Dupont**[3], **Antoine Branca**[1], **Tatiana Giraud**[1]

1. Ecologie Systematique et Evolution, Univ. Paris-Sud, CNRS, AgroParisTech, Université Paris-Saclay, Orsay, France

2. Fungal Biology and Pathogenicity, Institut Pasteur, Paris, France

3. ISEB, Museum National d'Histoire Naturelle, Paris, France

4. IGM, Univ. Paris-Sud, Orsay, France

5. LIPM, Institut National de la Recherche Agronomique, Toulouse, France

Studying parallel evolution on the same environment is a powerful approach for understanding the genomic processes of adaptation. Several species of the ubiquitous *Penicillium* molds have been independently domesticated for the maturing of cheese since at least the 18th century. Two key Penicillium species used for cheese-making are *P. roqueforti* for blue cheeses and *P. camemberti* for soft surface-ripened cheeses. These distantly related species have been independently selected for optimal growth and enzymatic activities in a human-made nutrient-rich environment containing numerous bacterial and fungal competitors. Domestication is an excellent model for studies of adaptation because it involves recent and strong selection on a few, identified traits, here we present part of our studies on the genomic footprints of domestication in chese-making fungi. First, by comparing the genomes of 29 Penicillium species, we found that adaptation to the cheese medium was associated with parallel gains, losses and positive selection on genes involved in the utilization of the cheese nutrients and competition with other microorganisms. A substantial part of the gene family expansions correspond to previously identified horizontally-transferred regions recently acquired by the cheese-making fungi[1]. Second, by focusing on the intraspecific genetic diversity at the whole genome level in 30 *P. roqueforti* strains and 3 closely related species from different sources (used for chese-making and growing in the wild), we reveal two independently domesticated groups of *P. roqueforti*, thus giving yet another example of the ease of *Penicillium* species to adapt on a short evolutionary timescale. Experiments validated fitness differences between the genetic groups. Overall, we show that adaptation can occur rapidly in eukaryotes, through the acquisition of a novel genomic metabolic toolbox. These findings contribute to improve our understanding of the genomic processes of adaptation to rapid environmental changes.

1. Ropars, Rodríguez de la Vega, et al. (2015). Adaptive horizontal gene transfers between multiple cheese-associated fungi. Current Biology 25(19):2562-2569 http://dx.doi.org/10.1016/j.cub.2015.08.025.

139

Selection for tameness in Red Junglefowl causes wide changes in brain gene expression

Johan Bélteky[1], **Per Jensen**[1], **Beatrix Agnvall**[1]

1. Linköping University, Linköping, ÖSTERGöTLAND, Sweden

Domestication, an accelerated evolution driven by man, puts pressure on organisms to adapt to their new environment. This has generated a large variety of phenotypes in a short time-span, as seen in a number of domestic animals including the domestic chicken.

Our hypothesis is that domestic phenotypes may to have developed partly as correlated responses to early selection for tameness. By reenacting early selection pressure on a population of wild-type animals we can study the associated phenotypic and genetic changes. We have bred three lines of the wild ancestor of the domestic chicken, the Red Junglefowl (Gallus gallus). By selecting solely for the trait fear of humans, we have two lines displaying high or low fear, and one unselected showing intermediate fear levels. Currently five generations of offspring have been bred within each line, and several phenotypic changes, correlated to the selection trait, can be observed: low fear birds gain weight faster, show higher feeding motivation and more aggressive behaviour, while the high fear birds show significantly more avoidance behaviour. Low fear birds also display higher dominance ranking, lay larger eggs with heavier chicks, and have better plumage condition.

We performed genetic analysis by means of transcription microarray and PCR on hypothalamus and cerebral hemisphere from selected animals. Whilst the unselected group did not differ from the original parental population, the high and low fearful groups in generation S5 differed from each other significantly for a number of genes in both tissues. Amongst the differentially expressed genes, we detect functions ascribed to immunology, weight and behaviour. We also detected a number of transcripts previously annotated with sperm-associated functions. The results support the hypothesis that domesticated phenotypes may evolve because of correlated effects related to reduced fear of humans.

140

Genome analysis of the red fox reveals genetic basis of domesticated behavior.

Anna V Kukekova[1], **Jennifer L Johnson**[1], **Halie M Rando**[1], **Xueyan Xiang**[2], **Shaohong Feng**[2], **Shiping Liu**[2], **Anastasiya V Kharlamova**[3], **Anastasiya V Vladimirova**[3], **Rimma G Gulevich**[3], **Jessica P Hekman**[1], **Zijun Xiong**[2], **Gregory M Acland**[4], **Xu Wang**[5], **Andrew G Clark**[5], **Lyudmila N Trut**[3], **Guojie Zhang**[2, 6, 7]

1. University of Illinois at Urbana-Champaign, Urbana, IL, United States

2. China National Genebank, BGI -Shenzhen, Shenzhen, Guangdong, China

3. Institute of Cytology and Genetics, Russian Academy of Sciences, Novosibirsk, Russian Federation

4. Baker Institute for Animal Health, Cornell University, Ithaca, NY, USA

5. Department of Molecular Biology and Genetics, Cornell University, Ithaca, NY, USA

6. Department of Biology, University of Copenhagen, Copenhagen, Denmark

7. State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, China

Differences in behavior of domesticated animals from their wild ancestors provide one of the best examples of the influence of genes on behavior but the identification of the underlying genes has been proved to be extremely difficult. Unlike the species, which were domesticated historically and selected for many different traits including morphology and appearance, the red fox (*Vulpes vulpes*) was domesticated in a controlled experiment by selection solely for behavior. The fox model may therefore provide advantages for the identification of genes implicated in domesticated behavior. We sequenced and assembled the genome of the red fox and re-sequenced a subset of foxes from domesticated population, population selected for aggressive behavior, and conventional farm-bred population to identify genomic regions associated with selection for behavior. The sequence analysis identified 30 regions of reduced heterozygosity in domesticated population and 34 regions of extreme divergence between domesticated and other fox populations, 13 identified regions overlapped between the two analyses. The nonsynonymous substitutions were identified in several genes located in identified regions including genes involved in synaptic plasticity and signal transduction. One region contained a single gene, *SorCS1*, a trafficking regulator of neurexin and AMPA receptors. The effect of this gene on behavior was confirmed using fox experimental three-generation pedigrees constructed by breeding domesticated foxes and foxes selected for aggressive behavior. The analysis of the fox brain transcriptome suggested the effect of *SorCS1* genotypes on gene expression. The fox findings may be further validated in dogs and may help to shed light on the intriguing question of whether domesticated behavior in different species is regulated through similar gene networks.

141

The genome sequence of a 5300-year-old maize cob recovered from the Tehuacan Valley provides insights into the early stages of maize domestication

Jazmin Ramos Madrigal[1], Bruce D Smith[2], Jeffrey Ross Ibarra[3], Marcus Thomas Gilbert[1], Nathan Wales[1]

1. Centre for GeoGenetics, Natural History Museum of Denmark, University of Copenhagen, Copenhagen, Denmark

2. Program in Human Ecology and Archaeobiology, Department of Anthropology, National Museum of Natural History, Smithsonian Institution, Washington DC, USA

3. Department of Plant Sciences, University of California, Davis, California, USA

Domesticated from a wild teosinte grass in southern Mexico more than 6,500 years ago, and perhaps much earlier, maize (*Zea mays* ssp. *mays*) is one of the world's most important crops. Over a span of more than six millennia, the domesticate has undergone considerable and rapid evolution under human selection, as it responded to a wide range of different growing conditions and selective pressures.

Genetic analysis of well-preserved maize samples recovered from archaeological contexts of different ages and from different regions of the Americas hold the promise of providing a detailed understanding of the complex evolutionary history of maize, and the timing and sequence of development of different selective traits.

Here we present the genome sequence of a 5300-year-old cob recovered from an archaeological site located ~450 km from the domestication center, in the Tehuacan Valley of Mexico. We compared our data to a reference dataset comprising 23 modern landraces and 17 wild teosinte genomes. By using D-statistics, a model-based clustering algorithm, and a multidimensional scaling analysis, we show that the specimen derives from the same source population that gave rise to modern maize, and that it represents an evolutionary branch during the early stages of domestication. Moreover, we show that 5,300 years ago, the maize genome was already more similar to modern maize than to its wild counterpart. Finally, we find that this ancient maize carries several domestication genes in the ancestral state, and only few in the domesticated state, supporting the idea of a punctuated domestication process.

142

The effect of archaic admixture on inferred human population history

Isabel Alves[3, 1, 2], Aboriginal Australian Consortium, Laurent Excoffier[3, 1]

1. Swiss Institute of Bioinformatic, Lausanne, Switzerland

2. Instituto Gulbenkian de Ciência, Oeiras, Portugal

3. Institute of Ecology and Evolution, Bern, Switzerland

The availability of high-quality archaic human genomes has revealed a complex admixture history of modern humans. However, it is still unclear to which extent neglecting past admixture events with archaics has biased our view of the evolutionary relationship (e.g divergence times) between modern humans (MH). To address this question, we first performed computer simulations under various scenarios including archaic admixture and show that if two modern human populations are differentially admixed with an archaic population their divergence time is strongly overestimated. We then quantified this bias by using a data set including high coverage genomes of Aboriginal Australian (AA), Eurasians, African and archaic humans. A likelihood analysis of the multidimensional site-frequency spectrum reveals that neglecting archaic contribution into MH leads to a divergence time between AA and Eurasians that is considerably older (~20 ky) than what is estimated under a much better supported scenario that explicitly accounts for differential archaic introgression into these populations. Finally, under our best scenario, we estimate that admixture from Denisovans into MH occurred as recently as 31-55 kya. We conclude with a discussion on the implications of our admixture times for the population history of our species.

143

Mutation, migration, standing variation: the where and how of convergent adaptation

Kristin Lee[1], Graham Coop[1]

1. University of California, Davis, Davis, CALIFORNIA, United States

Geographically separated populations can convergently adapt to the same selection pressure. Convergent evolution at the level of a gene may arise via three distinct modes. The selected alleles can (1) have multiple independent mutational origins, (2) spread throughout subpopulations via gene flow, or (3) be shared due to shared ancestral standing variation. To understand the impact of convergent positive selection on neutral diversity at linked loci, we make use of the fact that hitchhiking can be modeled as an increase in the variance in neutral allele frequencies around a selected site within a population. We develop coalescent theory to show how shared hitchhiking events between subpopulations act to increase covariance in allele frequencies between subpopulations at loci near the selected site, and extend this theory under different models of migration

and selection on the same standing variation. We incorporate this hitchhiking effect into a multivariate normal model of allele frequencies that also accounts for population structure. Based on this theory we present a composite likelihood-based approach that utilizes genomic data to identify loci involved in convergence, and distinguishes among alternate modes of convergent adaptation. We illustrate our method on genome-wide polymorphism data from four populations of killifish (*Fundulus heteroclitus*) that show very rapid convergent adaptation for tolerance to industrial pollutants. We identify a single locus at which both independent mutation events and migration play a role in adaptation across the species' range.

144

Uniparental inheritance promotes adaptive evolution in the mitochondrial genome

Joshua Christie[1], Madeleine Beekman[1]

1. School of Life and Environmental Sciences, The University of Sydney, Camperdown, NSW, Australia

In addition to their nuclear genome, eukaryotic organisms carry multiple asexual mitochondrial genomes. Lacking recombination, asexual genomes should struggle to accumulate beneficial substitutions and to purge those that are deleterious. Yet, empirical evidence of adaptive evolution in mitochondria suggests that they suffer none of these limitations of asexual reproduction. Here I use a computational model to show that the unique biology of mitochondrial genomes—in particular their large copy number combined with uniparental inheritance—enable them to readily undergo adaptive evolution. Uniparental inheritance increases variation in fitness between individuals, selecting against individuals with a high deleterious substitution load and for individuals that carry multiple beneficial substitutions. I will show that uniparental inheritance decreases competition between different beneficial substitutions (clonal interference), reduces genetic hitchhiking of deleterious substitutions during selective sweeps, and promotes adaptive evolution by increasing the level of beneficial substitutions relative to deleterious substitutions. When assuming that mitochondria inherit biparentally, the presumed ancestral state, decreasing the number of genomes transmitted during gametogenesis (transmission bottleneck) aids adaptive evolution. However, uniparental inheritance is required to maintain variation in fitness between individuals on which selection can act. As a result even a tight transmission bottleneck combined with biparental inheritance leads to less efficient adaptive evolution than when inheritance is uniparental. These findings explain the empirical observations that mitochondria—despite their asexual mode of reproduction—can readily accumulate beneficial substitutions and resist the accumulation of deleterious substitutions.

145

Holy Chicken! Selection analysis applied to time series of ancient genotype data reveals how medieval religious reform shaped the genomes of modern chickens.

Liisa Loog[2, 1], Mark Thomas[3], Andrea Manica[2], Greger Larson[1], Anders Eriksson[2]

1. Research Laboratory for Archaeology and the History of Art, University of Oxford, Oxford, UK

2. Department of Zoology, University of Cambridge, Cambridge, UK

3. Department of Genetics, Evolution & Environment, UCL, London, UK

A number of statistical approaches have been developed in recent years to detect and quantify the strength of natural selection using modern genomic data. However, these approaches all have poor temporal resolution and limited power to detect selection acting on standing genetic variation. Ancient DNA allele frequency data provides the most direct and sensitive alternative for detecting selection at specific loci, and offers the possibility of resolving temporal variation in selection strength. However, ancient DNA sample sizes are typically small, and samples tend to be sparsely and unevenly distributed in space and time. In addition, all approaches are sensitive to confounding effects of demography. Here we present a Bayesian framework for reconstruction allele frequency trajectories through time from ancient allele frequency data that can explicitly accommodate the confounding effects of gene flow between populations and uncertainty in sample ages. We applied this method to ancient European domestic chicken genotype data from TSHR locus, which has been argued to be under strong and recent selection in domestic chickens. We find that the hypothesized sweep allele shows a pattern of strong selection starting 1100 years ago, coinciding with a European-wide known shift in poultry management between the mid-ninth to mid-eleventh century. This shift is associated with religious but also a legal rule that required people to abstain from meat, brought in as part of the Benedictine reform. This directly highlights the importance of ancient DNA and statistical modeling for understanding how cultural practices in the past have shaped modern domesticated species.

146

A novel algorithm for selective sweeps detection in bacteria

Oren Avram[1], Yaara Oren[1], Eli Levy Karin[1], Tal Pupko[1]

1. Cell Research and Immunology, Tel Aviv University, Tel Aviv, Israel

Selective sweeps occur when a beneficial mutation spreads rapidly throughout the population due to natural selection. Searching for selective sweeps has proved to be one of the most fruitful ways to detect the footprints selection leaves on the genome.

In the last decade, methods to identify selective sweeps were developed as a powerful tool for uncovering the genomic basis of adaption. With a plethora of detection tools, the study of selective sweeps in eukaryotic systems is a well-established field of research. However, the search for selective sweeps in bacteria received little to no attention.

In our work we demonstrate that selective sweeps can also be detected in bacteria. Focusing on a comparative genomics *E.coli* database, we first study the performance of a commonly used selective sweep detection method in eukaryotes over these data and we discuss its limitations. Subsequently, we devise SAP, a novel phylogeny-based method for incomplete selective sweeps detection. We apply it to the *E.coli* database and detect several cases of incomplete selective sweeps. Using simulations, we demonstrate that most of these cases cannot be explained by neutral evolution under a model of no selection and no recombination, suggesting a bona fide signal for sweeps.

Since our methodology is not strain-specific but rather general, it can be applied to many other bacterial species, as long as they are able to share genetic material with their "neighbors". Thus, we expect our new method should contribute to the effort of understanding bacterial phenotypic variation and adaptation at the genomic level.

147

High-throughput experimental estimation of the effects of all amino-acid mutations to HIV's envelope protein on viral replication in cell culture

Hugh K Haddox[1, 2], **Jesse D Bloom**[2]

1. *Molecular and Cellular Biology, University of Washington, Seattle, Washington, United States*
2. *Basic Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, Washington, United States*

HIV is notorious for its ability to evade immunity and anti-viral drugs through rapid evolution. Thus, efforts to create an HIV vaccine aim to target conserved parts of the virus likely to have a limited evolutionary capacity for escape. However, our understanding of HIV's evolutionary capacity is incomplete. To further this understanding, we used deep mutational scanning to experimentally estimate the effects of all $\approx 10^4$ possible single amino-acid mutations to most of HIV's envelope protein (Env) on viral replication in cell culture. First, we made a library of *env* genes with ~1-2 codon-level mutations per gene and with nearly all possible single amino-acid mutations. Next, we selected for functional variants that supported viral replication in human T-cells. We then used next-generation sequencing to measure the amount that each mutation was enriched or depleted upon selection. From this data, we estimated each site's preference for each of the 20 amino acids using a statistical framework and software package developed by our laboratory. We compared our estimates of Env's site-specific amino-acid preferences in cell culture to the actual frequencies of these amino acids in naturally occurring HIV sequences. Our measurements are strongly correlated with amino-acid frequencies in natural sequences for most sites known to be important for conserved protein functions such as receptor binding. While this correlation is also high for most buried sites, it is dramatically lower for surface-exposed sites that are subject to pressures not present in our experiments such as antibody selection. We plan to further explore the mutational tolerance at antibody epitopes to identify possible antibody-escape mutations. Ultimately, we plan to repeat this experiment in the presence of antibody selection to identify actual escape mutations. Overall, our estimates of Env's mutational tolerance in cell culture provide a basis to better understand HIV's evolutionary capacity in nature.

148

Functional Sites Induce Long-Range Evolutionary Constraints in Enzymes

Benjamin Jack[1], **Austin G Meyer**[1], **Julian Echave**[2], **Claus O Wilke**[1]

1. *University of Texas at Austin, Austin, TX, United States*
2. *Escuela de Ciencia y Tecnologia, Universidad Nacional de San Martin, San Martin, Buenos Aires, Argentina*

The basic biochemical functions of life are carried out by large molecules called enzymes. Enzymes consist of long chains of amino acids folded into a three-dimensional structure. Within that structure, a specific cluster of amino acids, known as the active site, performs the biochemical function. Substituting one amino acid for another in the active site typically results in a defective, non-functional enzyme, and therefore mutations at or near enzyme active sites are often lethal. Moreover, even mutations far from the active site have been found to disrupt function. Nonetheless, as organisms evolve, enzymes accumulate random mutations. Where in enzymes' structures do these mutations accumulate without causing harm? We observe evidence for extensive interactions between active sites and distant regions of the enzyme structure, in a comprehensive set of over 500 enzymes. We show that active sites tightly control the substitutions that an enzyme can tolerate. This control extends far beyond regions of the enzyme immediately adjacent to the active site, covering over 80% of a typical enzyme structure. Our findings have broad implications for molecular evolution, for enzyme engineering, and for the computational prediction of active-site locations in novel enzymes.

149

Weakly selected standing variants dominate adaptation for 1000 generations in sexual, laboratory evolved yeast populations

Katya Kosheleva[1], **Michael Desai**[1]

1. *Harvard University, Cambridge, MASSACHUSETTS, United States*

Populations adapt by acquiring and fixing beneficial mutations. However, the typical number and strength of such mutations remains a subject of debate. In this work, we tackle this question by investigating the genetic basis of long-term adaptation in hybrid yeast populations evolved under different rates of outcrossing. We track the dynamics of adaptation on new and standing variants by resequencing populations and individuals at fixed times over the course of 1000 generations. Additionally, we assay changes in the mean and variance in fitness over time. We find that populations undergoing any amount of sex continue to adapt and maintain substantial variance in fitness through the duration of the experiment, while accumulating few new mutations. Despite this sustained adaptation, allele frequency changes stagnate after several hundred generations, and only a handful of regions that initially experience rapid change proceed to fix. We propose that this phenomenon of sustained adaptation despite stagnating genetic change may be explained by a model of dense, weakly selected sites. In this model, rapid allele frequency changes are the result of many alleles of small effect linked over ten or hundreds of kilobases hitchhiking together. As recombination breaks associations between distant sites, the effect of a typical linkage block declines. However, because individuals at later timepoints sample a greater variety of genotypes, the population continues to adapt. This work suggests that adaptation on weakly selected variants can dominate adaptation for hundreds of generations, and may result in dynamics that may deviate strongly from the selective sweep paradigm.

150

The Koala Genome Consortium – the utilization of *de novo* genome and transcriptome sequencing for applied conservation genomics of an iconic Australian marsupial

Rebecca Johnson[1, 2]

1. *Australian Museum Research Institute, Australian Museum, Sydney, NSW, Australia*
2. *Koala Genome Consortium, http://data.koalagenome.org ,*

The koala, *Phascolarctos cinereus*, is a biologically unique and evolutionarily distinct Australian arboreal marsupial that is frequently regarded as an 'iconic symbol of conservation' due to a range of threatening processes such as disease and habitat loss. The Koala Genome Consortium is a multi-disciplinary collaboration utilizing both genomic and transcriptomic data to investigate conservation in this species via a genomics approach.

This presentation will report *de novo* koala genome and transcriptome assembly and annotation, including the unique molecular attributes we have discovered for this iconic, monotypic marsupial. Three geographically separate koalas (two female and one male), have been genome sequenced at 50x-100x coverage using a range of platforms (including illumina and PacBio) and RNAseq for >15 tissues will be included. We report evidence for copy number expansion of the alpha amylase gene, aldehyde reductase gene and an assessment of koala phylogeographic history, which despite recent dramatic population declines, shows surprisingly weak levels of genetic differentiation

Version 1 of the koalal genome release will not only allow comparison between the relative benefits of hybrid and single platform genomes but will also form the foundation of evolutionary genetic studies of marsupials in general, thorough analysis of koala immune genes and importantly the opportunity for genome level investigation of the koala retrovirus.

151

Colourful conservation – conservation genetics and selection on colour polymorphism in the threatened Gouldian finch

Peri E Bolton[1], Lee A Rollins[2], James Brazill-Boast[1], Kang-Wook Kim[3], Kimberly L Maute[4], Sarah Legge[5], Simon C Griffith[1]

1. Macquarie University, Sydney, NSW, Australia

2. Deakin University, Geelong, VIC, Australia

3. University of Sheffield, Sheffield, Yorkshire, United Kingdom

4. University of Wollongong, Wollongong, NSW, Australia

5. Australian Wildlife Conservancy, Perth, WA, Australia

Distinct strategies associated with colour morphs can have effects on population dynamics, and on the partitioning and population structure of genetic variation. Antagonistic interactions and hybrid incompatibilities between morphs could have negative consequences for population fitness, and its effects may be magnified in a small population. Therefore, covariation between colour and sexual, behavioural and physiological traits may be relevant to conservation management. The Gouldian finch (*Erythrura gouldiae*) has two sympatric head colour morphs, and with a history of population declines this species is a flagship for conservation in Northern Australia. Captive studies have shown that the sympatric colour morphs correspond to different behavioural strategies, and interbreeding between morphs leads significant offspring mortality. This selection on head colour polymorphism may be a detriment to recovery of Gouldian finch populations, and could lead to genetic substructuring with respect to colour morph. We explore population structure and selection on head colour using traditional population genetics and a recently developed novel marker that determines the underlying genotype for head-colour. We demonstrate the utility of this marker to examine hypotheses with respect to selection on head colour and incompatibility. Across five geographically disparate sampling localities there is evidence of extensive gene-flow between them and between colour morphs. Gene-flow and head-colour frequencies suggest that incompatibility between colour morphs is not a threatening process in the wild.

152

Effects of selection on genetic diversity in threatened species breeding programs: empirical data from Tasmanian devil

Catherine E Grueber[2, 1], Belinda Wright[1], Katrina M Morris[1, 3], Carolyn J Hogg[4, 1], Katherine Belov[1]

1. School of Life and Environmental Sciences, Faculty of Veterinary Science, The University of Sydney, NSW, Australia

2. San Diego Zoo Global, San Diego, USA

3. Roslin Institute, The University of Edinburgh, Edinburgh, United Kingdom

4. Zoo and Aquarium Association, Sydney, Australia

Studies of laboratory populations and commercial species have shown that populations bred in captivity can experience massive genetic changes, relative to wild source populations, over surprisingly short timeframes. These changes may be attributable to several processes, including founder effects, genetic drift, and adaptation to captivity. The impact of captive breeding on the diversity of threatened species in conservation contexts remains vastly under-studied, but has serious implications for the large number of programs that aim to release captive-bred animals to the wild to support natural populations. The Tasmanian devil insurance population was established as a captive breeding program in 2006 in response to the decimating impact of devil facial tumour disease, which has caused severe population crashes in Tasmania since the disease emerged in 1996. The program now numbers over 700 devils, in more than 35 institutions across Australia, and presents an ideal opportunity for addressing questions relating to the impact of captive breeding in conservation. We measured the genetic effects of captive breeding using molecular genotyping of more than 300 SNPs across 22 candidate regions of interest (including immune genes) via a novel SNP typing assay developed by our group. We used parent-offspring trios to investigate changes in diversity by comparing observed offspring genotypes to expectations under neutral, Mendelian inheritance, using custom Monte Carlo (gene-drop) simulations. Preliminary findings revealed varying distributions of diversity across different genomic regions, with some regions indicating deficits of heterozygosity that deviate from neutral expectations. In this talk, I will outline the results that we found, how they may be interpreted in terms of possible selection pressures, and their implications for genetic management of not just devil, but all threatened species maintained in captivity.

153

Broken barriers: anthropogenic disturbance leads to continental-scale levels of genetic diversity

Katie L Millette[1], Andrew Gonzalez[1], Melania E Cristescu[1]

1. McGill University, Montreal, QUEBEC, Canada

Environmental selection in disturbed habitats are expected to erode diversity. However, habitat disruption can also unlock the evolutionary potential of populations by breaking ecological barriers and opening populations up to introgression of novel genotypes. We investigate a natural aquatic system historically exposed to a mosaic of anthropogenic stressors (heavy metals, acidification) and give evidence that localized disturbance can maintain levels of genetic diversity typically seen at regional and even cross-continent spatial scales. Using nuclear and mitochondrial genetic markers, and a population genetic survey of species of the *Daphnia pulex* complex from contaminated and control lakes, we show that contaminated lakes harbour high levels of haplotype and nucleotide diversity. Despite the geographic proximity and joint watershed of many lakes, the observation of shared multi-locus genotypes among populations is rare. Yet the observation and dominance of hybrid genotypes in three highly stressed habitats favours the hypothesis that ecological transitions can trigger hybridization and the emergence of novel asexual (obligately parthenogenetic) clones.

154

Lack of genetic diversity across diverse immune genes in an endangered mammal, the Tasmanian devil

Katrina M Morris[2, 1], Belinda Wright[2], Catherine E Grueber[2, 3], Carolyn Hogg[2, 4], Katherine Belov[2]

1. *Roslin Institute, University of Edinburgh, Easter Bush, MIDLOTHIAN, United Kingdom*
2. *The University of Sydney, Camperdown, NSW, Australia*
3. *San Diego Zoo Global, San Diego, CA, USA*
4. *Zoo and Aquarium Association, Mosman, NSW, Australia*

Genetic diversity of immune genes is critical for population-level resistance against infectious diseases. Studies in diversity of immune loci in wild species have almost exclusively focused on Major Histocompatibility Complex (MHC) genes; however these genes only account for a fraction of immune gene diversity. Tasmanian devils (*Sarcophilus harrisii*) are threatened with extinction by the spread of Devil Facial Tumour Disease. Devils are known to lack diversity at MHC and Toll-like receptor genes. Whether there are polymorphisms in devil immune genes more broadly is unknown. In this study, we characterised diverse immune gene families in the devil genome, and identified SNPs in a wide range of important immune genes. Using genome-level data from ten devils we identified SNPs within the exons, introns and flanking regions of 167 immune genes, including cytokines, chemokines and natural-killer cell receptors. From this data, we developed long-amplicon assays to target nine key genes which were sequenced in up to 220 devils. We found an extreme paucity of genetic diversity across a broad range of immune genes. Despite this, signatures of balancing selection were exhibited by one chemokine gene, suggesting that remaining diversity may hold adaptive potential. The low functional diversity may leave devils highly vulnerable to infectious disease; therefore monitoring and preserving remaining diversity is critical for the long-term management of this species. Examining genetic variation in diverse immune genes should be a priority for threatened wildlife species. This study can act as a model for broad-scale immunogenetic diversity analysis in wild populations.

155

Genomics-led biodiversity discovery in reptiles across the monsoonal tropics of northern Australia

Craig Moritz[1], Mozes Blom[1], Jason Bragg[1], Sally Potter[1], Renae Pratt[1], Dan Rosauer[1], Ana Silva[1]

1. *The Australian National University, Acton, ACT, Australia*

The dramatic increase in scale allowed by next-gen sequencing coupled with emerging coalescent-based analysis methods offers great scope for increasing our understanding of biodiversity pattern and its evolutionary foundations. We have combined mtDNA phylogeography with sequence and SNP data for 1000s of loci obtained via custom exon capture to uncover the diversity of widespread lizard taxa across Australia's tropics. The results indicate that traditional morphology-based taxonomy underestimates diversity several-fold, especially in geckos. Mapping of phyloendemism confirms several well-known biodiversity hotspots, but also reveals new ones, especially on islands and in presumed climatic refugia in areas with harsher climates. Much of this diversity occurs in indigenous managed lands, which have higher biodiversity value than currently realised. In some areas, initial results point to a high frequency of past introgression across multiple taxa. Understanding how speciation-extinction dynamics have shaped these various patterns remains a challenge, and we propose that stability of local micro-climates has played a key role. If so, this bears on how landscapes are managed for conservation

157

Genomic insights into the mechanisms of priority effects in nectar yeast

Manpreet K Dhami[1], Thomas Hartwig[2], Tadashi Fukami[1]

1. *Department of Biology, Stanford University, Stanford, CA, United States*
2. *Department of Plant Biology, Carnegie Institution for Science, Stanford, CA, USA*

Floral nectar is a harsh environment that hosts a complex community of microbes. Floricolous yeast, *Metschnikowia reukaufii*, is a dominant species in this community, competing with other fungi and bacteria. Specifically, this species exerts a strong priority effect, excluding other microbes and deterministically influencing community structure. We sequenced the genome of *M. reukaufii* and identified the genetic mechanisms by which it specialises in nectar. We found a high rate of tandem gene duplication in the genome, with majority of duplicated genes involved in nitrogen metabolism and transport. The two high-capacity amino acid transport genes, *GAP1* and *PUT4,* involved in amino acid scavenging when nitrogen is scarce, were present in tandem gene arrays. We confirmed that all four copies of *GAP1* were expressed in nectar conditions and in fact regulated via the nitrogen catabolite repression pathway, establishing their role in nitrogen scavenging. *M. reukaufii* is able to efficiently deplete up to 90% of amino acids present in nectar soon after colonisation, limiting potential competition from late-arriving species. These genes, expressed only in nitrogen-limited environment provide an example of adaptation to high carbon, low nitrogen nectar resource, and explain the strong priority effects exhibited by *M. reukaufii*.

158

Next generation suite for molecular evolutionary genetics analysis (MEGA)

Sudhir Kumar[1], Koichiro Tamura[2]

1. *Temple University, Philadelphia, PENNSYLVANIA, United States*
2. *Biology, Tokyo Metropolitan University, Tokyo, Japan*

Molecular Evolutionary Genetics Analysis (MEGA) software contains many sophisticated methods and tools for phylogenomics and phylomedicine. In this talk, the focus will be on major new developments that will enable researchers to use MEGA for (a) phylogenetic inference of very large datasets, (b) building very large timetrees quickly and accurately using innovative methods, (c) automatically predicting gene duplication events in gene family trees, and (d) conducting high-throughput and scripted analyses using a computational core. The new MEGA software suite is now 64-bit and is available in two interfaces: graphical and command line. The graphical user interface (GUI) is a native Microsoft Windows application that can also be used on Mac OSX. The command line computational core MEGA-CC is available as native applications for Windows, Linux, and Mac OSX. All MEGA versions are available from www.megasoftware.net free of charge.

159

What is new in BEAST

Remco Bouckaert[1]

1. University of Auckland, Auckland

Bayesian phylogenomic methods advance at a rapid pace, and this talk will be a whirlwind tour of the latest developments in BEAST and its packages. In particular, we will have a look at

o bModelTest, a Bayesian approach to dealing with site models,

o multi gamma site models for dealing with heterotachy.

o the fossil birth death model and CladeAge, two new ways for dealing with fossils calibrations.

o spherical phylogeography, which allows whole world analysis with somewhat inhomogeneous diffusion so that landscape features can be taken in account.

o various ways to do species delimitation, including STACEY, BFD (for *BEAST) and BFD* (for SNAPP).

o structured coalescent and approximations, allowing larger number of demes.

o ARG trees for bacterial data through BACTER.

We will have a look at computational advances, making BEAST faster than before. If you have any further questions or problems with BEAST, you can visit me at the `BEAST clinic' in the trade area.

160

IQ-TREE: efficient phylogenomic software for maximum likelihood analysis

Arndt von Haeseler[1], **Olga Chernomor**[1], **Bui Quang Minh**[1], **Lam-Tung Nguyen**[1], **Heiko A Schmidt**[1], **Jana Trifinopoulos**[1]

1. University of Vienna, Vienna, VIENNA, Austria

We present IQ-TREE, the successor of the TREE-PUZZLE and IQPNNI software, as an efficient and versatile phylogenomic tool for maximum likelihood analysis of large alignments. IQ-TREE supports multiple sequence types (DNA, protein, codon, binary and morphology) and a wide range of evolutionary models including mixture and partition models. IQ-TREE performs fast model selection, partition scheme finding, likelihood mapping, efficient tree reconstruction, terrace aware phylogenomic inference, ultrafast bootstrapping, branch tests, and tree topology tests. IQ-TREE supports parallel computation and provides check-pointing to restart phylogenetic analysis if the original run was interrupted. IQ-TREE infers usually higher likelihood trees than RAxML and PhyML while requiring similar amount of computing time. However, the performance of all three likelihood tools depends on the data set. Thus, it is useful to run several programs.

IQ-TREE is open-source, it supports all major platforms (Windows, Mac OSX, Linux) and is freely available from http://www.cibiv.at/software/iqtree. For novice users a user-friendly web server is available at http://iqtree.cibiv.univie.ac.at.

161

Beyond software tuning: scaling up comparative coding sequence analysis using approximations and models that adapt their complexity to the data.

Sergei LK Pond[1], **Spencer V Muse**[2], **Ben Murrell**[3]

1. Temple University, Philadelphia, PA, United States

2. Statistics, North Carolina State University, Raleigh , NC, United States

3. Medicine, University of California , San Diego, CA, United States

A practical upper bound on the number of sequences that can be analyzed with many popular comparative methods is ~10^3, especially if codon-substitution models are used. This number can be raised by several orders of magnitude, enabling the study of gene-sized alignments with 10^4 – 10^5sequences, much more extensive model testing, or the implementation of more realistic models with added complexity.

We describe a relatively general approximation technique to limit the number of expensive likelihood function evaluations *a priori,* by discretizing a part of the parameter space to a fixed grid, estimating other parameters using much faster simpler models, and integrating over the grid using MCMC or a variational Bayes approach. With FUBAR[1], we demonstrate how this technique can achieve 100× or greater speedups for detecting sites subject to positive selection, while improving statistical performance. Other analyses where there are only a 2-3 parameters of interest (e.g. detection of directional selection in protein sequences) can be accommodated.

When discretization is not approproate, it is often possible to develop methods that employ variable parametric complexity chosen with an information theoretic criterion. For example, in the Adaptive Branch Site Random Effects model [aSBREL, 2], we quickly select and apply models of different complexity to different branches in the phylogeny, and deliver statistical performance matching or exceeding best-in-class existing approaches, while running an order of magnitude faster.

1. Ben Murrell, Sasha Moola, Amandla Mabona, Thomas Weighill, Daniel Sheward, Sergei L. Kosakovsky Pond, and Konrad Scheffler FUBAR: A Fast, Unconstrained Bayesian AppRoximation for Inferring Selection Mol Biol Evol (2013) 30 (5): 1196-1205
2. Martin D. Smith, Joel O. Wertheim, Steven Weaver, Ben Murrell, Konrad Scheffler, and Sergei L. Kosakovsky Pond Less Is More: An Adaptive Branch-Site Random Effects Model for Efficient Detection of Episodic Diversifying Selection Mol Biol Evol (2015) 32 (5)

162

MALT: Fast alignment and analysis of metagenomic DNA sequence data applied to the Tyrolean Iceman

Alexander Herbig[1], **Frank Maixner**[2], **Kirsten I. Bos**[1], **Albert Zink**[2], **Johannes Krause**[1], **Daniel H. Huson**[3]

1. MPI for the Science of Human History, Jena, THURINGIA, Germany
2. Institute for Mummies and the Iceman, Bolzano, Italy
3. Center for Bioinformatics, University of Tuebingen, Tuebingen, Germany

Modern sequencing technologies produce vast amounts of DNA sequence data in the context of large-scale metagenomic studies revealing the complexity of microbial communities in unprecedented detail. These analyses require high-throughput computational methods that allow for a fast processing of sequencing data while retaining a high level of sensitivity and precision.

Here we present MALT (**M**egan **AL**ignment **T**ool) a program for the fast alignment of DNA sequencing reads to large reference genome databases such as GenBank. MALT is able to process hundreds of millions of reads within only a few hours. In addition to the alignment procedure MALT implements a taxonomic binning algorithm that specifically assigns reads to bacterial species or strains. Its tight integration with the interactive metagenomic analysis software MEGAN allows for visualization and further analyses of results. These analyses can be performed in a comparative manner for studying the dynamics of microbial communities over time, or from different habitats or hosts. In particular the human microbiome is of major interest as it is comprised not only of a large number of commensals, but potentially also pathogens that have evolved with their human host. To gain insights into these evolutionary relationships, the field of paleogenetics aims to study ancient DNA extracted from archaeological remains.

In this context we demonstrate MALT by its application to the metagenomic analysis of two ancient microbiomes from oral cavity and lung samples of the 5,300-year-old Tyrolean Iceman.

163

Assessment of substitution model adequacy for phylogenomics

David A. Duchene[1], Simon Y.W. Ho[1], Sebastian Duchene[1]

1. University of Sydney, Camperdown, NSW, Australia

Genome sequences offer a rich source of data for studying evolutionary relationships and biological processes. They present a number of challenges to phylogenetic methods, including complex patterns of variation and large computational demands. To improve the feasibility of phylogenetic analysis, genomic data are sometimes filtered according to a chosen criterion. One approach is to filter data according to how well the evolutionary model describes them, using one of several tests of model adequacy. However, the efficacy of these tests for identifying when phylogenetic inferences will be unreliable remains unknown. We propose a framework for assessing substitution model adequacy using fast likelihood methods. Based on a simulation study, we find that some test statistics can identify particular sources of bias. Other test statistics are highly conservative, frequently rejecting the model when the inferences are not inaccurate or imprecise. We demonstrate our framework by analysing three large data empirical sets, and find that selecting data using our approach can lead to different phylogenetic inferences. Model-adequate data according to our approach produce more congruent inferences than model-inadequate data, which has also been identified in previous research. Filtering genomic data using the test statistics identified in our simulation study improves the reliability of inferences, and can be useful tool for phylogenomic studies.

164

Bacterial viruses, can't live with them, can't live without them

Francisco Rodriguez-Valera[1]

1. Universidad Miguel Hernandez, Alicante, ALICANTE, Spain

Viruses and cells are among the most interwoven partnerships in biology. we have information from bacterial genomics and metagenomics indicating that the structure of bacterial populations is determined by their viral predators. Populations of planktonic bacteria are multi-clonal with different lineages that carry different viral receptors, i.e. have different viral susceptibility, they are also different in metabolic and environmental interaction. Like different tissues in a multicellular eukaryote they cooperate to exploit efficiently their habitat. The genetic wealth actually present in a drop of seawater within a single bacterial species is surprisingly large, rivalling, if not exceeding their eukaryotic counterparts, but to preserve this genetic wealth viral predation is of essence. Without viruses a selective sweep in which a clone becomes dominant is hard to avoid and would lead to irretrievable loss of genetic diversity and, eventually, much poorer ecological performance. This model of prokaryotic populations will be presented and evidence supporting it from analysis of genomes and metagenomes of marine bacteria.

165

Swapping partners mid-dance: Symbiotic replacement in a tightly integrated intrabacterial, intracellular nested mutualism

Stephanie R Weldon[1], Mark Ladinsky[2], John P McCutcheon[3]

1. NASA Postdoctoral Program, Missoula, MT, United States
2. Division of Biology and Biological Engineering , California Institute of Technology, California
3. Division of Biological Sciences, University of Montana, Missoula, MT, United States

Approximately ten percent of insect species require maternally inherited intracellular bacteria to survive and reproduce, generally because these bacterial mutualists provide nutritional supplementation to an insect surviving on a limited substrate. These intracellular insect symbionts include bacteria with the smallest cellular genomes yet sequenced, which have an organelle-like structural stability and mutation rate, attributable to their small effective population sizes and host-limited lifestyles. Furthermore, some insects are reliant on more than one obligate heritable symbiont, and generally these multi-symbiont systems show a division of labor in terms of their nutrient production. A subset of mealybugs (Pseudococcidae) rely on a binary obligate symbiont set where one partner, a gammaproteobacterium, is embedded in the cytoplasm of the other, a betaproteobacterium: this is the only known extant intracellular mutualism between two bacteria, within or outside of an insect host. The genomes of both bacterial partners show immense reduction and complementarity in the insect host where they have been best studied, *Planococcus citri*; despite this tight interdependence, however, the gammaproteobacterial partner has been replaced repeatedly in the mealybug lineage by relatives of the widespread insect endosymbiont *Sodalis*. This series of subsititutions has provided us with a unique opportunity to enhance our understanding of the process by which free-living bacteria are biochemically and physiologically integrated into highly specialized, multi-partite symbioses.

167

Disentangling diet and phylogeny reveals both horizontal and vertical evolution of mammalian microbiomes

Mathieu Groussin[1], Florent Mazel[2], Jon Sanders[3], Chris Smillie[1], Sebastien Lavergne[2], Wilfried Thuiller[2], Eric Alm[1]

1. Massachusetts Institute of Technology, CAMBRIDGE, MASSACHUSETTS, United States

2. University Grenoble Alpes, Laboratoire d'Écologie Alpine (LECA), Grenoble, -, France

3. Organismic and Evolutionary Biology, Harvard University, Cambridge, Massachusetts, United States

The persistence and intimacy of mammalian host-microbe interactions over evolutionary timescales is controversial. This is partly because host phylogeny and diet are so deeply confounded that it may not be possible to disentangle their individual contributions to microbiome evolution. Here, we show that host phylogeny and diet influence the distribution of independent gut bacterial lineages, and do so on vastly different timescales. Diet mostly influences the acquisition of deeply divergent (>300 Ma) microbial lineages, while associations with host phylogeny are seen across more recent lineages. Considering microbiome at appropriate phylogenetic scales allows us to model the evolution of the microbiome along the mammalian tree, and to accurately infer ancient diets from the predicted microbiomes of mammalian ancestors. More detailed phylogenetic analysis reveals for the first time, large-scale patterns of co-speciation between mammals and their gut symbionts, some of which are associated with immune diseases in humans. Surprisingly, the bacterial genera with the greatest amounts of co-speciation with their hosts have been widely overlooked in previous studies, laying a path for future studies that probe these newly discovered host-microbe associations for signs of co-evolution.

168

The Genomic Footprint of Lichenization: Comparative Genomics of Lecanoromycetes

Bastian Greshake[1], Andreas Blaumeiser[2], Anjuli Meiser[3], Francesco Dal Grande[4], Imke Schmitt[3, 4], Ingo Ebersberger[2]

1. Goethe University, Frankfurt am Main, Offenbach Am Main, HESSEN, Germany

2. Department for Applied Bioinformatics, Goethe University, Frankfurt am Main, Germany

3. Institute of Ecology, Evolution and Diversity, Goethe University, Frankfurt am Main, Germany

4. Senckenberg Biodiversity and Climate Research Centre, Frankfurt am Main, Germany

Lichenization, the symbiosis of a fungus with photosynthesizing alga or cyanobacteria is a highly successful lifestyle. It allows lichens to conquer ecological niches that are otherwise too extreme to support eukaryotic life. Most lichenized fungi are members of the Lecanoromycetes, accounting for about half of the so far described ascomycetes. Their degree of dependence on their symbiotic partner varies and has in many instances evolved to an extent that the respective fungus grow very poorly – if at all – in isolation. This renders lichen communities excellent examples to study the genomic footprint of the evolutionary process turning an autonomously living organism into an obligate symbiont. However, the data acquisition itself is a challenge. For lichens growing poorly in culture, the only feasible method of genome analysis is the metagenomic sequencing of the eukaryotic and bacterial components.

Building on a benchmarking study about potentials and pitfalls of lichen metagenomics, we here present the first high-quality genomes of a lichenized fungus, *Lasallia pustulata*, and its photobiont, *Trebouxia sp.*, generated purely from metagenomic samples. We compared the genome of *Lasallia pustulata* to two cultureable Lecanoromycetes, *Cladonia grayi* and *Xanthoria parietina*, and further 18 non-lichenized Pezizomycetes. This revealed a pronounced loss of otherwise highly conserved pezizomycete core-genes specifically on the lineage of *L. pustulata.* We confirmed the absence of these through gene order analyses. Their functional annotation through KEGG showed a widespread loss of metabolic capacity in *L. pustulata* across all metabolic pathways. This metabolic reduction can be at least partially mitigated by the bacterial communities present in *L. pustulata,* as metagenomic samples collected from a wide geographic range, showed a stable core-set of bacteria and metabolic functions across those. We conclude that lichenization enables reductive evolution in the individual symbionts, potentially made possible by the genomic redundancy across the lichen hologenome.

169

Nitrogen fixation by the conifer foliar microbiome

Carolin Frank[1], Lara M Kueppers[2], Andrew B Moyes[2], Dana L Carper[1], James Kupihea[1], Paola Saldierna[1], Dianne Quiroz[2], Jennifer Pett-Ridge[3]

1. University of California, Merced, Caucasian, CALIFORNIA, United States

2. Climate and Ecosystem Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, California, United States

3. Physical and Life Sciences Directorate, Lawrence Livermore National Lab, Livermore, California, United States

Mature temperate and boreal forest are nitrogen (N) limited, yet N budgets indicate unknown sources of N in these ecosystems. Symbiotic N_2-fixing plants are notably absent from coniferous forests, and sources to overcome N limitation are not well understood, but include epiphytic N_2-fixation in mosses, free-living N_2 fixation in litter and soil, bedrock nitrogen, and deposition of nitrogen pollution.

Via Illumina sequencing of the 16S RNA gene, we have found a consistent association between pines growing in nutrient limited ecosystems and specific bacteria, most notably potential N_2-fixing acetic acid bacteria (AAB), and Rhizobiales spp. These associations appear to be conserved across host species, time, and geographic distance, suggesting selection on the part of the tree, the bacteria, or both, potentially reflecting a functional partnership based on N_2-fixation. Using the acetylene reduction assay on surface sterilized foliar samples, we have confirmed nitrogenase activity in the subalpine conifer *Pinus flexilis* (limber pine) growing at Niwot Ridge, Colorado, as well as more recently, in in bishop pine (*Pinus muricata*) and lodgepole pine (*Pinus contorta* ssp. *bolanderi* and ssp. *contorta*) growing along a gradient in soil age and associated variation in soil fertility at the "ecological staircase" in Mendocino, California.

Not surprisingly, N_2-fixation rates of needle endophytes are much lower than those of nodulating N_2-fixers, and comparable to rates of free-living fixation in soils in temperate and boreal ecosystems. So far, we have found no evidence that local differences in soil N availability affect rates of foliar N fixation. Together, these results suggest that foliar endophytes represent a low-cost, evolutionarily stable N_2-fixing strategy for long-lived conifers that never fully alleviates N limitation in temperate and boreal ecosystems. These results open up the possibility that hidden symbiotic N_2-fixers hide in other N-poor ecosystems as well.

170

Moving from genetics to genomics in understanding climate change adaptation in Drosophila: what else can we learn?

Ary Hoffmann[1], John Oakeshott[2], Rahul Rane[3], Stephen Pearce[2], Renee Catullo[2], Pip Griffin[3], Ronald Lee[3], Carla Sgro[4], Marina Telonis-Scott[4]

1. University of Melbourne, Parkville, VIC, Australia

2. CSIRO, Canberra, ACT, Australia

3. The University of Melbourne, Parkville, VIC, Australia

4. Biological Sciences, Monash University, Clayton, Vic

Genetic studies of Drosophila have featured prominently in research to understand the potential of organisms to adapt through evolution to recent climate change. The early phase of this work focussed on quantitative genetic approaches, which was followed by later work focussing on specific genetic polymorphisms and temporal studies. Much of this work took advantage of widespread Drosophila species distributed along climate gradients. One of the core messages is that insects, like plants, can be genetically differentiated along gradients, reflecting a balance of local adaptation and movement as well as historical factors that remain poorly defined. In several instances, the adaptive significance of differentiated traits has been obvious, and in other cases it has been cryptic. The clinal studies have also proved to be a good source of information on adaptive processes acting on specific polymorphisms, and in some cases links between this work and genetic mapping of traits from variation within populations has been possible. Recently, this work has moved into an –omics phase, based on genomic assessments of clines and selection lines. Early results from this effort have highlighted the complexity of adaptation along clines but also the importance of chromosomal structural variation in climate adaptation. The results have also indicated where overlap in the genetic basis of adaptive shifts might be expected, based on comparisons of different selection lines and mapping efforts. New insights are now emerging from combining information from the intraspecific level to comparative genomics across species. We illustrate the insights emerging from this work by considering early results from a comparative analysis

171

Comparative Evolutionary Genomics of Adaptation to Environmental Change in Ecologically Important Aquatic Organisms

Luciano B Beheregaray[1]

1. Flinders University, Adelaide, SA, Australia

Understanding whether natural populations will be able to adapt to selective pressures associated with rapid environmental and climatic change is a research priority. Measuring the strength and characteristics of selection in natural populations remains a daunting task, particularly for non-model species. In this talk I present results (and several unresolved challenges) from three research programs that study adaptation in non-model marine and freshwater organisms by integrating population genomics with environmental modelling and common-garden experiments. Each program explores natural replicates of the adaptation process by comparing closely related lineages and populations in geographically separate environments or in shared environments. Our datasets include ddRAD, RNA-seq, candidate adaptive genes, habitat mapping and trait phenotyping. The main aim of the first program (23 populations of two species of abalones, n=732) was to assess the relative contributions of space and selection in large, well-connected marine populations. Substantial neutral gene flow was the norm in both species, but their adaptive datasets showed marked population structure associated with environmental heterogeneity; in particular, with thermal gradients. In the second program (50 populations of two perch species, n=638), we tested for associations between neutral and adaptive diversity and gradients of environmental disturbance. Both species showed neutral population structure linked to the riverine network. However, it appears that long-term environmental instability (measured by natural hydroclimatic disturbance) has promoted adaptive diversity and evolutionary resilience in these lineages. In the third program (39 riverine populations of two rainbowfish species, n=1035) we experimentally assessed adaptive potential to climate change and tested landscape predictions from the 'climatic variability hypothesis'. We showed heritability and heritable plasticity for the expression of candidate genes in future climates. At a landscape level, populations from more variable habitats showed higher adaptive resilience to climate change. Strategies for cataloguing adaptive resilience to environmental change in ecologically important organisms are discussed.

172

Molecular signatures of transgenerational brain response to ocean acidification in a reef fish

Celia Schunter[1], Megan Welch[2], Taewoo Ryu[3], Huoming Zhang[1], Michael L Berumen[1], Göran E Nilsson[4], Philip Munday[2], Timothy Ravasi[1]

1. King Abdullah University of Science and Technology, Thuwal, JEDDAH, Saudi Arabia

2. James Cook University, Townsville, Australia

3. APEC Climate Center, Busan, South Korea

4. University of Oslo, Oslo, Norway

The impact of ocean acidification on marine ecosystems will depend on the capacity for species to adapt. Recent studies show that the behaviour of reef fishes is impaired at projected future CO_2 levels; however there is also individual variation that might promote adaptation. We used a genome-wide approach to evaluate the potential heritability of this variation in CO_2 sensitivity in the spiny damselfish, *Acanthochromis polyacanthus*. Offspring of CO_2 tolerant and CO_2 sensitive parents were reared at near-future CO_2 (754 µatm) or present-day control levels (414 µatm). By integrating 36 brain transcriptomes and proteomes with a *de novo* assembled genome we investigated the molecular responses of the fish brain to increased CO_2 and the expression of parental tolerance to high CO_2 in the offspring molecular phenotype. Exposure to high CO_2 resulted in differential regulation of 173 and 62 genes and 109 and 68 proteins in the tolerant and sensitive groups respectively. Importantly, the majority of differences between offspring of tolerant and sensitive parents occurred in high CO_2 conditions. Consequently, there was a clear signature of parental sensitivity to high CO_2 in the molecular phenotype of the offspring, primarily driven by circadian rhythm genes. This transgenerational molecular signature suggests that individual variation in CO_2 sensitivity could facilitate adaptation of fish populations to ocean acidification.

173

Transgenerational epigenetic inheritance: mechanisms and biology in C. elegans

Eric A Miska[1]

1. University of Cambridge, Cambridge, CAMBRIDGESHIRE, United Kingdom

Since August Weismann (1834-1914) formulated the distinction between innate and acquired characteristics at the end of the 19th century, the debate relating to the inheritance of acquired traits has raised many controversies in the scientific community. Following convincing arguments against (*e.g.* William Bateson) this debate was then set aside by the majority of the scientific community. However, a number of epigenetic phenomena involving RNA, histone modification or DNA methylation in many organisms have renewed interest in this area. Transgenerational effects likely have wide-ranging implications for human health, biological adaptation and evolution, however their mechanism and biology remain poorly understood. We recently demonstrated that a germline nuclear small RNA/chromatin pathway can maintain epi-allelic inheritance for many generations in *C. elegans*. This is a first in animals. We named this phenomenon RNA-induced epigenetic silencing (RNAe). We are currently further characterizing the mechanism of RNAe. In addition, we are testing the hypothesis that RNAe provides a transgenerational memory of the environment ("Lamarckism"). We will present new data suggesting a role for RNAe in sensing nutrition and the environment.

1. Ashe, A., Sapetschnig, A., Weick, E.-M., Mitchell, J., Bagijn, M. P., Cording, A. C., et al. (2012). piRNAs can trigger a multigenerational epigenetic memory in the germline of C. elegans. Cell, 150(1), 88–99. http://doi.org/10.1016/j.cell.2012.06.018

174

Purifying selection causes genetic canalization of gene expression at high and low larval densities in *Drosophila melanogaster*

Ana Marija Jakšić[2, 1], **François Mallard**[2], **Viola Nolte**[2], **Christian Schlötterer**[2]

1. Graduate School of Population Genetics, Vetmeduni Vienna, Vienna, Austria

2. Institute for Population Genetics, Vetmeduni Vienna, Wien, ÖSTERREICH, Austria

Gene expression differences between genotypes are not homogeneous across environments. Across a temperature gradient expression differences are most pronounced at extreme temperatures. The similarity of gene expression between genotypes may be the outcome of higher efficacy of purifying selection in frequently encountered environments. However, the alternative explanation of decanalization at extreme, stressful environments cannot be ruled out. To distinguish between these two explanations we contrasted gene expression differences between two *D. melanogaster* strains cultivated at high or low larval density. We find more differences in gene expression for flies grown at low density, than for high density flies. Since low larval density is less stressful and the *D. melanogaster* strains were maintained for >80 years at high density conditions, we conclude that the similar gene expression profile of the two genotypes at high larval density is the outcome of purifying selection. Genetic variation causing expression differences at low larval density were probably not purged because this environment was rarely encountered. Our results suggest that expression differences between genotypes may frequently not be the outcome of adaptive changes, but may also reflect lack of purifying selection.

175

Phenotypic changes in organismal adaptation to new environments: plasticity distorts while evolution restores

Wei-Chin Ho[1], **Jianzhi Zhang**[1]

1. Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, Ann Arbor, MI, United States

At the phenotypic level, adaption to a new environment involves plastic changes followed by evolutionary changes, but neither the relationship between the two changes nor the underlying reason for the relationship is well understood. Here we address these questions by studying plastic and evolutionary changes of *Escherichia coli* metabolic fluxes upon switches of the nutritional environment. We computationally mimic plastic changes using minimization of metabolic adjustment (MOMA) and mimic evolutionary changes using flux balance analysis (FBA), because these mathematical tools have been shown to respectively predict plastic and evolutionary responses to perturbations. We find that an environmental alteration typically plastically distorts the fluxes of many reactions, a large fraction of which are then restored via evolutionary changes. Rarely are reactions subject to plastic and evolutionary changes in the same direction or subject to evolutionary changes only. This prominent pattern of distortion-restoration generally holds regardless of whether the new environment is richer or poorer than the original environment. These findings echo recent transcriptome-based observation that evolutionary changes in gene expression level tend to reverse plastic changes. Because the mechanisms of flux changes in MOMA and FBA are known, we are in the process of discerning the underlying cause of the plastic distortion followed by evolutionary restoration in environmental adaptations.

176

Genomic insights from two emerging model systems: Polygenic adaptation to simian immunodeficiency virus (SIV) infection in vervet monkeys, and the role of gene flow in rapid speciation in Lake Malawi cichlids

Hannes Svardal[2, 1], **Anna Jasinska**[3], **Milan Malinsky**[2, 4], **Christian Apetrei**[5], **Béatrice Jacquelin**[6], **Michaela Müller-Trutwin**[6], **Martin Genner**[7], **George Turner**[8], **Wesley Warren**[9], **Eric Miska**[2, 4], **Nelson Freimer**[3], **Richard Durbin**[2], **Magnus Nordborg**[1]

1. Gregor Mendel Insitute, Austrian Academy of Sciences, Vienna, Austria

2. Wellcome Trust Sanger Institute, Cambridge, United Kingdom

3. University of California Los Angeles, Los Angeles, USA

4. University of Cambridge, Cambridge, United Kingdom

5. University of Pittsburgh, Pittsburgh, USA

6. Institut Pasteur, Paris, France

7. University of Bristol, Bristol, United Kingdom

8. University of Bangor, Bangor, United Kingdom

9. Washington University in St. Louis, St. Louis, USA

Vervet monkeys (Chlorocebus sp.) are amongst the most widely distributed non-human primates and are natural hosts of simian immunodeficiency virus (SIV) with a high viral prevalence observed in all sampled African populations. We use whole-genome sequencing data from 163 individuals

covering six subspecies across the continent to infer relationships and demonstrate cross-taxon gene flow. A scan for diversifying selection across subtaxa yields strong enrichment for viral response genes and genes whose human orthologs interact with human immunodeficiency virus (HIV). Selection scores are also highly elevated in genes that show a response to experimental SIV infection in vervet monkeys but not in rhesus macaques. Because the latter species is not a natural SIV host, and develops immunodeficiency disease upon SIV infection, this interaction likely reflects adaptation to SIV. We discuss the strongest candidate genes for repeated selection for SIV defence.

I will also discuss more briefly results from whole genome sequencing data from 156 individuals from 71 species of the Lake Malawi cichlid radiation. We find that Lake Malawi cichlid species are much more closely related than previously thought and D-statistics (ABBA-BABA test) reveals massive signatures of ancient genetic exchange both with riverine out-groups and across the major clades within the radiation. Both paleogeological records of low lake level and a striking excess of haplotypes that are approximately 10 times more diverged than the genomic average are consistent with an ancestral hybrid swarm giving rise to the current radiation.

177

The population history of Aboriginal Australia

Anna Malaspinas[1], Aboriginal Australia Consortium, David Lambert[2], Eske Willerlsev[1]

1. Centre for GeoGenetics, Copenhagen, COPENHAGEN, Denmark

2. Environmental Futures Research Institute, Griffith University, Nathan, Australia

While Australia is unique in having one of the longest continuous period of human occupation outside Africa, its population history remains largely uncharacterized. The distinctiveness of the Australian archaeological record has led to the suggestion that the ancestors of Aboriginal Australians and Papuans are modern representatives of an early dispersal of modern humans out of Africa. To describe the origins of Aboriginal Australians, as well as their differentiation, adaptation and relatedness to other populations, we collected DNA sampled in collaboration with Aboriginal Australian communities and individuals in Australia. We sequenced genomes at high-depth from Aboriginal Australian individuals. The sequenced individuals represent a number Pama Nyungan languages and originate from regions geographically widespread across the Australian continent. Combining this dataset with whole genome data from Africans and Eurasians we investigated the number of migration waves out of Africa, explicitly taking archaic introgression into account. Based on the site-frequency spectrum we estimate that Aboriginal Australians and Eurasians derive mostly from a single out of Africa wave. Furthermore, we find that Papuan and Aboriginal Australian ancestors diverged long before Australia and New Guinea were separated by higher sea levels, suggesting early population structure in the ancient continent of Sahul (Australia, New Guinea and Tasmania). As expected, we detect European and East Asian admixture across most of the modern Aboriginal Australian groups included in the study. Finally, once we account for this recent admixture we find that genetics mirrors both geography and linguistics.

178

Ancestry-specific estimation of recent effective population size in the Americas

Sharon R Browning[1], Brian L Browning[1]

1. University of Washington, Seattle, WA, United States

Most existing methods for estimating effective population size over time have low accuracy for the recent past. In contrast, methods such as IBDNe that use segments of identity by descent (IBD) between individuals in a population-based sample can estimate recent effective population size as a function of time, with high accuracy for the past 3-200 generations.[1]

Many current-day American populations are admixtures of African, European and/or Native American ancestry, and the ancestry of an individual's genetic material can be estimated at each point the genome. Here we extend IBDNe to utilize these local ancestry calls in order to estimate the ancestry-specific recent effective population sizes for admixed populations. This allows one to estimate the past effective sizes of the ancestral African, European, and Native American populations, as well as the founding sizes at the time of colonization and the post-admixture effective population sizes.

We demonstrate the efficacy of our method on simulated admixed data, and we apply it to admixed American populations from the 1000 Genomes Project and to African American samples from two US cities. We estimate that the pre-colonization ancestral population sizes were 1-3 orders of magnitude larger than the effective population sizes immediately after colonization. In most cases, population sizes rebounded quickly after colonization. We also estimate that prior to the colonization events, the growth rates of the Native American ancestral populations were 2-4% per generation, which are similar to the growth rates estimated for the European and African ancestral populations over the same time periods.

1. Browning, SR and BL Browning (2015). Accurate Non-parametric Estimation of Recent Effective Population Size from Segments of Identity by Descent. AJHG 97: 404-418.

179

Human population structure and genetic adaptation in the Himalayan region

Elena Arciero[1], Thirsa Kraaijenbrink[2], Asan .[3], Marc Haber[1], Qasim Ayub[1], Mark Jobling[4], George van Driem[5], Yali Xue[1], Peter de Knijff[2], Chris Tyler-Smith[1]

1. The Wellcome Trust Sanger Institute, Hinxton, United Kingdom

2. MGC Department of Human and Clinical Genetics, Leiden University Medical Centre, Leiden, The Netherlands

3. BGI-Shenzhen, Shenzhen, China

4. Department of Genetics, University of Leicester, Leicester, United Kingdom

5. Institut für Sprachwissenschaft, University of Bern, Bern, Switzerland

The Himalayas in South Asia provide a diversity of environments for humans, some of which have required substantial genetic adaptation. There is, however, little understanding of the genetic history of the Himalayans and how culture, geography and genetic selection have shaped Himalayan genomes. In this study, we are using a combination of population-genetic and functional analyses to explore these topics. We analysed ~600,000 genome-wide SNPs in 948 Himalayan individuals from 49 different autochthonous groups from Nepal, Bhutan, North India and Tibetan Plateau in China. We find that the Himalayan populations share a component derived from a common ancestral population, followed by the development of local fine genetic structure correlating with language and geographical distribution. High altitude adaptation phenotype seems to have originated in a single ancestral population and spread across several areas in Himalaya. Genetic signatures of adaptation to high altitude are observed in *EPAS1* and *ATP6V1E2*. We find signatures of adaptation to low altitude within the *TRIM67* region, possibly associated with anti-microbial activity

in the tropical forest. We are exploring functional validation of variants in the *EPAS1* region using 1000 Genomes Project lymphoblastoid cell lines with and without the Denisovan introgressed haplotype under both normoxia and hypoxic conditions and analysing expression of candidate hypoxia responsive genes in order to characterize the response of cell lines with the introgressed haplotype. We are also investigating whether this haplotype confers resistance to the DNA damage experienced at high altitude through induction of DNA strand breakage and oxidative DNA damage in cultured cells.

180

Fine-scale identity-by-descent and birth records in Finland provide insights into recent population history

Alicia R Martin[1, 2], Sini Kirminen[3], Aki S Havulinna[4], Antti-Pekka Sarin[3], Aarno Palotie[1, 3, 2], Veikko Salomaa[4], Samuli Ripatti[3], Matti Pirinen[3], Mark J Daly[1, 2]

1. Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA, United States

2. Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, Boston, USA

3. Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, Finland

4. National Institute for Health and Welfare (THL), Helsinki, Finland

Finland provides unique opportunities to investigate both population and medical genomics because of its adoption of unprecedented uniformity in national electronic health records, concerted coordination of research centers across the country, detailed historical records, as well as recent population bottlenecks that drove specific disease alleles to high frequency. We investigate recent population history (up to ~50 generations ago), particularly relevant to rare, disease-conferring alleles, using identity-by-descent (IBD) haplotype sharing in >10,000 Finns. We compare IBD sharing in Finland to nearby Scandinavian countries with considerably different population histories, including >8,000 Swedes and >30,000 Danes. We find drastically more sharing on average in Finns, including many long tracts. By leveraging fine-scale birth record data, we find a non-linear decay of pairwise IBD sharing with increasing distance across Finland. This arises from pockets of excess IBD sharing; e.g. pairs of individuals from northeast Finland share on average several-fold more of their genome IBD than pairs from southwest regions containing the major cities of Turku and Helsinki. We demonstrate inference of recent migration patterns from IBD sharing patterns. For example, high IBD sharing in northeast Finland radiates from north to south rather than to the west, indicating a coastal wave of migration. We also investigate recent effective population size changes across regions of Finland and find evidence supporting the distinction between early and late settlement areas. However, our results indicate a more continuous flow of migration than previously posited, with a minimum N_e occurring ~12 generations ago in the northernmost Lapland region and moving further back in time to the south, with a bottleneck detectable in the early settlement area ~40 generations ago. Lastly, we leverage IBD sharing for genetic disease mapping and show that rare, functional haplotypes show more significant association via IBD mapping than single variants with linear mixed effect models.

181

Demographic history and population structure of chimpanzees (P*an troglodytes*) with implacations for global conservation strategies

Peter Frandsen[1, 2], Marc de Manuel Montero[3], David D. Hughes[3], Claudia Fontsere[3], Jessica Hernandez[3], Vitor Sousa[4], Martin Kuhlwilm[3], Chimpanzee Diversity Consortium, Laurent Excoffier[4], Aylwyn Scally[5], Yali Xue[6], Chris Tyler-Smith[6], Hans Redlef Siegismund[1], Christina Hvilsom[2], Tomas Marques-Bonet[3, 7]

1. Copenhagen University, Copenhagen, COPENHAGEN, Denmark

2. Copenhagen Zoo, Copenhagen, Denmark

3. Institut de Bilogia Evolutiva (CSIC/UPF), PRBB, Barcelona, Spain

4. Institute of Ecology and Evolution, University of Bern, Bern, Switzerland

5. Department of Genetics, University of Cambridge, Cambridge, United Kingdom

6. Welcome Trust Sanger Institute, Cambridge, United Kingdom

7. CNAG/CRG, Barcelona, Spain

Large-scale population genomics have recently provided novel insight to the diversity and evolutionary history of the great apes, including our own closest living relative, the chimpanzee. However, limited by a lack of precise geographical information, our knowledge on the local demographic history and fine-scale population structure of chimpanzees is still incomplete. Such knowledge is crucial when setting future conservation strategies for chimpanzees, both *in situ* and *ex situ*.To fill this knowledge gap, we have analyzed a comprehensive dataset of 60 wild born chimpanzee genomes, covering all four subspecies sampled across their natural distribution range. From this project, we present an unpreceded fine scale inference of complex demographic histories and a tight link between geography and local layers of genetic population structure. Apart from valuable insights to the local evolutionary past of the chimpanzee, these findings have allowed us to identify ancestry informative markers (AIM). We will show how this panel of AIMs can be used to re-assign confiscated individuals to their geographical origin, demonstrating a novel tool to combat illegal trafficking of chimpanzees along with the means to provide an accurate genetic guidance for global *ex situ* conservation management plans.

182

Rapid identification of phylogenetically informative data from high-throughput sequencing reads

Rachel S Schwartz[1], Reed A Cartwright[1]

1. Arizona State University, Tempe, ARIZONA, United States

Datasets for phylogenetics have grown dramatically in recent years. However, while large datasets contain a lot of information, they also contain noise. The challenge for phylogenomics is to extract information from large datasets rapidly and efficiently. We have developed easy-to-use, open-source software called SISRS (Site Identification from Short Read Sequences) to identify such data from raw reads, and have demonstrated the success of this approach. SISRS assembles a composite reference genome consisting primarily of loci that are conserved across species. Aligning reads to this genome and calling genotypes results in a large dataset of phylogenetically informative sites. We have evaluated approaches to generate the composite genome and thereby identify phylogenetically informative regions of the genome. To-date SISRS has been overly conservative in calling sites in order to avoid downstream effects of erroneous base calls due to error in sequencing and alignment. This approach results in significant loss of information. We have evaluated approaches to jointly genotype samples given read information to produce a larger number of accurately called genotypes. Additionally, we discuss the potential for including read information to jointly call the genotypes and phylogeny. By identifying conserved yet variable loci directly from raw sequence data, we can provide accurate alignments for phylogenetic analysis

at any taxonomic level. New approaches to rapidly identify these loci and the sequence of each locus for each species will allow us to generate accurate, well-resolved phylogenies, particularly for non-model organisms lacking reference genomes.

183

RADpainter and fineRADstructure: population inference from RAD-seq data

Milan Malinsky[1, 2], Emiliano Trucchi[3], Daniel J Lawson[4], Daniel Falush[5]

1. University of Cambridge, Cambridge, United Kingdom

2. Wellcome Trust Sanger Institute, Cambridge, United Kingdom

3. University of Vienna, Vienna, Austria

4. University of Bristol, Bristol, United Kingdom

5. University of Swansea, Swansea, Wales, United Kingdom

Understanding of shared ancestry in genetic datasets is almost always key to their interpretation. The fineSTRUCTURE package (Lawson et al., 2012) represents a powerful model-based approach to investigating population structure using genetic data. It offers especially high resolution in inference of recent shared ancestry, as evidenced for example in its application to investigation of genetic structure of the British population (Leslie et al., 2015). The high resolution of this method derives from utilizing haplotype linkage information and from focusing on the most recent coalescence (common ancestry) among the sampled individuals to derive a "co-ancestry matrix" - a summary of nearest neighbor haplotype relationships in the dataset. Further advantages when compared with other model-based methods (e.g. STRUCTURE and ADMIXTURE) include the ability to deal with a very large number of populations, explore relationships between them, and to quantify ancestry sources in each population.

The existing pipeline for co-ancestry matrix inference was designed to meet the needs of analyzing large scale human genetic SNP datasets, where chromosomal location of the markers are known and haplotypes are typically assumed to be correctly phased. Therefore, these methods have so far been inaccessible to users without high quality genome-wide haplotypes. With a boom in non-model organism genomics, there is a pressing need to bring these approaches to communities without access to such data.

Here we present RADpainter, a program designed specifically to infer the co-ancestry matrix from RAD-seq data, taking full advantage of its unique features. We package this new program together with the fineSTRUCTURE MCMC clustering algorithm into fineRADstructure - a complete, easy to use, and fast population inference package for RAD-seq data (https://github.com/millanek/fineRADstructure).

1. Lawson, D. J., Hellenthal, G., Myers, S. & Falush, D. Inference of population structure using dense haplotype data. PLoS Genet. (2012).
2. Leslie, S. et al. The fine-scale genetic structure of the British population. Nature 519, 309–314 (2015).

184

Envision: A computational tool for predicting mutational effect magnitudes

Vanessa Gray[1], Ronald J HAUSE[1], Douglas M Fowler[1], Jay Shendure

1. University of Washington, Seattle, WASHINGTON, United States

Current computational predictors for mutational effect focus on the binary consequences of mutations, *e.g.,* deleterious or not. Recently, technological advances have afforded high-throughput methods to quantify mutational effects on protein function. Here, we leverage large-scale mutagenesis data sets comprising tens of thousands of quantitative mutational effect scores for several proteins and protein domains to train a computational tool for predicting mutational effect scores. Our tool, Envision, was trained using gradient boosting machine learning and uses evolutionary conservation, biochemical, and structural annotations to predict both categorical and quantitative effects of single amino acid mutations. Envision is highly accurate both for classification and regression in 10-fold cross validation. We validated Envision in several ways, including on large-scale mutagenesis data not included in model training and on other mutational databases like the Protein Mutant Database. In all cases, we find that Envision outperforms other predictors, except when those predictors were trained on the testing data in question.

185

Addicted? Reduced host resistance in populations with defensive symbionts

Rodrigo Cogni[1], Julien Martinez[2], Francis Jiggins[2]

1. Universidade de São Paulo, São Paulo, SP, Brazil

2. Genetics, Cambridge University, Cambridge, UK

Heritable symbionts that protect their hosts from pathogens have been described in a wide range of insect species. By reducing the incidence or severity of infection, these symbionts have the potential to reduce the strength of selection on genes in the insect genome that increase resistance. Therefore, the presence of such symbionts may slow down the evolution of resistance. Here we investigated this idea by exposing *Drosophila melanogaster* populations to infection with the pathogenic *Drosophila* C virus (DCV) in the presence or absence of *Wolbachia*, a heritable symbiont of arthropods that confers protection against viruses. After nine generations of selection, we found that resistance to DCV had increased in all populations. However, in the presence of *Wolbachia* the resistant allele of *pastrel*—a gene that has a major effect on resistance to DCV—was at a lower frequency than in the symbiont-free populations. This finding suggests that defensive symbionts have the potential to hamper the evolution of insect resistance genes, potentially leading to a state of evolutionary addiction where the genetically susceptible insect host mostly relies on its symbiont to fight pathogens.

186

Both co-evolution with protein binding-partners and RNA structural plasticity influences the evolution of bacterial RNA regulators.

Michelle M Meyer[1], Betty L Slinger[1], Shermin Pei[1]

1. Boston College, Chestnut Hill, MA, United States

In bacteria, regulation in response to environmental cues is mediated by many mechanisms. These include a variety of specific interactions between complex tertiary structures formed by the mRNA and a variety of cellular signals. Despite their complexity, such RNA regulators appear to be rapidly evolving and are often not conserved across broad taxonomic groups. There are several documented instances where analogous biological functions are regulated by very distinct RNA regulators in diverse bacterial species. Our laboratory is utilizing the structured RNA cis-regulators that respond to ribosomal proteins as a model to explore the factors contributing the diversity of analogous RNA regulators observed. These regulators typically act by interacting with a protein to inhibit translation initiation and allow stoichiometric production of ribosome components. We have focused on ribosomal protein S15 and its interactions with four different mRNA structures that occur in diverse bacterial species. We find that the diversity of mRNA structures apparent across different bacteria phyla is partially driven by differences in the recognition surfaces of diverged S15 homologs. However, *in vitro*selection experiments have also demonstrated that such regulators are relatively frequent in sequence space. A population of RNAs selected to interact with a single S15 homolog is incredibly diverse. High-throughput sequencing revealed a population of >4 million distinct sequences with few easily recognizable motifs apparent. Furthermore, 50% of sequences tested not only interact with S15 *in vitro*, but also allow regulation of an *Escherichia coli* reporter in response to S15, in some cases with characteristics comparable to the native regulator. These experiments suggest that the diversity of RNA regulators observed in nature is a consequence of both co-evolution between diverging protein homologs and their corresponding RNA regulators, as well as the plasticity of RNA structure that allows many potential solutions to exist within sequence space.

187

A new realistic codon model for genome-scale positive selection analysis with variation in DNA constraints and in selection

Iakov I Davydov[1, 2], Nicolas Salamin[1, 2], Marc Robinson-Rechavi[1, 2]

1. University of Lausanne, Lausanne, VAUD, Switzerland

2. Swiss Institute of Bioinformatics, Lausanne, Switzerland

The fixation of mutations in genes is due to a balance of selection, mutation and drift. Codon models have proven very useful in distinguishing selection, including positive selection, from drift. Synonymous substitution rates are assumed to capture all variation that is not under selection, and thus the ratio of non synonymous (dN) to synonymous (dS) substitutions should indicate selection. There are many models allowing selection (and thus dN) to vary across the gene, but dS is assumed to be constant over all positions of one gene. Yet significant variation of dS has been observed inside genes.

We have developed a simple new model which takes into account variations in mutation rate in addition to selection levels. We combine this both with site variation in selection (M8 of PAML) and with branch-site variation in selection. Thus we present the first integration of dS variation with episodic selection, thanks to the fact that our model is more computationally efficient than previous efforts to capture dS variation. We use our improved positive selection models to scan genome-scale data. We show that the new model provides a better fit on the real data. We present different effects of the gene sequence constraints on the performance of the codon models in mammals vs bacteria.

188

Virological Factors That Increase the Transmissibility of Emerging Human Viruses

Jemma L Geoghegan[1], Alistair M Senior[1], Francesca Di Giallonardo[1], Edward C Holmes[1]

1. University of Sydney, Sydney, NSW, Australia

The early detection of pathogens with epidemic potential is of major importance to public health. Most emerging infections result in dead-end 'spill-over' events in which a pathogen is transmitted from an animal reservoir to a human but is unable to achieve the sustained human-to-human transmission necessary for a full-blown epidemic. It is therefore critical to determine why only some virus infections are efficiently transmitted among humans while others are not. We sought to determine which biological features best characterized those viruses that have achieved sustained human transmission. Accordingly, we compiled a database of 203 RNA and DNA human viruses and used an information theoretic approach to assess which of a set of key biological variables were the best predictors of human-to-human transmission. The variables analysed were: taxonomic classification; genome length, type and segmentation; the presence or absence of an outer envelope; recombination frequency; duration of infection; host mortality; and whether or not a virus exhibits vector-borne transmission. This comparative analysis revealed multiple strong associations. In particular, we determined that viruses with low host mortality, that establish long-term chronic infections, that are non-segmented, non-enveloped and, most importantly, not transmitted by vectors, were more likely to be transmissible among humans. In contrast, variables including genome length, genome type and recombination frequency had little predictive power. In sum, we have identified multiple biological features that seemingly determine the likelihood of inter-human viral transmissibility, in turn enabling general predictions of whether viruses of a particular type will successfully emerge in human populations.

189

Thaumarchaeotal Metalloenzymes: Evolutionary History, Influence of Oxygen and Role in Nitrogen and Greenhouse Gas Cycles

Jennifer B Glass[1], Jessica A Kozlowski[2], Lisa Y Stein[2], Martin G Klotz[3, 4], Frank J Stewart[1], Christopher L Dupont[5], Gregory P Fournier[6]

1. Georgia Institute of Technology, Atlanta, GA, USA

2. University of Alberta, Edmonton, Alberta, Canada

3. Queens College of the City University of New York, Flushing, NY, USA

4. Xiamen University, Xiamen, China

5. J. Craig Venter Institute, La Jolla, CA, USA

6. Massachusetts Institute of Technology, Cambridge, MA, USA

Thaumarchaeota are the most globally significant archaea in the nitrogen cycle, and one of the most numerically abundant phyla today. The first thaumarchaeote to be isolated was a marine ammonia-oxidizing archaeon (AOA; Könneke, et al. 2005). Although thaumarchaeotal catabolism requires molecular oxygen, AOA retain activity at low oxygen levels, and peak in abundance at oxic-anoxic interfaces where ammonia and oxygen are available, albeit at low concentrations. AOA produce elevated levels of gaseous N compounds (NO, a toxic yet vital metabolic intermediate, and N_2O, a potent greenhouse gas) under oxygen depletion (Kozlowski et al., 2016), although the mechanism of archaeal N_2O

production remains debated. The discovery of numerous genes encoding copper-containing enzymes in AOA genomes is particularly intriguing because ammonia-oxidizing bacteria (AOB), which commonly dominate in ecosystems with plentiful ammonia, rely on a multitude of iron-based proteins for their metabolism. It has thus been hypothesized that AOA evolved after the Great Oxidation Event (~2.4 billion years ago) and a concurrent rise in environmental copper levels under more oxidizing conditions (Klotz & Stein, 2008). Many unanswered questions remain about the timing of marine AOA evolution and the consequences of AOA diversification and ocean oxygenation for nitrogen and greenhouse gas cycling. This talk will address AOA metabolism, phylogeny, and evolution in the context of changing seawater iron and copper availability over Earth history, as well as depth-and oxygen-dependent correlations between iron and copper concentrations, and environmental genes and transcripts encoding archaeal iron and copper proteins, in modern seawater samples (Glass et al., 2015). *This research was supported by the NASA Astrobiology Institute Alternative Earths team (NNA15BB03A).*

1. Glass JB, Kretz CB, Ganesh S, Ranjan P, Seston SL, Buck KN, Morton PL, Landing WM, Moffett JW, Giovannoni SJ, Vergin KL, Stewart FJ. 2015. Meta-omic signatures of microbial metal and nitrogen cycling in marine oxygen minimum zones. Front Microbiol 6: 998.
2. Klotz MG, LY Stein. 2008. Nitrifier genomics and evolution of the nitrogen cycle. FEMS Microbiol Lett 278: 146-156.
3. Könneke M, Bernhard AE, de la Torre JR, Walker CB, Waterbury JB, Stahl DA. 2005. Isolation of an autotrophic ammonia-oxidizing marine archaeon. Nature 437: 543-546.
4. Kozlowski JA, M Stieglmeier, C Schleper, MG Klotz, LY Stein. 2016. Pathways and key intermediates required for obligate aerobic ammonia-dependent chemolithotrophy in bacteria and Thaumarchaeota. ISME J doi:10.1038/ismej.2016.2.

190

5. Mapping the Physiological and Metabolic Adaptations to Oxygen across the Archaea

6. **Gregory Fournier[1], Abigail Caron[1]**
7. *1. Massachusetts Institute of Technology, Cambridge, MA, United States*
8. Over billions of years, Archaea have adapted to a wide diversity of environments on Earth. During this time, one of the greatest challenges and opportunities was the rise of molecular oxygen within the atmosphere. Several archaeal lineages have independently evolved aerobic metabolisms across the Crenarchaeota, Thaumarchaeota, and Euryarchaeota. While the most distinctive archaeal group, methanogens, have generally retained their anaerobic metabolism, this pathway has been lost several times, in the cases of Halobacteriales, Thermoplasmatales, and Archaeoglobales. Each of these lineages has a metabolism that either directly or indirectly depends on the presence of molecular oxygen within the environment, suggesting that these transitions were in response to rising oxygen levels. We map the history of oxygen tolerance across these groups via phylogeny of a gene encoding the superoxide dismutase (SOD) enzyme, which detoxifies free radical oxygen species. We find that this gene was independently acquired via HGT many times within Archaea, especially within aerobic groups. These acquisitions appear to be singular adaptive events, without subsequent transfers or losses after the initial acquisitions. We propose that these transfers document the rising levels of oxygen across environments, with the earliest acquisitions immediately following the Great Oxygenation Event, and subsequent, more recent transfers tied to higher levels of oxygen experienced since the Neoproterozoic. The evolutionary history of additional oxygen-associated gene families in Archaea will further assist in reconstructing this narrative of the oxygenation of the microbial world.

191

Evolution and global dissemination of the multidrug resistant Escherichia coli ST131 clone

Nouri Ben Zakour[1], Areej Alsheikh-Hussain[1], Nhu TK Nguyen[1], Melinda Ashcroft[1], Leah Roberts[1], Brian Forde[1], Mitchell Stanton-Cook[1], Mark A Schembri[2], Scott A Beatson[1]

1. Australian Centre for Ecogenomics & School of Chemistry and Molecular Biosciences, The University of Queensland, St Lucia, Australia

2. School of Chemistry and Molecular Biosciences & Australian Infectious Diseases Centre, The University of Queensland, St Lucia, Australia

Escherichia coli sequence type 131 (ST131) is predominantly associated with urinary tract and bloodstream infections, and was first reported in the United Kingdom in 2008. Since then, it has rapidly disseminated worldwide to become the most frequently isolated fluoroquinolone-resistant (FQR) *E. coli*clone. *E. coli* ST131 is strongly associated with several factors, including resistance to fluoroquinolone, high virulence gene content, the possession of the type 1 fimbriae FimH30 allele and the production of the CTX-M-15 extended-spectrum beta-lactamase. Yet, the sequence of events leading to the rapid emergence and successful spread of this multi-drug resistant clone remains largely undetermined, mainly due to the lack of geographical and temporal diversity of strain collections studied so far.

In order to reconstruct the evolutionary scenario of the emergence of FQR *E. coli* ST131, we combined publically available genomic data of 188 ST131 strains, spanning the years 1967 to 2011 and from 9 geographical regions. We investigated the ST131 clonal structure and identified the genetic changes that define the global phylogeny of *E. coli* ST131. Contrary to some initial reports, we confirmed the defining role that recombination has played in shaping the evolution of the distinct lineages. Thanks to the increased temporal resolution achieved by our combined dataset, divergence time estimation using Bayesian Evolutionary Analysis by Sampling Trees (BEAST) could be performed and revealed that the multidrug resistant ST131 lineage originated around 1987, coincidently following the first clinical usage of fluoroquinolone in 1986. Geographical ancestral reconstruction also suggested that FQR ST131 most likely emerged in North America before spreading globally. Finally, we were able to identify key strains harboring intermediate states leading to more resistant lineages. Taken altogether, we propose that following the progressive acquisition of mobile genetic elements and recombination events, which likely increased the virulence and fitness of ST131, point mutations conferring resistance to fluoroquinolone were the pivotal events leading to a rapid population expansion in 1990's to early 2000's.

This work highlights the challenges and rewards of combining publicly available genomic datasets and how this approach can synergistically provide better resolution into the series of events leading to the emergence of highly successful multi-drug resistant bacteria.

192

Genome-scale rates of evolutionary change in bacteria

Sebastian Duchene[1], Kathryn E Holt[1], Francois-Xavier Weill[2], Simon Le Hello[2], Jane Hawkey[1], David J Edwards[1], Matthieu Fourment[3], Edward C Holmes[3]

1. University of Melbourne, Melbourne, VIC, Australia

2. Institut Pasteur, Paris, France

3. University of Sydney, Sydney, NSW, Australia

Evolutionary methods can be used to infer the time scale of groups of organisms. In some pathogens it is possible to estimate the time of origin of infectious outbreaks or of disease emergence. The accuracy of these estimates, however, relies on our understanding of the rate at which genetic change accumulates over time. The recent surge of genomic data presents an unprecedented opportunity to improve our understanding of pathogen evolution, with the potential of improving future inferences of their evolutionary time scale. We estimated the rate of evolution for 36 complete genomes of different bacterial pathogens using a range of computational methods. We find large differences in the rates. For example, some bacteria, such as those that causes hospital-derived infections, evolve much faster than those that cause tuberculosis, which undergo extended periods of latency. We investigated characteristics of the bacteria that may explain variation in their rates. We find that genome size, genome composition, and the sampling time appear to play an important role in determining their rate. Our results provide the first genomic perspective of bacterial rates of evolution, thereby improving our understanding of the time scale over which they evolve.

193

Phylogenomic networks reveal limited phylogenetic range of recent lateral gene transfer by transduction

Ovidiu Popa, Giddy Landan, Tal Dagan[1]

1. University of Kiel, Kiel, SH, Germany

Bacteriophages are recognized DNA vectors and transduction is considered as a common mechanism of lateral gene transfer (LGT) during microbial evolution. Anecdotal events of phage-mediated gene transfer were studied extensively, however, a coherent evolutionary viewpoint of LGT by transduction, its extent and characteristics, is still lacking. Here we report a large-scale evolutionary reconstruction of transduction events in 3,982 genomes. We inferred 17,158 recent transduction events linking donors, phages and recipients into a phylogenomic transduction network view. We find that LGT by transduction is mostly restricted to closely related donors and recipients. Furthermore, a substantial number of the transduction events (9%) are best described as gene duplications that are mediated by mobile DNA vectors. We propose to distinguish this type of paralogy by the term *autology*. A comparison of donor and recipient genomes revealed that genome similarity is a superior predictor of species connectivity in the network in comparison to common habitat. This indicates that genetic similarity, rather than ecological opportunity, is a driver of successful transduction during microbial evolution. A striking difference in the connectivity pattern of donors and recipients shows that while lysogenic interactions are highly species-specific, the host range for lytic phage infections can be much wider, serving to connect dense clusters of closely related species. Our results thus demonstrate that DNA transfer via transduction occurs within the context of phage-host specificity, but that this tight constraint can be breached, on rare occasions, to produce long range LGTs of profound evolutionary consequences.

194

Cellular anthropology: using stem cell models to explore human development and evolution

Joanna Wysocka[1]

1. Stanford University School of Medicine, Stanford, CA, United States

While studies in model organisms have led to great progress in unveiling the conserved mechanisms of gene regulation, many aspects of development that are unique to humans and other primates remain unexplored, as are regulatory principles underlying emergence of human-specific traits. I will discuss some of our recent progress in understanding transcriptional mechanisms governing human development and evolution, such as those involving the activity of transposable elements in early embryogenesis or our recent quantitative analyses of cis-regulatory divergence in the human and chimpanzee neural crest, an embryonic cell population that is most relevant for evolution of human craniofacial form.

195

Interplay of cis and trans mechanisms driving transcription factor binding, chromatin, and gene expression evolution

Emily S Wong[1, 2], Bianca Schmitt[3], John Marioni[1], Duncan Odom[3], Paul Flicek[1]

1. European Bioinformatics Institute (EMBL-EBI), Cambridge, UK

2. University of Queensland, St Lucia, QLD, Australia

3. Cancer Resesarch UK, University of Cambridge, Cambridge, UK

Noncoding regulatory variants play a central role in the genetics of human diseases and in evolution. To elucidate the regulatory mechanisms underlying transcription factor (TF) binding variations in mammals, we measured allele-specific TF binding affinity of three liver-specific TFs between crosses of two inbred mouse strains. We classified over 45,000 binding events to one of four regulatory categories: conserved/non-differential-acting, cis, trans, both cis and trans. Our results highlight the dominance of additively-inherited cis-driven variation in TF occupancy variation. Trans-acting variations are most often dominantly inherited. Cis-acting variants lead to local coordination of TF occupancies that decays with distance and distal coordination which may be modulated by long-range chromatin contacts. Our results reveal the interlinking regulatory mechanisms that interplay to drive TF occupancy, chromatin state, and gene expression in complex mammalian cell states.

196

Transcription Factor Evolution as a Mechanim for Modifying Developmental Gene Regulatory Networks.

Veronica Hinman[1], Greg Cary[1], Alys Jarvela[2, 1], Rene Francolini[1]

1. Biology, Carnegie Mellon U., Pittsburgh, PENNSYLVANIA, United States

2. Biology, U. Maryland, Baltimore, Maryland, USA

It is well documented that gene regulatroy networks can evolve extensively through mutations to cis-regulatory modules. Transcription factor proteins that bind these cis-regulatory modules may also evolve to produce novelty. Such coding changes, however, are considered to be more

rare, because transcription factors are highly pleiotropic and hence are more constrained to evolve in ways that will not produce widespread detrimental effects. Recent technological advances have unearthed a surprising variation in DNA-binding abilities, such that individual transcription factors may recognize both a preferred primary motif and an additional secondary motif. This provides a source of modularity in function. In this talk, we will present recent work that shows that orthologous transcription factors can also evolve a changed preference for a lower affinity secondary binding motif, thereby offering an unexplored mechanism for GRN evolution. We demonstrate that this difference may allow for greater evolutionary change in timing of regulatory control and provide a mechanism through which organisms can evolve a changed response to signaling gradients. This uncovers a layer of transcription factor binding divergence that could exist for many pairs of orthologs and provides a mechanism through which transcription factors can evolve to produce subtle changes in phenotype.

197

Genomic basis of the evolution of wing pigmentation in Drosophila

Mathilde PARIS[1], Anaïs Landais[1], Emma Lemoine[1], caroline minervino[1], Philippe Veber, Benjamin Prud'homme[1]

1. CNRS, MARSEILLE, BOUCHES DU RHONE, France

The past 20 years of comparative developmental genetic studies have shown that blueprints of animal forms are encoded in complex gene regulatory and signaling networks made of evolutionary conserved components. Morphological evolution is thought to largely result from modifications in the interactions among components of genetic networks, rather than from the diversification of gene repertoires or gene functions. However it remains unclear how novel morphological traits were gained, and how the underlying complex genetic regulatory network appeared. The evolution of a pigmentation spot located on the wing of some Drosophila males is a great model to tackle this question, notably because it is genetically tractable and has evolved several times in the melanogaster group. Using a comparative and integrative approach our goal is to map on the drosophila tree the genomic changes that caused changes in the genetic regulatory network of wing pigmentation, which in turn led to the evolution of this morphological trait. In order to identify genes whose changes in spatio-temporal expression underlie the evolution of wing pigmentation, we have compared gene expression using RNA-seq in 3 different species and between sexes, at 11 time points spanning wing pupal development and in a spatially resolved manner. In order to understand the genetic determinism of these gene expression changes we have employed functional genomics methods such as ChIP-seq and FAIRE. We are testing the involvement of a dozen of candidates in the evolution of wing spot formation by using a combination of CRISPR/Cas9-targeted mutagenesis, RNAi knockdown and overexpression assays in pupae of the spotted species *D. biarmipes*.

198

Controlling for phylogenetic relatedness improves discovering the genomic basis underlying species' phenotypic differences

Michael Hiller[1]

1. Max Planck Institute of Molecular Cell Biology and Genetics & Max Planck Institute for the Physics of Complex Systems, Dresden, Saxony, Germany

The growing number of sequenced genomes allows us now to address a key question in genetics and evolutionary biology: What is the genomic basis that underlies phenotypic differences between species? We developed a computational framework called Forward Genomics that associates phenotypic to genomic differences by focusing on phenotypes that are repeatedly lost in independent lineages. Here, we present two new Forward Genomics methods that (i) control for the phylogenetic relatedness between the species of interest, (ii) control for differences in evolutionary rates and (iii) compute the significance of the association between phenotypic and genomic differences. We systematically compare these methods on simulated and on real data and demonstrate that the new methods significantly improve the sensitivity to detect such associations.

We use these methods to discover genomic loci that underlie the degeneration of the visual system in blind subterranean mammals. This genome-wide screen identifies many loci that are enriched in functions related to eye development and the perception of light as well as loci associated with eye diseases in human. In addition, we find genomic loci with a function in the circadian rhythm, which might be an adaptation to the subterranean environment.

The Forward Genomics framework has broad applicability to many other phenotypic differences. The new methods presented here significantly advance our ability to discover the genomic basis underlying phenotypic differences between species, which will contribute our understanding of how nature's phenotypic diversity has evolved.

199

Polygenic basis for ecological adaptation in the Atlantic herring revealed by genome sequencing

Leif Andersson[1]

1. Uppsala University, Uppsala, UPPSALA, Sweden

The Atlantic herring (*Clupea harengus*) is one of the most abundant vertebrates on earth and constitutes a huge biomass in the North Atlantic. We have generated a high-quality draft assembly of the herring genome and carried out whole genome sequencing of more than 20 pooled populations samples (50 fish in each) and of individual fish from certain localities. The results revealed nearly identical allele frequencies among subpopulations in the East Atlantic implying minute amount of genetic drift and possibly gene flow between subpopulations. However, highly significant differentiation at 1-2% of all SNPs was found. We identified almost 500 independent loci associated with a recent niche expansion from marine (Atlantic Ocean) to brackish waters (Baltic Sea), and more than 100 independent loci showing genetic differentiation between spring- and autumn-spawning populations. Interestingly, autumn-spawning Baltic herring from the Baltic Sea show extensive haplotype sharing with a spring-spawning Baltic herring at loci associated with adaptation to the Baltic Sea but extensive haplotype sharing with autumn-spawning Atlantic herring from the North Sea although these has been classified as different subspecies. Thus, different ecotypes of herring carry different combinations of adaptive haplotypes. The results show that both coding and non-coding changes contribute to ecological adaptation. Haplotype blocks, often spanning multiple genes and maintained by selection, are associated with genetic differentiation.

200

The genetic architecture of age at maturity in 57 Atlantic salmon populations: a large-effect locus with sex dependent dominance reduces sexual conflict and shows signals of local adaptation

Nicola J Barson[1], Tutku Aykanat[2], Kjetil Hindar[3], Matthew Baranski[4], Geir H Bolstad[3], Peder Fiske[3], Celeste Jacq[4], Arne J Jensen[3], Susan E Johnston[5], Sten Karlsson[3], Matthew Kent[1], Thomas Moen[6], Eero Niemelä[7], Torfinn Nome[1], Tor Naesje[3], Panu Orell[7], Atso Romakkaniemi[7], Harald Saegrov[8], Kurt Urdal[8], Jaakko Erkinaro[7], Sigbjorn Lien[1], Craig R Primmer[2]

1. Norwegian University of Life Sciences, Ås

2. Department of Biology, University of Turku, Turku, Finland

3. NINA, Norway

4. Nofima, Norway

5. Edinburgh University, UK

6. Aquagen, Norway

7. Finnish Natural Resources Institute, Finland

8. Rådgivande Biologer, Norway

Males and females share many traits that have a common genetic basis, however selection on these traits often differs between the sexes leading to sexual conflict. Under such sexual antagonism, theory predicts the evolution of genetic architectures that resolve this sexual conflict. Yet, despite intense theoretical and empirical interest, the specific loci underlying sexually antagonistic phenotypes have rarely been identified, limiting our understanding of how sexual conflict impacts genome evolution and the maintenance of genetic diversity. Here, we identify a large effect locus controlling age at maturity in Atlantic salmon, an important fitness trait in which selection favours earlier maturation in males than females, and show it is a clear example of sex dependent dominance reducing intralocus sexual conflict and maintaining adaptive variation in wild populations. Using high density single nucleotide polymorphism (SNP) data across 57 wild populations and whole genome re-sequencing, we found that the vestigial-like family member 3 gene (*VGLL3)* exhibits sex-dependent dominance in salmon, promoting earlier and later maturation in males and females, respectively. There were also signs of spatially varying selection consistent with selection towards local optima. *VGLL3*, an adiposity regulator associated with size and age at maturity in humans, explained 39% of phenotypic variation, an unexpectedly large proportion for what is usually considered a highly polygenic trait. Such large effects are predicted under balancing selection from either sexually antagonistic or spatially varying selection. Our results provide the first empirical example of dominance reversal permitting greater optimisation of phenotypes within each sex, contributing to the resolution of sexual conflict in a widespread evolutionary trade-off between age and size at maturity in Altantic salmon. They also provide key empirical evidence for how variation in reproductive strategies can be maintained over large geographical scales.

201

Population genomics of *Paramecium* species

Parul Johri[1], Tom G Doak[1], Sascha Krenek[2], Georgi K Marinov[1], Michael Lynch[1]

1. Indiana University, Bloomington, IN, United States

2. Institute of Hydrobiology, Technische Universität Dresden, Dresden, Germany

Population-genomic analyses are essential to understanding factors shaping genomic variation and lineage-specific sequence constraints. The dearth of such analyses for unicellular eukaryotes prompted us to assess variation in *Paramecium*, one of the most well-studied ciliate genera. The a*urelia*complex consists of ~15 morphologically indistinguishable species that diverged subsequent to two rounds of whole-genome duplications (WGDs), as long as 320 MYA, and are well known for their streamlined genomes. We examine patterns of polymorphism by sequencing whole genomes of 10-12 worldwide isolates of each of three species belonging to the *Paramecium aurelia* complex: *P. tetraurelia*, *P. biaurelia*, *P. sexuarelia,* and two outgroup species that do not share the WGDs: *P. caudatum* and *P. multimicronucleatum*. An apparent absence of strong global geographic population structure suggests continuous or recent dispersal of *Paramecium* over long distances. Introns and intergenic regions are highly constrained relative to 4-fold degenerate sites, more so in species with smaller intergenic regions. Nuclear genome diversity is reduced up to ~100-150 bp both upstream and downstream of genes, suggesting the presence of densely packed regulatory modules. Comparison of sequence variation at non-synonymous and synonymous sites suggests similar recent selective pressures on paralogs within and orthologs across the deeply diverging species and allows identification of possible candidates of duplicate genes that might be undergoing non-functionalization. This study serves as a first attempt at a genome-wide population-genomic analysis in *Paramecium*, and provides a valuable resource for future studies in evolutionary and functional genetics in ciliates.

202

Inferring population dynamics from high throughput genomic analysis of *Plasmodium falciparum* field samples

Sha Zhu[1], Gil McVean[1], Pf3k consortium

1. Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, United Kingdom

Genetic variation within the malarial parasite *Plasmodium falciparum* affects key phenotypes including drug resistance and risk of severe disease. Advances in technology and experimental protocol mean that obtaining high coverage genome sequencing data from routine blood samples taken in the field is now possible. However, interpreting such data is difficult because of high rates of mixed infections and highly variable data quality.

To provide a framework for analysing genetic variation in *P. falciparum*, the Pf3k project is working to build a global reference map of genome sequence and tools that can enable rapid analysis of data from field samples, with 2,512 samples available to date.

Here, we describe and validate methods for inferring the structure and identity of strains present in a sample by combining a reference panel of known haplotypes with data from an additional sequencing experiment. In particular, we describe Monte Carlo methods for inferring haplotypes present in a sample that generalise techniques developed for diploid samples, but which can cope with multiple strains and the over-dispersion of allele counts that results from experimental protocol. The approach is validated through analysis of experimentally generated mixed samples.

When applied to the Pf3k data, our approach demonstrates substantial variation in local parasite population dynamics. For example, we find that while 642 out of 934 cases from Asia present evidence of infection by a single parasite strain, this the case for only 657 out of 1490 cases from Africa. Moreover, we find evidence for substantial within-continent variation in population structure, indicative of epidemiological heterogeneity and the effects of drug-induced selection pressure.

Our results demonstrate the feasibility of inferring genome-wide patterns of haplotype structure in malarial parasites taken from clinical field samples and establish a resource for driving the development of new approaches for integrating population genetic and epidemiological modelling.

203

Genomic correlates of inbreeding depression in outcrossing Caenorhabditis nematodes

Luke M Noble[1], Victoria V Cattani[1], Max R Bernstein[1], Jia Shen[1], Annalise B Paaby[2], Matthew V Rockman[1]

1. Biology, New York University, New York, NY, United States

2. Biology, Georgia Institute of Technology, Atlanta, GA, United States

Population size, structure and demographic history are central, complicating factors in prediction of inbreeding depression, the decline in fitness commonly seen with increasing homozygosity. The effects of inbreeding have been extensively studied, largely theoretically, and supporting data within and across species has not always conformed to expectation. We made highly replicated measurements of inbreeding depression for multiple isolates of globally distributed Caenorhabditis species. Selfing has evolved repeatedly in this genus, yet inbreeding depression is extremely strong in outcrossing species and it is unclear how the short-term benefits of reproductive assurance have been accessed. From genomes assembled and annotated *de novo*, we estimated effective population size, and fixed and segregating loads from variant impact prediction. Some of these estimates are highly predictive of inbreeding depression across morphologically similar, yet deeply diverged, species.

204

Investigating the evolution of new biochemical pathways in baker's yeast *Saccharomyces cerevisiae*

Åsa Pérez-Bercoff[1], Tonia L. Russell[2], Philip J. L. Bell[3], Paul V. Attfield[3], Richard J. Edwards[4]

1. School of Biotechnology and Biomolecular Sciences, University of New South Wales, Sydney, NSW, Australia

2. Ramaciotti Centre for Genomics, University of New South Wales, Sydney, NSW, Australia

3. Microbiogen Pty Ltd, Sydney, NSW, Australia

4. School of Biotechnology and Biomolecular Sciences, University of New South Wales, Sydney, NSW, Australia

Understanding how new biochemical pathways evolve in a sexually reproducing population is a complex and largely unanswered question. We have successfully evolved a novel biochemical pathway in yeast using a sex based population approach.

For over 30 years, wild type *Saccharomyces* has been widely reported to not grow on xylose at all, but we discovered that most strains can grow, albeit at almost undetectable rates. A mass mated starting population of *Saccharomyces cerevisiae* strains was evolved under selection on Xylose Minimal Media (XMM) with forced sexual mating every ~two months for 1463 days. This produced a population that could grow on xylose as a sole carbon source. Initial studies show the xylose growth trait is quantitative and presumably governed by many genes. To investigate the evolution of the xylose phenotype, a xylose utilising strain MBG11a was isolated. MBG11a was sequenced with PacBio RSII long read sequencing at the Ramaciotti Centre for Genomics. A high quality complete genome was assembled *de novo* using the hierarchical genome-assembly process (HGAP3) using only PacBio non-hybrid long-read SMRT sequencing data, corrected using Quiver, and compared to the genome of the *S. cerevisiae* S288C reference genome.

Approximately 98.5% of the MBG11a genome could be aligned to S288C at 99.5% sequence identity, with over 15,000 non-synonymous and 200 nonsense SNP differences. We have crossed MBG11a with a reference wild type yeast strain (X2180 gal2, Xyl-) and are testing offspring on different minimal media in an attempt to identify MBG11a variants responsible for the novel growth phenotype.

Understanding what has occurred in the evolving yeast population, and how the yeast genome adapted under the selection pressures is of broad interest as it allows experimental analysis of how novel complex biological functions can evolve in an organism.

205

Co-functioning between horizontally transferred segments reveals exaptation in the evolution of new metabolic phenotypes in E. coli

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

1. Heinrich Heine University Duesseldorf, Düsseldorf, NRW, Germany

Bacterial adaptation proceeds predominantly through horizontal gene transfers (HGTs). Often, the evolution of a new phenotype (such as the ability to grow on a new set of nutrients) requires the acquisition of multiple DNA segments. It is unlikely that these are acquired via HGT simultaneously. Thus, the first HGT event is either neutral and becomes beneficial only after the second HGT event, or it provides a benefit by itself in another environment and thus represents an exaptation for the second adaptation. In this work, we studied the metabolic networks of 53 *E. coli* strains and their ancestors. Using flux balance analysis to define metabolic phenotypes, we found that (i) in contrast to expectations from a neutral model of HGT, new phenotypes acquired along a single phylogenetic branch are less likely to require the acquisition of multiple DNA segments than expected in a neutral model; (ii) new phenotypes acquired relative to a more distant ancestor often relied on multiple HGT events spread over different phylogenetic branches; and (iii) this split of the acquisition of adaptive gene sets over several phylogenetic branches was more frequent than expected from a neutral model of HGT. We conclude that exaptation is widespread in metabolic adaptation of *E. coli* strains.

206

Dietary and Environmental Factors Shaping African Gut Microbiomes

Meagan A Rubel[1], Matthew EB Hansen[2], Aubrey Bailey[3], Kyle Bittinger[3], William Beggs[2], Alessia Ranciaro[2], SImon R Thompson[2], FD Bushman[3], Sarah A Tishkoff[2, 4]

1. Department of Anthropology, University of Pennsylvania, Philadelphia, PA, USA

2. Department of Genetics, University of Pennsylvania, Philadelphia, PA, USA

3. Department of Microbiology , University of Pennsylvania, Philadelphia, PA, USA

4. Department of Biology, University of Pennsylvania, Philadelphia, PA, USA

African populations have adapted to a range of environments and foods as they spread through the continent, and their gut microbiomes (GMs) may have co-evolved with them. The GMs from sixty ethnically diverse Tanzanians practicing different subsistence strategies (Hadza hunter-gatherers, Burunge agriculturalists, Maasai pastoralists, recently-settled Sandawe hunter-gatherers) with urban American individuals were sequenced using the V1/V2 regions of 16s rRNA from fecal samples to interrogate potential covariates shaping GM composition including traditional diet, quantification of gut parasites, geography, nutritional and ethnographic surveys, and genetic data from an Illumina 5M Omni SNP Array.

IBS relatedness for 1,066 Africans, including microbiome-sequenced individuals, was calculated from SNP array data and visualized using Multi-Dimensional Scaling plots of genetic similarity, and showed that Hadza are genetically differentiated from other Tanzanian ethnic groups. Bacterial compositional analysis showing that Hadza have lower within-group diversity than other populations indicates that they have a strong within-group bacterial composition. The Sandawe had higher-within group diversity and similar bacterial composition in heatmap analyses to Burunge, which may relate to their recent settlement and adoption of agriculture. Principal-coordinate analysis of bacterial families revealed that Tanzanians have two predominant bacterial gradients associated with broad global gut enterotypes: A strong Prevotellaceae-Ruminococcaceae gradient and a weak Bacteroidales-Ruminococcaceae gradient. Bacteroidales is associated with diets high in protein and fats, whereas Prevotellaceae and Ruminococcaceae are associated with plant and fiber rich diets, which affirms expectations for populations eating rural, nonwestern foods. The Hadza are enriched in Treponema, which may assist with nutrient extraction from fibrous plants. Maasai show variation in Prevotella relative abundance and an enrichment of Bacteroidales, which may assist with digestion of dairy and meat. This represents one of the largest GM studies to date of Africans, including the first pastoralist GM results, and provides novel microbiome data from sparsely characterized African groups.

207

Population and sex-specific transcriptomics of east Australian *Drosophila*

Marina Telonis-Scott[1], Jeremy RB Newman[2], Lauren M McIntyre[2, 3], Carla M Sgro[1]

1. *School of Biological Sciences, Monash University, Clayton, VICTORIA, Australia*
2. *Department of Molecular Genetics and Microbiology, University of Florida, Gainesville, Florida, USA*
3. *Department of Statistics, University of Florida, Gainesville, Florida, USA*

Gene expression divergence is widespread and thought to be crucial for environmental adaptation, despite largely uncharted genome-wide fitness consequences. General patterns using population transcriptomics however can provide broader insights into gene expression evolution in heterogeneous environments. *Drosophila* populations sampled from clinal gradients have been used to infer variation in gene regulation likely maintained by spatially varying selection, but so far little attention has been payed to the contribution of sex despite pervasive sexually dimorphic expression. Using RNAseq, we address the importance of sex-specific expression between populations that are thought to differ in ecology. The eastern Australian temperate-tropical latitudinal gradient is an excellent resource to study intraspecific local adaptation given the diverse climates, clines in thermal tolerance, fitness and morphology traits, and gene expression. We utilized the well-established 'cline-end' sampling strategy to survey gene expression in tropical and temperate *D. melanogaster* reared at 25°C. We discuss gene expression, alternative isoform expression and sequence variation between males and females from climatically diverse populations along the same environmental gradient in the context of climatic adaptation.

208

protTrace: Predicting the evolutionary traceabilities for proteins and pathways

Arpit Jain[1], Arndt Prof von Haeseler[2], Ingo Prof Ebersberger[1]

1. *Goethe University, Frankfurt Am Main, HESSEN, Germany*
2. *CIBIV, University of Vienna, Vienna, Austria*

The identification and functional characterization of functional protein networks concentrates on few and often only distantly related model organisms. Integrating these individual insights into comprehensive picture of molecular and functional evolution is typically subject of downstream bioinformatics analyses. Here gene sets of broad and phylogenetically diverse species collection are screened for homologs to previously described pathway components. The resulting phylogenetic profiles serve as basis for inferences regarding distribution of the corresponding pathways across species, their evolutionary origins and functional diversification. Typically ignored during this interpretation is the circumstance that sensitivity of homolog searches decrease as function of evolutionary time. Gaps in phylogenetic profiles especially for distantly related or fast evolving species may therefore either represent genuine absence of the corresponding proteins or an artefact of limited sensitivity in homolog search. Here we introduce the concept of evolutionary traceability to facilitate informed interpretation of phylogenetic profiles. We present a framework to compute for a protein and an evolutionary time the probability to detect homolog by means of significant sequence similarity if it's present. Specifically, we simulate protein sequence change over time considering each protein's specific constraints on evolutionary process. We monitor decay of similarity to the original sequence and determine the time when a significant similarity is no longer detected. Repeating the simulation 1,000 times and fitting a logistic growth curve to the observed data obtains then for each protein a detection probability distribution over time. Mapping this information onto a species tree determines for any protein of interest whether or not sequence similarity is likely to suffice for homolog identification in a given species. We have exemplified our approach by tracing the evolution of *S. cerevisiae* genome across the three domains of life. Further, we resolved the error bias in estimating gene age with reliable interpretations of phylogenetic profiles.

209

Term pregnancy in marsupials is homologous to implantation in eutherian mammals: a hypothesis

Oliver W Griffith[1], Stella A Protopapas[1], Jamie Maziarz[1], Wagner P Gunter[1]

1. *Department of Ecology and Evolutionary Biology, Yale University, West Haven, CONNECTICUT, United States*

The evolution of live birth in vertebrates requires the formation of a placenta, which is an organ formed by the apposition of maternal and embryonic tissues to support the exchange of materials between mother and embryos during pregnancy. Whilst the first mammals were egg laying, live birth evolved in the common ancestor of marsupial and eutherian mammals approximately 170 mya. In this common ancestor pregnancy is believed to have been short, with mothers giving birth to small, under-developed young. This mode of reproduction has been retained in marsupial mammals. To understand the genetic processes that support the formation of the placenta in marsupials we measured transcriptome wide gene expression in uterine tissues of non-, mid-, and late-pregnant grey short-tailed opossums (*Monodelphis domestica*). On this data we performed differential gene

expression and gene ontology analysis. We found that during late gestation there was a significant over representation of genes involved in inflammation, which is an important component of the implantation pathway of eutherians. We also identified several key genes involved in eutherian implantation that were highly expressed in late pregnant uterine tissue. Using immunohistochemistry we show that these changes occur in the uterine epithelium, and hence at the maternal-fetal interface. Together our data suggest that the late gestation phase of pregnancy in opossums is homologous to the implantation phase of pregnancy in eutherian mammals. Further, we suggest that the processes that facilitate implantation in eutherians, may have had fundamentally different roles when they were co-opted to pregnancy in early viviparous mammals.

210

Geographic patterns of virulence in a stickleback – tapeworm system

Agnes Piecyk[1], **Martin Kalbe**[1]

1. *Max Planck Institute for Evolutionary Biology, Ploen, Germany*

Species interactions in its extreme form of reciprocal co-evolution of hosts and parasites generate and maintain biodiversity. We aimed at a better understanding of potentially different co-evolutionary trajectories of a vertebrate host, the three-spined stickleback *Gasterosteus aculeatus*, and its naturally infecting cestode *Schistocephalus solidus* on a large geographic scale. *S. solidus* highly specifically infects *G. aculeatus* as intermediate host and undergoes almost its complete somatic growth in the stickleback, sometimes even exceeding the fish's weight. The parasite's size serves as a proxy for virulence and the growth is assumed to be an adaptive trait. Using controlled experimental infections, we studied the infection phenotype of different sympatric and allopatric host-parasite combinations. Naïve lab-bred sticklebacks from a German and a Norwegian reference population were individually infected with single tapeworm larvae from ten different localities across the Northern Hemisphere. The infection success, the parasite's size and fish condition and immunological parameters were determined to estimate virulence. We found clusters of worm populations with similar levels of virulence that correlate with geographic areas and with the parasite's phylogeny. *S. solidus* strains grew generally larger in German fish, however, within each host population, we found similar clusters. These results emphasize that the parasite's growth can better be linked to genetic clustering of *S. solidus* and ecological characteristics of each habitat than to latitude or geographical distance between the source populations. In order to learn more about the molecular mechanistic basis of this interplay, we are currently analyzing expression levels of stickleback immune genes in response to infection with the different parasite strains.

211

Snakes, plugs and mating balls: Telomere dynamics in red-sided garter snakes

Nicky Rollings[1], **Heather L Waye**[2], **Randolph W Krohmer**[3], **Emily J Uhrig**[4], **Robert T Mason**[4], **Mats Olsson**[5], **Camilla Whittington**[1], **Chris R Friesen**[1]

1. *University of Sydney, Camperdown, NSW, Australia*
2. *University of Minnesota, Morris, Minnesota, USA*
3. *Saint Xavier University, Chicago, Illinois, USA*
4. *Oregon State University, Corvallis, Oregon, USA*
5. *Department of Biological & Environmental Sciences, University of Gothenburg, Gothenburg, Sweden*

Sexual conflict may arise when male and female individuals of a species have different reproductive strategies. These different strategies mean the sexes experience different costs and trade-offs and it may be possible to observe these differences at a molecular level. Telomeres are short tandem repeats found at the ends of chromosomes that shorten with stress and growth. The rate of shortening is often correlated with the lifespan of an organism. We therefore chose to investigate the telomere dynamics of a species that experiences sexual conflict, the red-sided garter snake *Thamnophis sirtalis parietalis*. Male garter snakes emerge from hibernation with high levels of corticosterone, are aphagous during the mating season, and invest all of their time in trying to mate. Conversely, females often only reproduce every second year, and invest more energy into maintaining a stable thermal environment in order to produce fit offspring. As males invest more highly into reproduction while females prioritise self-maintenance we would predict that male snakes would experience more telomere loss than females. We investigated this by determining the ages of individuals using skeletal chronology and used qPCR to determine blood telomere length. Telomere length decreased with age in male garter snakes but remained stable in female snakes. Furthermore, we assessed the relationship between blood and sperm telomere length in male garter snakes, finding a moderate correlation (R^2 = 0.285). This suggests that older male garter snakes may pass shorter telomeres on to their offspring, potentially affecting their fitness.

212

Ancient DNA evidence supports a 'Founder-Takes-All' model for spatial genetics.

Jonathan Waters[1]

1. *University of Otago, Dunedin, New Zealand*

Density-dependent processes often play a key role in the spatial structuring of genetic variation. In particular, related processes such as gene surfing and high-density blocking can generate striking geographic contrasts in the distributions of genes. Well-characterised phylogeographic patterns associated with postglacial recolonization, progressive island colonization, microbial sectoring, and the 'Out of Africa' pattern of human expansion, are fundamentally similar, and arguably underpinned by a 'founder takes all' density-dependent principle. In the current study, comparisons of ancient DNA and modern DNA from New Zealand's pinniped and penguin assemblages reveal sudden spatio-temporal genetic shifts, apparently in response to human-mediated extirpation events. These rapid phylogeographic transitions underscore the role of 'founder takes all' processes in constraining the distributions of genes and species.

213

What is the role of within-host evolution in *Staphylococcus aureus* infection?

Daniel J Wilson[1]

1. *University of Oxford, Oxford, OXON, United Kingdom*

Is the body's natural flora capable of evolving over time to become more virulent? Could evolvable virulence explain why some people suffer invasive bacterial infections while others do not? Despite its notoriety as a dangerous hospital-associated pathogen, *Staphylococcus aureus* is a commonly carried commensal, found in the noses of 30% of adults. From the perspective of the bacteria, invasive disease occurs rarely compared to asymptomatic carriage. In previous work, we investigated the evolutionary dynamics of nasal carriage, and discovered that in one long-term nasal carriage population an excess of protein-truncating substitutions was associated with the transition to a life-threatening invasive blood stream infection. Here, we report results of population-based and molecular studies we are conducting into the role of genetics and within-host evolution in the progression of invasive *Staphylococcus aureus* disease.

214

Genome-wide association study of carriage versus invasive disease in *Neisseria meningitidis*

Sarah G Earle[1], Holly B Bratcher[2], Odile B Harrison[2], Dorothea MC Hill[2], Paula Kriz[3], Christoph M Tang[4], Martin CJ Maiden[2], Daniel J Wilson[1, 5]

1. Nuffield Department of Medicine, University of Oxford, Oxford, UK

2. Department of Zoology, University of Oxford, Oxford, UK

3. National Reference Laboratory for Meningococcal Infections, National Institute of Public Health, Prague, Czech Republic

4. Sir William Dunn School of Pathology, University of Oxford, Oxford, UK

5. Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK

The human pathogen *Neisseria meningitidis*, the meningococcus, is a major cause of bacterial septicaemia and meningitis worldwide. However, *N. meningitidis* is principally a commensal. Although the commensal nature of the majority of meningococcal infections is well understood, the factors promoting the transition of asymptomatic carriage to invasive disease remain to be fully elucidated. Host factors such as carriage state, complement deficiencies, social behavior, and geographic location are associated with increased disease risk. Colonization with hyperinvasive meningococci is also a major risk factor for invasive disease, but the bacterial genetic mechanisms underlying invasiveness are not well understood.

Here, we investigated the genetic basis of carriage versus invasive disease in 261 isolates of *N. meningitidis* by a genome-wide association study (GWAS), applying methods we adapted to bacteria to capture both lineage-associated differences and locus-specific effects on phenotype. Associations were tested between carriage versus disease and both SNPs and the presence/absence of 31bp "kmers".

We found significant associations at variants within genes involved in the synthesis of the polysaccharide capsule, a well established virulence factor and a major component contributing to the survival of the bacteria in the blood stream and cerebrospinal fluid, within possible phase variable regions, plus other regions representing potentially novel virulence factors. This has important implications for understanding virulence in *N. meningitidis*, and why only some strains cause disease. Association studies of this kind have the potential to provide insight into identifying virulent strains, and could further provide candidate targets to assist approaches in treating and preventing meningococcal disease.

215

Profile-based comparison of *Salmonella* genomes reveals signatures of invasive potential

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

1. University of Canterbury, Christchurch, CANTERBURY, New Zealand

2. Institute for Molecular Infection Biology, University of Wurzberg, Wurzberg, Germany

Comparative genomics investigations have yielded an abundance of information on the genetic variation between organisms, however understanding the consequences of this variation for protein function has proved challenging. We present a profile HMM-based method for assessing the functional significance of mutations in protein coding sequences. We demonstrate an application of the method to comparative analysis of bacterial genomes to scan for functionally significant genetic variation. As a model system, we have chosen the well-studied species *Salmonella enterica*, of which several lineages are known to have undergone adaptation to an invasive infection style, associated with a narrowing of host tropism with a concurrent accumulation of pseudogenes. We show that our method is able to detect functional degradation of genes associated with host-adapation that are not detected by conventional pseudogene analysis, and suggest that this approach offers a sensitive measure of the loss-of-function mutations that may occur as a result of adaptation to a new niche.

216

Genomic analysis of adaptation during chronic colonization in *Helicobacter pylori*

Sean Schneider[1]

1. Fred Hutchinson Cancer Research Center, Seattle, WASHINGTON, United States

Helicobacter pylori is a human-specific bacterial pathogen that infects and colonizes the stomach for decades. *H. pylori* causes chronic gastritis in hosts, that in 10-20% of cases will progress to severe disease including peptic ulcers and/or gastric cancers. Because *H. pylori* is thought to infect ~50% of humans in the world and 33% in the US, it is a major global cause of gastric cancers. My project uses whole genome sequencing to investigate the processes that *H. pylori* uses to adapt to new hosts. I compare the results of two different adaptation experiments: 1) Adapting the bacterium to a colonize the stomach of mice in a laboratory setting (wild type is not capable of long term infection in mice). *H. pylori* is serially passaged in mice for increasing colonization loads. Mice are infected with wild type bacteria as well as several mutants defective for recombination mechanisms. *H. pylori* typically has very high rates of recombination and this study examines whether recombination aids in adaptation to a new host. 2) Examining long term adaptation to a human host. Patient J99 was infected with *H. pylroi* and samples collected by endoscopy, refused treatment, and came back to the clinic 6 years later and had many further samples collected from specifically labeled anatomic regions. I have sequenced dozens of strains from this collection from the different time points and regions. This study examines differences in adaptation resulting from colonizing different niches of the gastric anatomy. Comparing this approach with the mouse adaptation study gives insights into how different environments affect adaptation in this organism.

217

Insertion sequence dynamics in the global *Shigella sonnei* population

Jane Hawkey[1, 2, 3], **Helen Billman-Jacobe**[2], **Kathryn E Holt**[1, 3]

1. Department of Biochemistry and Molecular Biology, Bio21 Molecular Science and Biotechnology Institute, The University of Melbourne, Parkville, VIC, Australia

2. Department of Agriculture and Food Systems, The University of Melbourne, Parkville, VIC, Australia

3. Centre for Systems Genomics, The University of Melbourne, Parkville, VIC, Australia

Insertion sequences (IS) are small, transposable elements, commonly found in bacterial genomes. They are amongst the most dynamic components of bacterial genomes and can drive evolution through functional changes (interrupting or up-regulating genes); mediating gene deletion; or creating structural variation. The *Shigella sonnei* reference genome 53G contains over 200 different IS insertions. Whole genome phylogenetic analysis shows that there are three common lineages of *S. sonnei*. Variation caused by IS is often ignored in such analyses as these regions are difficult to extract from short read data. We developed a tool, ISMapper, to detect IS from short read data and used it to screen a global collection of *S. sonnei* isolates for the presence of 12 different IS. We found that each lineage has a unique IS profile and distinct IS dynamics. We used ancestral state reconstruction to infer the timing of IS gain and loss events on the tree. We estimate that the common ancestor of *S. sonnei* (circa 1669) carried 129 IS insertions, and that lineages 2 and 3 each accumulated ~120 additional IS by the mid 19th century and are now saturated at ~300 IS per genome. Lineage 1 accumulated only ~20 additional IS by the mid 19th century, but is now undergoing IS expansion, with current isolates still accumulating IS at a rate of ~0.3 copies/genome/year. Additionally, we detected hotspots for IS insertions, and showed that IS are major contributors to gene inactivation which has played a significant role in *S. sonnei* evolution.

218

Ancient leprosy genomics: Retracing the evolutionary history of *Mycobacterium leprae* from medieval genomes

Verena J Schuenemann[1, 2], **Charlotte Avanzi**[3], **Alexander Seitz**[4], **Ben Krause-Kyora**[5], **Alexander Herbig**[6], **Andrej Benjak**[3], **Jesper L Boldsen**[7], **G. Michael Taylor**[8], **Pushpendra Singh**[3], **Sarah Inskip**[9], **Simon Mays**[10], **Helen D Donoghue**[11], **Sonia Zakrzewski**[12], **Kay Nieselt**[4], **Stewart T Cole**[3], **Johannes Krause**[1, 2, 6]

1. Institute for Archaeological Sciences, University of Tuebingen, Tuebingen, BW, Germany

2. Senckenberg Center for Human Evolution and Paleoenvironment, University of Tuebingen, Tuebingen, Germany

3. Global Health Institute, Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland

4. Center for Bioinformatics, University of Tuebingen, Tuebingen, Germany

5. Institute of Clinical Molecular Biology, Kiel University, Kiel, Germany

6. Max Planck Institute for the Science of Human History, Jena, Germany

7. Department of Anthropology (ADBOU), Institute of Forensic Medicine, University of Southern Denmark, Odense S, Denmark

8. Department of Microbial and Cellular Sciences, Faculty of Health and Medical Sciences, University of Surrey, GU2 7XH Guildford, UK

9. Department of Archaeology, Leiden University, Leiden, The Netherlands

10. Historic England, Portsmouth, UK

11. Centre for Clinical Microbiology, Division of Infection and Immunity, University College London, London NW3 2PF, UK

12. Department of Archaeology, University of Southampton, Southampton, UK

Ancient DNA studies can provide new perspectives for evolutionary history of pathogens such as for *Mycobacterium leprae*, the causative agent of leprosy, one of the oldest recorded and most feared diseases in human history, which was prevalent in Europe until the 16th century and is still endemic in many countries with over 200,000 new cases reported annually.

Previously we observed an exceptional DNA preservation of *M. leprae* in medieval skeletons, that enabled us to successfully reconstruct a late medieval leprosy genome by *de novo* assembly, thus offering the prospect to retrace *M. leprae's* pre-historic origin. Furthermore, the analysis of medieval *M. leprae* genomes pointed to a pre-medieval origin of most contemporary human and armadillo leprosy lineages and suggested a prevalence of two distinct lineages in medieval northwestern Europe.

Here we analyzed novel medieval *M. leprae* genomes from different time points and geographic locations including the so far oldest *M. leprae* genome derived from one of the earliest known cases of leprosy in the UK, a skeleton from the Great Chesterford cemetery dated to 415–545 AD, in order to reconstruct the last 1500 years of *M. leprae's* evolutionary history. A phylogenetic comparison revealed the contemporary presence of at least 4 distinct lineages and suggests a high diversity of *M. leprae* strains in medieval Europe. In addition, the 1500-year-old Great Chesterford genome allowed us to trace one of the lineages, lineage 3, back to the 6th century.

These results develop and refine previous models for the geographic distribution of *M. leprae* lineages in the past indicating a higher complexity and point out the necessity of studying ancient *M. leprae* strains to understand the history of leprosy worldwide.

219

Tracing the aftermath of the Black Death through analyses of ancient genomes.

Kirsten Bos[1], **Maria Spyrou**[1], **Rezeda Tukhbatova**[2], **Joachim Wahl**[3], **Ilgizar R Gazimzyanov**[4], **Alexander Herbig**[1], **Johannes Krause**[1]

1. Max Planck Institute for the Science of Human History, Jena, THURINGIA, Germany

2. Laboratory of Paleoanthropology & Paleogenetics, Kazan Federal University, Kazan, Russian Federation

3. Department of Archeological Sciences, University of Tuebingen, Tuebingen, Germany

4. Institute of Archaeology named after A. Kh. Khalikov, Tatarstan Academy of Sciences, Kazan, Russian Federation

Yersinia pestis infections have had a long history with humans, with earliest confirmed cases dating back as far as the Bronze Age. The mid-14th century Black Death is the most famous of these outbreaks, claiming anywhere from 30 to 50% of the European population in only five short years.

Evidence is accumulating that reveals the presence of extinct daughter populations of the Black Death in Europe as the cause of subsequent epidemics until the mid 18th century. In addition, it has recently been suggested that one of these daughter populations traveled east, settled in Southeast Asia, and gave rise to modern plague lineages that have a near worldwide distribution. Genomic data from post Black Death outbreaks are essential to determine the paths traveled by the pathogen after the Black Death, and to determine the potential sources for European epidemics that persisted until the Early Modern Era.

Here we present *Y. pestis* genomes from 14th century Tatarstan, Russia and 16th century Ellwangen, Germany. Together these genomes reveal important steps along the path traveled by the Black Death, and support the notion of an extinct European reservoir of plague that persisted for several centuries.

220

Building a bigger brain: the genetic basis for the human neocortex

Gregory Wray[1]

1. Duke University, Durham, NORTH CAROLINA, United States

Among the many morphological changes that took place during human origins, an expansion in brain volume is conspicuous. My research group use functional genomic approaches to identify candidate loci underlying human-specific features of brain anatomy and a combination of mouse models and induced pluripotent stem cells to validate and understand the functional consequences of mutations within those loci. We recently discovered a novel enhancer containing human lineage-specific mutations that drive elevated expression of the Wnt receptor FZD8 within neural progenitor cells during early corticogenesis, decreasing their cell cycle time, and increasing cell number and overall brain volume. We have also identified changes in lipid metabolism in adipocytes that result in increased production of diacylglycerides essential for the greatly expanded surface area of neural- and glial cell membranes in the human brain. These results, along with those from other groups, highlight the complex interdependence of changes in development and physiology that underlie the evolution of a bigger brain.

221

Genomic insight into the evolution of larval body plans

William L Hatleberg[1], **Bernard M Degnan**[1], **Sandie M Degnan**[1]

1. School of Biological Science, The University of Queensland, Brisbane, QLD, Australia

The majority of animal species transition between two morphologically and ecologically distinct body plans: a free-swimming, dispersive larva and a benthic adult. Although this biphasic lifecycle is widely distributed across nearly all metazoan phyla, pointing to an ancient synapomorphy, each lineage is characterized by lineage- and taxon-specific morphological adaptations. This provides an excellent comparative framework for the evolution of novelty in animal body plans. Using developmental transcriptome data from the sponge *Amphimedon queenslandica* as a foundation for lifecycle evolution, we examined how biphasy is orchestrated on a genomic level. We find that the *Amphimedon* lifecycle is regulated by distinct suites of larval- and adult-specific genes that complement a core set of highly expressed genes that are shared between larval and adult body plans. Moreover, we find that novel poriferan-specific innovations are enriched in the adult transcriptome, while larval and juvenile stages are characterized by the up-regulation of well-conserved metazoan developmental genes. By extending our analyses to the developmental transcriptomes of five eumetazoan phyla, we show that genes differentially expressed in each taxon are largely characterized by lineage-specific co-expression modules, which are driven by both novel and well-conserved transcription factors. We discuss the implications of these results in the context of larval biology and historical embryological theory to provide a new genomically-informed hypothesis for animal lifecycle evolution.

222

Evolution of modulatory regulatory programs across different tissues in cichlids

Tarang K Mehta[1], **Wiktor Jurkowski**[2], **Sara A Knaack**[3], **Christopher Koch**[4], **Sushmita Roy**[3, 4, 5], **Federica Di Palma**[1]

1. Vertebrate and Health Genomics, The Genome Analysis Centre (TGAC), Norwich, Norfolk, UK

2. Integrative Genomics, The Genome Analysis Centre (TGAC), Norwich, Norfolk, UK

3. Wisconsin Institute for Discovery (WID), Madison, Wisconsin, USA

4. Dept. of Biostatistics and Medical Informatics, UW Madison, Madison, Wisconsin, USA

5. Dept. of Computer Sciences, UW Madison, Madison, Wisconsin, USA

In vertebrates, the East African cichlid radiations represent arguably the most dramatic examples of adaptive speciation. In the great lakes Victoria, Malawi and Tanganyika and within the last few million years, one or a few ancestral lineages of haplochromine cichlid fish have given rise to over 1500 species exhibiting an unprecedented diversity of morphological and ecological adaptations. Such explosive phenotypic diversification of East African cichlids is unparalleled among vertebrates and implies the rapid evolution of regulatory regions and networks underlying the traits under selection.

Comparative functional genomics, transcriptomics and epigenomics are powerful tools to study the evolution of tissue and species divergence. We recently developed *Arboretum*, an algorithm to identify modules of co-expressed genes across multiple species in a phylogeny. By integrating inferred modules with nucleotide variation, predicted *cis* regulatory elements and miRNA profiles from five East African Cichlids, we investigated the evolution of tissue-specific gene regulation. Our analyses identified modules with tissue-specific patterns for which we reconstructed the evolutionary gene regulatory networks across the five cichlids species. We report striking cases of rapid network rewiring for genes known to be involved in traits under natural and/or sexual selection such as jaw morphology (*dlx2a*), visual systems (*rho*) and pigmentation (*ednrb1a*). Our unique integrative approach that interrogates the evolution of regulatory networks allowed us to identify the rapid regulatory changes associated with certain traits under selection in cichlids.

223

Oxytricha's mobile genome

Xiao Chen[1], **Kelsi Lindblad**[1], **Laura Landweber**[1, 2]

1. Princeton University, Princeton, NJ, United States

2. Columbia University, New York, NY, United States

The ciliate *Oxytricha* is a microbial eukaryote with two nuclei. One nucleus is the differentiated product of a zygotic nucleus produced at mating. Hence Oxytricha possesses a dynamic pair of genomes inside a single cell. Massive DNA deletions and rearrangement produce a highly fragmented but transcriptionally active somatic genome from a complex germline zygotic genome that provides an archive. The differentiation

process eliminates nearly all noncoding DNA, including *all* transposons, and rearranges over 225,000 short DNA segments to produce a mature, somatic genome containing over 16,000 tiny gene-sized chromosomes. Essential to the rearrangement process are thousands of germline copies of telomere-bearing elements (TBEs), a class of Tc1/mariner transposons. We recently annotated more than 10,000 complete and 24,000 partial TBE sequences in the reference germline genome. Phylogenomic analysis reveals that they cluster into four major families, with a preference for either insertion into DNA segments that are retained in the somatic genome or their maintenance at such sites. Availability of a draft germline genome assembly for a second *Oxytricha*strain has allowed us to characterize TBE insertion sites that differ between the two strains. This has identified novel TBE insertions and suggests that all four TBE families may still be mobile. Many recent insertions are in close proximity to precursor somatic sequences, and several interrupt gene loci, which necessitates their deletion during development. These findings demonstrate that TBE transposon insertion actively contributes to DNA fragmentation during genome evolution in *Oxytricha*.

224

Rapid establishment of a piRNA based defense system – evidence from the P-element invasion dynamics in experimentally evolving *D. simulans* populations

Robert Kofler[1], Viola Nolte[1], Kirsten-Andre Senti[1], Christian Schloetterer[1]

1. VetMedUni Wien, Wien, --- SELECT A STATE ---, Austria

The P-element, one of the best understood eukaryotic transposable elements (TEs) recently invaded natural *D. simulans* populations. We captured a natural *D. simulans* population from Florida at an early stage of the invasion and set up a replicated experimental evolution study in hot and cold environments. This opens the unprecedented opportunity to study a natural invasion of a TE with the aid of high throughput sequencing technologies in replicated populations evolving in different environments.We show that in all replicate populations of a given environment the P-element rapidly spreads with a remarkable consistency. In the hot environment P-element copy numbers increased 16-fold up to generation 20 and attained a stable copy number of about 30 per haploid genome. No further increase could be noted during the next 40 generations of experimental evolution. By contrast, at cold conditions the speed of the invasion is much slower, the P-element multiplied 4-fold by generation 30. Interestingly, the P-element invasion only results in a modest reactivation of resident TE families in *D. simulans*.The leveling out of the P-element invasion at hot conditions could be due to i) a balance between transposition events and negative selection against TE insertions ii) non-autonomous truncated copies of the P-element, which have been shown to down-regulate transposition activity and iii) piRNAs. RNA-seq and small RNA-seq analysis argues that the dominant factor containing the spread of the P-element is the emergence of piRNAs complementary to the P-element.

225

piRNA vs TEs: genomic and epigenomic variation in 16 strains of mice

Eric Miska[1]

1. University of Cambridge, Cambridge, CAMBRIDGESHIRE, United Kingdom

Piwi-interacting RNAs (piRNAs) are a class of small non-coding RNAs. Their length ranges from ~22 to ~34 nucleotides depending on the organism and developmental stage. They have been identi ed in both vertebrate and invertebrate animals. To date, the main role of piRNAs is the silencing of transposable elements. In animals de cient in the piRNA pathway, transposons move freely leading to genome instability and sterility [1]. Interestingly, piRNAs have been shown to mediate transgenerational epigenetic memory [2].

In mice, most piRNAs originate from long precursor transcripts known as piRNA clusters. These transcripts eventually get processed into mature piRNAs by a mechanism still poorly understood [3].

To better understand piRNA biogenesis and their mechanisms of action, we are exploring the genetic and epigenetic variations in piRNAs between 16 di erent strains of mice [4]. Taking advantage of recently developed, reference-quality de-novo assemblies for those genomes, we are conducting a genome-wide comparative study and investigating the degrees of structural variation in piRNA-abundant regions. So far, small-RNA sequencing on four developmental stages (two embryonic and two post- partum) of these 16 strains has been performed. Results from our rst analyses will be presented. These will help in understanding the interplay between genomic variation and changes in piRNA populations between strains, developmental timepoints, and individuals. Future directions of this work will be discussed, as transgenerational e ects of piRNAs will be studied by performing crosses between some of the strains. This project is timely and therefore we anticipate that its completion will provide the community with a comprehensive resource.

1. Siomi et al., 'PIWI-Interacting Small RNAs'.
2. Bagijn, M. P., Goldstein, L. D., Sapetschnig, A., Weick, E.-M., Bouasker, S., Lehrbach, N. J., et al. (2012). Function, targets, and evolution of Caenorhabditis elegans piRNAs. Science (New York, N.Y.), 337(6094), 574–578. http://doi.org/10.1126/science.1220952
3. Li et al., 'An Ancient Transcription Factor Initiates the Burst of piRNA Production during Early Meiosis in Mouse Testes'.
4. Mouse genomic variation and its effect on phenotypes and gene regulation. (2011). Mouse genomic variation and its effect on phenotypes and gene regulation., 477(7364), 289–294. http://doi.org/10.1038/nature10413

226

Estimating seven coefficients of pairwise relatedness using population genomic data

Matthew S Ackerman[1], Parul Johri[1], Ken Spitze[1], Sen Xu[2], Thomas Doak[1], Kimberly Young[1], Michael Lynch[1]

1. Indiana University, Bloomington, IN, United States

2. University of Texas at Austin, Austin, TX

Population structure is described by genotypic correlation coefficients between individuals, the most basic of which are Jacquard's nine condensed coefficients. These correlation coefficients form the basis of quantitative-genetic analysis, and geneticists perform experimental crosses or pedigree analysis in order to recover them. Molecular techniques can be used to recover these coefficients between individuals with unknown relationships, but previously could only recover four of these coefficients at best. I have developed a method for recovering seven coefficients using the full set of biallelic loci derived from whole-genome sequences and a maximum-likelihood method. This approach should allow for more robust estimation of the components of genetic variance from population-genomic data, and is potentially very useful for conservation genetics.

Simulations show that the procedure is nearly unbiased, even at the minimally informative 3$\times$ coverage, and that errors in five of the seven coefficients are statistically uncorrelated. The sum of the remaining two coefficients provides an unbiased assessment of the overall correlation of

heterozygosity between two individuals. These methods have been applied to four populations of the freshwater crustacean *Daphnia pulex*, revealing several interesting characteristics that are not apparent with other techniques. The use of a maximum-likelihood method also allows us to assess statistical significance of relationships using a log-likelihood ratio test, and we find statistically significant negative estimates of many of these pair-wise relatedness coefficients. Although these coefficients are traditionally regarded as measures of identity probabilities, which cannot be negative, we treat them as measures of conditional association, which can be negative. These methods are implemented as part of an expansive package of maximum-likelihood programs for the analysis of population genomic data (mapgd) that I have implemented, which we hope will greatly enhance the power of such studies (available from https://github.com/LynchLab/MAPGD).

227

Decay of Accuracy of Genomic Prediction with Genetic Distance

David J Balding[1], Marco Scutari[2], Ian McKay[3]

1. University of Melbourne, University Of Melbourne, VIC, Australia

2. Oxford University, Oxford

3. NIAB, Cambridge

Statistical models used for the prediction of quantitative traits from high-density genomic data are usually tested using cross-validation, which implicitly assumes that new individuals (whose phenotypes we would like to predict) originate from the same population that the genomic prediction model is trained on. However in many settings models trained in one population are used to predict phenotypes in different populations. We investigate the decay of predictive accuracy as the genetic distance between the training and target populations increases. We do this using clustering and resampling to construct a sequence of target populations of increasing genetic distance from the training population. We find that the correlation between true and predicted values decays approximately linearly with respect to FST (or mean kinship) between the training and the target populations. We illustrate this relationship using data sets from mice, wheat and humans. In addition to analysis of real-world data, we apply our approach to a simulated multi-generation genomic selection experiment and to simulated phenotypes using worldwide human genome data.

228

A tutorial on how (not) to over-interpret STRUCTURE/ADMIXTURE bar plots

Daniel Falush[1], Lucy van Dorp[2], Daniel Lawson[3]

1. University of Swansea, Swansea, United Kingdom

2. University College London, London, United Kingdom

3. University of Bristol, Bristol, United Kingdom

Genetic clustering algorithms, implemented in popular programs such as STRUCTURE and ADMIXTURE, have been used extensively in the characterisation of individuals and populations based on genetic data. A successful example is reconstruction of the genetic history of African Americans who are a product of recent admixture between highly differentiated populations. Histories can also be reconstructed using the same protocol for scenarios where groups have not experience recent admixture, where recent genetic drift is strong or other scenarios that deviate in some way from the underlying inference model. Unfortunately, such histories can be highly misleading. We have implemented a "differential palette" visualization of the fineSTRUCTURE coancestry matrix which facilitates easy comparison with STRUCTURE/ADMIXTURE bar plots and assessment of how good a fit the admixture model is for the dataset as a whole and for particular individuals and groups. Combining these complementary analyses with additional methods such as supervised clustering that are designed to test specific hypothesis should allow more robust analysis of recent demographic history based on genetic data.

229

Hill-Robertson Interference in the Genomes of Wild Mice, *Mus musculus castaneus*

Tom R Booker[1], Peter D Keightley[1]

1. The University of Edinburgh, Edinburgh, SCOTLAND, United Kingdom

Understanding how and why genetic diversity varies between species and across the genome are questions at the heart of population genetics. Studies in multiple eukaryotic species have shown that nucleotide diversity is reduced close to conserved, functional elements. This variation in nucleotide diversity is consistent with models of Hill-Robertson Interference, suggesting that natural selection has played a role in generating the observed patterns. It has, however, proven difficult to distinguish between two mechanisms that can lead to variation in the amount of nucleotide diversity: background selection due to selection against deleterious mutations and recurrent selective sweeps of advantageous variants. Using whole genomes of *Mus musculus castaneus* individuals sampled from the species' ancestral range we estimate the strength of selection acting on protein coding genes and conserved non-coding elements and construct a fine-scale recombination map. Using these we hope to tease apart the contribution of positive and negative selection to the patterns of nucleotide diversity across the genome.

230

Weird animal genomes, epigenetics and sex chromosome turnover

Jenny Graves[1, 2]

1. School of Life sciences, La Trobe University, Melbourne, Victoria, Australial

2. Institute of Applied Science, University of Canberra, Canberra, ACT, Australia

In humans and other mammals with XX females and XY males, the Y bears a gene (*SRY*) that induces testis differentiation in the embryo and switches on hormones that masculinize it. This is a very stable sex determining system, but comparison between humans and distantly related mammals reveals that it evolved fairly recently and has undergone major change. Birds, snakes and even monotreme mammals also have stable sex chromosome systems, but the sex chromosomes and sex determining genes quite different. Other retiles, amphians and fish have a great variety of sex determining genes and chromosomes as well as environmentally determined sex, revealing rapid turnover in many taxa. Here I will

discuss the evolution of novel sex determining genes and the differentiation of the chromososomes that bear them into differentiated XY or ZW pairs. I will explore what happens to ex-sex chromosomes, and how sex chromosome turnover might impact vertebrate evolution.

231

Mapping the Eukaryote Chromosome: From Primary Constriction to Monia Gap.

Gregory B Peters[1], **Chris Moran**[2]

1. Childrens Hospital at Westmead, North Sydney, NSW, Australia

2. Faculty of Veterinary Science, , University of Sydney, Sydney, NSW, Australia

Graham Charles Webb and his many colleagues studied the eukaryote chromosome in diverse contexts. Within this body of work, four themes appear:

- Structural & numerical mutation of karyotype: as an indicator of [and agent within] the evolution of insect populations.
- Chromosome mutation in human [clinical] cytogenetics.
- Mammalian gene mapping.
- Reproductive effects of chromosome mutation, in species of agricultural importance.

Theme 1 grew from Graham's early years in the laboratory of Prof. Michael White, with whom he was closely associated, from 1965. As his PhD student, Graham was involved in Prof. White's "...work on the coastal species of Morabine grasshoppers in Victoria and South Australia, which led to the formulation of a special form of Sympatric Speciation: *Stasipatric Speciation*, for which Michael was justifiably famous" [Webb, pers com]. In a separate study, Graham's own work led him to the belief [contra MJDW!] that the parthenogenetic species *Warramaba virgo* had an origin through species hybridization. This hybrid origin was indeed confirmed, by molecular study, in 1981.

Theme 2: From 1967, Graham established a parallel career, in detection of human chromosome mutations, via diagnostic testing. Through that work, with particular emphasis on the technique of *in situ* hybridisation**, Graham won the reputation as Australia's doyen of clinical cytogeneticists.

Theme 3: In the 1980s, Graham adapted his **ISH expertise to gene mapping, in humans and mice. This work expanded to include other mammals, and Graham came to be recognised throughout the world for his technical excellence. Not least of which was his uncanny ability to learn, in short order, the complex G-band karyotype of any given species, be it a primate, rat, goat, sheep, cow, pig, etc. Not surprisingly, "**Theme 4**" arose as a synthesis of all the above.

GC Webb's contributions to these themes will be discussed, in relation to current knowledge.

232

Chromosome rearrangements in a second transmissible tumour in Tasmanian devils

Maya Kruger-Andrzejewska[1], **Janine Deakin**[1]

1. Institute for Applied Ecology, University of Canberra, Canberra, ACT, Australia

Tasmanian devils are currently under the threat of extinction in the wild due to a transmissible tumour known as devil facial tumour (DFT) disease. Extensive cytogenetic and genetic characterization of tumours from different individuals have supported an allograft theory of transmission, where the tumour cells themselves are the infectious agent. Cases of transmissible tumours are rare but it appears that a second transmissible facial tumour (DFT2) has been identified in individuals from southern Tasmania (Pye et al. 2016). DFT2 is karyotypically and genetically distinct from DFT1 and also appears to be of a different cellular origin. The emergence of a second transmissible raises the question as to whether there are common genomic features shared by these two tumours. As a first step towards addressing this question, we mapped 57 bacterial artificial chromosomes, spread across all six autosomes and the X chromosome, by fluorescent in situ hybridisation. This enabled us to identify chromosome rearrangements in DFT2 and to make comparisons to DFT1. Like DFT1, chromosome 1 is the most rearranged chromosome in DFT2, having acquired fragments from all five other autosomes. Cytogenetically, this is probably the only feature shared between the two facial tumours, although the chromosome rearrangements are quite different between them. Chromosome 1 has also been highly rearranged in the course of marsupial evolution, making it tempting to suggest that this region of the genome in marsupials is more susceptible to breakage and rearrangement than other chromosomes.

1. Pye RJ, Pemberton D, Tovar C, et al. (2016) A second transmissible cancer in Tasmanian devils. Proc Natl Acad Sci 113:374–379

233

The evolutionary complexity of ecological networks

Daniel B Stouffer[1]

1. University of Canterbury, Christchurch, CANTERBURY, New Zealand

Present-day ecosystems face a variety of threats, such as invasive species, whose effects permeate entire communities. Partly for this reason, community ecology has been quick to embrace more holistic approaches that consider all species within an ecosystem and the collection of interactions between them within a network formalism. Perhaps because it draws on tools from graph theory and statistical physics, a cornerstone idea in this network approach is the notion that ecological networks are paradigmatic complex systems. In this talk, I will argue that the majority of research during the past few decades actually demonstrates that the exact opposite is true. To demonstrate this point, I will begin by outlining the growing body of network studies that describe ubiquitous structural patterns that link different ecological networks together, independent of details like where they come from or their particular species composition. I will then describe how adopting a more evolutionary perspective---built by the incorporation of species' phylogenetic relatedness or macro-evolutionary models into network studies---has provided evidence to support the mechanisms that underpin these structual patterns. Finally, I will conclude by explaining how the apparent "limits" to ecological complexity are not a drawback but instead direct us to an exciting new set of open research questions.

234

The biophysical origin of observed patterns of protein evolutionary divergence of sequence, structure, and motion

Julian Echave[2, 1]

1. Consejo Nacional de Investigaciones Científicas y Tecnológicas (CONICET), Buenos Aires, Argentina

2. Escuela de Ciencia y Tecnología, Universidad Nacional de San Martín (UNSAM), San Martín, Buenos Aires, Argentina

Protein evolutionary divergence is not random. A simple comparison of homologous proteins shows clear patterns of differential conservation/variation at the levels of amino-acid sequence, 3D structure, and protein motion. For instance, the rate of sequence evolution varies among sites; protein structures diverge more at some sites than others, deforming along the same coordinates that govern the softest internal protein vibrations, and some protein vibrations are more variable than others.

Here, I will describe a simple biophysical model of protein evolution that integrates the divergence of sequence, structure, and motion; it models the mutational effects on structure, dynamics, and stability, and natural selection as a function of protein stability. Despite its simplicity, the model matches the observed patterns of evolutionary divergence very well.

I will show that, as we usually rightly assume, sequence patterns emerge mainly from selective constraints. In contrast, stability-based natural selection has almost no influence on the divergence of structures and motions. Thus, there is a basic difference: **patterns of sequence variation among sites are due to selection, but patterns of differential conservation of structure and motion are mainly mutational.**

235

Stochastically varying environments promote evolution of modularity and hierarchy in simulated bacterial metabolic networks

Aaron Goodman[1], **Marcus Feldman**[1]

1. Stanford University, Stanford, CA, United States

A typical E. coli bacterium undergoes a complete cell cycle every 40 minutes, in doing so it hydrolyzes 10 to 50 * 10^9 ATP molecules. The process is carried out by around 500 metabolic enzymes and around 1200 distinct metabolites. A bacterium's ability to reproduce depends on the efficiency of its metabolism. The complex metabolisms of bacteria are often studied as a network of metabolites linked together by the enzymes that transform one metabolite into another. The properties of these networks vary across the bacterial kingdom and are influenced by the bacterial life histories. It has been found that bacteria evolve modular networks to survive in changing environments, and evolve hierarchical networks to optimally process metabolites when the environment is stable.

However, despite apparent opposing selective pressures for hierarchy in stable environments, and modularity in varying environments, degree of hierarchy and modularity are correlated in real world metabolic networks. We use an artificial chemistry approach to simulate the evolution of metabolic networks to show how evolution in varying environments can affect the topological properties of bacterial metabolic networks.

Using a simplified model of bacterial metabolisms in which the number of enzymes and metabolites is significantly restricted, we are able to simulate evolution of organisms and reconstruct metabolic networks. We find that the way in which environments vary, whether predictably, or randomly greatly impacts the topology of the optimal metabolic networks. In random unpredictable environments, hierarchy is advantageous and modularity disadvantageous. However artificial metabolic networks that evolve in stochastically varying environments have the same topological properties as real world bacterial metabolic networks.

1. Peter D. Karp, Monica Riley, Suzanne M. Paley, Alida Pellegrini-Toole, and Markus Krummenacker. EcoCyc: Enyclopedia of Escherichia coli genes and metabolism. Nucleic Acids Research, 25(1):43–50, 1997.
2. S. J. Pirt. The Maintenance Energy of Bacteria in Growing Cultures. Proceedings of the Royal Society B: Biological Sciences, 163(991):224–231, oct 1965.
3. J. Pramanik and J. D. Keasling. Stoichiometric model of Escherichia coli metabolism: Incorporation of growth-rate dependent biomass composition and mechanistic energy
4. F.C. Neidhardt, J.L. Ingraham, and M. Schaechter. Physiology of the bacterial cell : a molecular approach. Sinaurer Associates, Sunderland, MA, 1990.
5. D. E. Dykhuizen, a. M. Dean, and D. L. Hartl. Metabolic flux and fitness. Genetics, 115(1):25–31, 1987.
6. Hongwu Ma and AP Zeng. Reconstruction of metabolic networks from genome data and analysis of their global structure for various organisms. Bioinformatics, 19(2):270–277, 2003.
7. J Luo and CL Magee. Detecting evolving patterns of selforganizing networks by flow hierarchy measurement. Complexity, 2011.
8. Merav Parter, Nadav Kashtan, and Uri Alon. Environmental variability and modularity of bacterial metabolic networks. BMC evolutionary biology, 7:169, 2007.
9. Anat Kreimer, Elhanan Borenstein, Uri Gophna, and Eytan Ruppin. The evolution of modularity in bacterial metabolic networks. Proceedings of the National Academy of Sciences of the United States of America, 105(19):6976–81, may 2008.
10. Kashtan, N., Noor, E., & Alon, U. (2007). Varying environments can speed up evolution. Proceedings of the National Academy of Sciences of the United States of America, 104(34), 13711–6.

236

Physical origins and evolutionary effects of high-order epistasis in genotype-phenotype maps

Michael J Harms[1], **Zachary R Sailer**[1]

1. University of Oregon, Eugene, OREGON, United States

A key goal in molecular evolution is to understand why one evolutionary trajectory is taken rather than others. One important determinant of these outcomes is epistasis, where the effect of mutation depends on the presence or absence of other mutations. While pairwise epistasis has been studied extensively, much less is known about high-order epistasis between three or more mutations. If present, high-order interactions could lead to a profound memory effect in evolution, with early mutations strongly shaping evolutionary outcomes. To investigate high-order epistasis, we analyzed a series of published, experimental genotype-phenotype maps using a robust statistical model. We found extensive high-order epistasis, with statistically-significant interactions between up to six mutations. Removing these interactions dramatically altered outcomes of evolutionary simulations in these maps, revealing that high-order epistasis can indeed shape evolutionary outcomes. We next investigated the origins of this

epistasis, finding that the epistasis in some datasets could be explained by a globally nonlinear genotype-phenotype map. In others, it results from specific interactions between mutations. Using a biophysical model of proteins, we then showed that the observed patterns of high-order epistasis arise naturally from the ensemble nature of biomolecular systems. From these results, we propose that high-order epistasis is the rule rather than the exception in molecular evolution, and that this is a natural consequence of the physical properties of biomolecular systems. This implies that, in general, the effect of mutations will be different if they occur early or late in evolution, and that early evolutionary events strongly constrain future ones.

237

The use of mechanistic genotype-phenotype mapping models to simulate the evolution of transcriptional systems

Jayson Gutiérrez[2, 1], Bo Colruyt[2, 1], Steven Maere[2, 1]

1. Department of Plant Systems Biology, VIB, Ghent, Belgium

2. Department of Plant Biotechnology and Bioinformatics, Ghent University, Ghent, Belgium

Parallel to the development of increasingly mechanistic system-scale modeling approaches in molecular biology, there is a growing interest in attaining a more mechanistic perspective on the evolution of molecular systems. In recent years, several evolutionary simulation approaches and artificial life platforms have been developed to study the evolution of systems. However, modelling the genotype-phenotype relationships of biological systems to a degree of realism that is sufficient for studying the detailed molecular mechanisms of system evolution remains a challenge. We developed a novel genotype-phenotype mapping (GPM) framework to model the expression phenotypes of transcriptional systems from artificial genome sequences, inspired on statistical thermodynamics approaches used to model transcriptional regulation processes. We used this GPM model to study the capacity of several classes of gene regulatory networks (GRNs) to evolve novel phenotypes. We found that the evolvability of identically wired networks with qualitatively the same phenotype strongly depends on the exact genomic sequence of the system under study. Moreover, the evolvability of GRNs towards pre-specified target expression phenotypes is often surprisingly limited, and in most cases crucially hinges on the occurrence of neutral substitutions, with very few direct adaptive paths leading to higher fitness. We also tested whether genome duplication enhances the evolvability of GRNs, and found that, although genome duplication often does increase system evolvability, this is generally far less evident than previously assumed. In brief, the evolution of systems appears to be vastly more complex than anticipated on the basis of the highly abstracted evolutionary simulation models in current use, and we believe that fine-grained modelling approaches such as the one used here will become indispensable to shed more light on the mechanisms and constraints underlying system evolution.

238

Functional retention of protein-protein interactions despite substantial sequence divergence.

Ashley I. Teufel[1], Austin G. Meyers[2], Claus O. Wilke[1]

1. Institute for Cellular and Molecular Biology, AUSTIN, TX, United States

2. School of Medicine, Texas Tech University Health Sciences , Lubbock, TX, United States

Orthologous genes in divergent species are expected to perform similar or identical functions. Recently, experimental work demonstrated that nearly half of yeast genes can be functionally replaced with their human orthologs. To understand the evolutionary constraints that lead to the replicability of orthologous proteins in highly divergent genomes we performed a set biochemically realistic simulations on interacting proteins. Using a novel method of efficient forward simulation we evolved several sets of interacting proteins under different selective, stability, and population genetic scenarios. We then evaluated their ability to bind evolutionarily divergent ancestral partners. We find that selection for protein-protein interactions preserves the ability to bind to divergent partners despite extensive sequence divergence, though eventually the accumulation of mutations causes binding incompatibilities. These findings have implications for humanizing entire cellular processes in yeast, which could simplify drug discovery and studies of human genetic polymorphisms. These findings also shed light on how the co-evolution of residues at binding interfaces maintain function over long evolutionary periods.

239

Out in the (not so) cold: ancient DNA and the tropics

Kirsten Ziesemer[1], Marcela Sandoval-Velasco[2], Jazmin Ramos-Madrigal[2], Christina Warinner[3], Corinne Hofman[1], Eske Willerslev[2, 4], Hannes Schroeder[1, 2]

1. Faculty of Archaeology, Leiden University, Leiden, the Netherlands

2. Natural History Museum of Denmark, University of Copenhagen, Copenhagen, Denmark

3. Department of Anthropology, University of Oklahoma, Oklahoma, USA

4. Department of Zoology, University of Cambridge, Cambridge, United Kingdom

Ancient DNA studies have had a major impact in archaeology and ecology. However, the vast majority of ancient DNA studies have been performed on samples from cold or temperate regions, where DNA preservation is generally much better than in non-temperate settings. With average annual temperatures over 25°C and high levels of ambient humidity, the Caribbean represents a particularly challenging environment for ancient DNA research, which explains why so few ancient DNA studies have been conducted in the region to date.

Nonetheless, previous studies have shown that ancient DNA does preserve in the Caribbean, in some cases possibly up to several thousand years; however, the factors influencing DNA preservation in this challenging environment are as yet not well understood. High-throughput sequencing data offer an excellent way to study the molecular preservation of archaeological specimens, using such indicators as the endogenous DNA content, average fragment lengths, molecular decay rates, and DNA fragmentation and deamination patterns.

Using low-coverage, high-throughput sequencing data from over 100 archaeological bone and tooth specimens from the Caribbean, we systematically explore the effects of temperature, age, microbial action, time since excavation, sample and soil type, and burial setting on DNA preservation. Results indicate that the level of DNA preservation varies dramatically across the region and that factors such as the burial setting, the age of the sample, and sample type can have a strong influence on the level of preservation.

240

20,000 bones and counting - Insights into Past Biodiversity and Ancient DNA Preservation using Bulk-Bone Metabarcoding

Michael Bunce[1], Dáithí C Murray[1], Alicia Grealy[1], Dalal Haouchar[1], Daniel Werndly[1], Kealoha Kinney[2], Morten Allentoft[3]

1. Trace and Environmental DNA (TrEnD) laboratory, Curtin University, Perth, WA, Australia

2. Brown University, Providence, RI, USA

3. GeoGenetics, University of Copenhagen, Copenhagen, Denmark

Fossil bones provide a unique window into the past but they are often difficult to interpret. Only a small proportion of animals are preserved as fossils – an even smaller fraction are then recovered and able to be identified morphologically. We have developed a globally applicable next generation DNA sequencing method that offers a genetic perspective on fossil assemblages with the aim of rapidly overlaying genotype data over more traditional methods of study. Our approach is called bulk-bone metabarcoding (BBM). BBM involves the conversion of largely non-diagnostic bone fragments into powder which is then genetically indexed, amplified and sequenced on NGS platforms.

This presentation will showcase some BBM data from a variety of sites across Australia, New Zealand, Hawaii, USA, Madagascar and Armenia. The data generated using BBM provides some key insights into past biodiversity and faunal turnover. Moreover the approach is an efficient way to assess DNA preservation both within and between fossil sites. Taken together, bulk-bone metabarcoding provides a powerful and cost-effective way to study past biodiversity with tangible benefits in conservation science, paleobiology and archaeology.

241

Ancient DNA preservation in tropical pre-contact archaeological sites in the Americas

Maria A Nieves Colon[1], Anne C Stone[1]

1. School of Human Evolution and Social Change, Arizona State University, Tempe, Arizona, United States

Paleogenomics research on populations from the tropical Americas has historically been constrained by poor recovery of ancient DNA. While the effects of environmental and thermal conditions on DNA survival have been studied and modeled in ancient samples from temperate contexts, similar methodological research in ancient tropical samples is still lacking. Here we compare DNA preservation in human skeletal remains from two pre-contact tropical contexts in the Americas: the Yaxuná site in the Yucatán peninsula (250-550 CE) (n=6), and the Punta Candelero, Paso del Indio and Tibes sites in Puerto Rico (500-1300 CE) (n=29). We compare the application of two different extraction methods on teeth and, in the Yaxuná samples, petrous portion tissue. All samples were extracted, transformed into libraries, captured for the complete mitochondrial genome and sequenced on the Illumina MiSeq. Preliminary results suggest that endogenous DNA content is lower in the Yaxuna samples than in the Puerto Rican samples irrespective of extraction method. Yields from petrous portions were also unexpectedly low in these samples. There was no statistically significant difference in endogenous DNA recovery from either extraction method across samples from both sites ($t_{(8)}$=-0.0095, p=0.99). These early findings suggest that DNA survival may be highly dependent on site-specific processes and that efforts to recover ancient genetic material in tropical samples may need to be tailored on a case-by-case basis. We are currently testing these inferences further by exploring the relationship between deposition age and temperature on overall DNA preservation in these samples through a modified model of DNA decay.

242

Using whole-genome in-solution capture to infer the geographic origin of Indian Ocean enslaved people in the historical cemetery of Le Morne (Mauritius)

Rosa Fregel[1], Marcela Sandoval-Velasco[2], Martin Sikora[2], Maria C Avila-Arcos[1, 3], David Poznik[1], Fernando Mendez[1], Meredith Carpenter[1], Alexandra Sockell[1], Christopher R Gignoux[1], Krish Seetah[1], Sasa Caval[1], Diego Calaon[1], Carlos Bustamante[1]

1. Stanford University, Stanford, CALIFORNIA, United States

2. Centre for GeoGenetics, University of Copenhagen, Copenhagen, Denmark

3. International Laboratory for Human Genome Research, National Autonomous University of Mexico, Queretaro, Mexico

Until recently, paleogenomic techniques were restricted to cold environments, as DNA degradation in warmer regions makes sequencing costs unaffordable. However, key chapters of human history happened in warm areas, and technical advances in capturing endogenous DNA have been developed for extending the geographic limits of ancient DNA research.

One key period of history involved the forced migration of millions of people due to slavery. Recent paleogenomic studies have helped to shed light on the origins of slaves within the trans-Atlantic trade. However, the Indian Ocean slave trade has remained understudied. In this project, we seek to elucidate the history of slavery across the Indian Ocean by focusing on Mauritius, where hundreds of thousands of slaves were imported during colonial times. We used next-generation sequencing of ancient DNA to estimate the genome-wide ancestries of individuals sampled in the historical cemetery of Le Morne (n=26), which is thought to contain the remains of Malagasy slaves. As tropical climate conditions in Mauritius are an obstacle for ancient DNA recovery, we used whole-genome in-solution capture (WISC) to enrich endogenous DNA.

Endogenous DNA accounted for 0.14–45.1% of the total. On average, through WISC we were able to increase endogenous DNA content by a factor of 13.0 (4.5–34.5), demonstrating the importance of capture-enrichment methods for paleogenomics in tropical regions. Although genome coverage was low (median=2.4%, IQR=1.1%–7.8%), for 80% of the samples, at least 2,000 SNPs intersected with a reference panel of 555 modern samples. Principal component analysis and admixture estimates indicate that samples from Le Morne had ancestries related to not only mainland Africa and Madagascar, but we also detected European and South Asian admixture. Our results represent the first genomic study to recover data from individuals involved in the Indian Ocean slave trade, and they provide insight into the complex demographic history of Mauritius.

243

The genomic enigma of two Medieval North Africans

Torsten Günther[1], Cristina Valdiosera[2], Juan Carlos Vera-Rodríguez[3], Ricardo Rodriguez-Varela[4], Emma Svensson[1], Rafael M Martínez Sánchez[5], Rafael Carmona Ávila[6], Leonor Peña Chocarro[7], Guillem Pérez Jordà[8], Youssef Bokbot[9], Eneko Iriarte[10], Colin Smith[2], Mattias Jakobsson[1]

1. Uppsala University, Uppsala, Sweden

2. La Trobe University, Melbourne, Australia

3. Departamento de Historia, Geografía y Antropología, Universidad de Huelva, Huelva, Spain

4. Universidad Complutense de Madrid, Madrid, Spain
5. Departamento de Prehistoria y Arqueología, Universidad de Granada, Granada, Spain
6. Museo Histórico Municipal, Córdoba, Spain
7. Escuela Española de Historia y Arqueología en Roma/CSIC, Rome, Italy
8. G. I. Arqueobiología, Instituto de Historia, Madrid, Spain
9. Institut National des Sciences de l'Archéologie et du Patrimoine, Rabat, Morocco
10. Departamento Historia, Geografía y Comunicación, Universidad de Burgos, Burgos, Spain

The trans-Saharan gold and salt trade as well as the trans-Saharan slave trade played an important role in population movements connecting sub-Saharan and Mediterranean economies during the Middle Ages. The slave trade alone is said to have transported more than 9 million slave soldiers and domestic servants along the trans-Saharan route. In this study, we present the genomic analysis of two human individuals from a cave site in the area of present-day Morocco which were directly dated to the Medieval period. The samples were processed in a designated ancient DNA lab and the genomic data obtained shows standard patterns of authentic ancient DNA with low levels of contamination. Both individuals – which represent the first ancient genome sequence data from North Africa – do not exhibit particular genetic affinities to modern North Africans or any other present-day population in published genotype data sets despite relatively extensive data has been produced from many areas of Africa. In fact, the most parsimonious way to model them genetically is as two-source admixture between Mediterranean Europeans and Southern Africans. The lack of archaeological context of the two individuals opens up various alternatives to explain their genomic pattern. Both individuals could represent a Medieval African population without population continuity to modern-day populations. Alternatively, both Mediterranean Europe and Southern Africa are known source regions in the Arab slave trade, thus they could potentially represent the offspring of slaves of different origin. The Arab slave trade extended over a longer period and may have involved more slaves than its transatlantic counterpart and our data might provide the first genetic insight into this historical process and the people who suffered in it. Our results highlight how archaeogenetic research can shed lights into historical events and long-distance population movements while opening new questions for the interpretation of the data.

244

Eggshell palaeogenomics: palaeognath evolutionary history revealed through ancient nuclear and mitochondrial DNA from the Madagascan elephant bird *Aepyornis*

Alicia Grealy[1], Matthew Phillips[2], Gifford Miller[3], Marcus Thomas Gilbert[4], David Martin Lambert[5], Michael Bunce[1], James Haile[6]

1. Curtin University, Perth, WA, Australia
2. Queensland University of Technology, Brisbane, QLD, Australia
3. University of Colorado, Boulder, Colorado, USA
4. Centre for GeoGenetics, Natural History Museum, University of Copenhagen, Copenhagen, Denmark
5. Griffith University, Brisbane, QLD, Australia
6. University of Oxford, Oxford, UK

Palaeognaths, the sister group of all other living birds, were long considered to be relics from the breakup of the Gondwanan supercontinent. However, there is renewed skepticism of the extent to which vicariance explains palaeognath biogeography, with recent molecular studies instead arguing for dispersal of volant ancestors across marine barriers. Resolving this debate hinges upon accurately reconstructing the evolutionary relationships and timing of divergence among this group, which remain contentious. Recently, mitogenome sequences from the extinct elephant birds of Madagascar have further informed the palaeognath phylogeny; however, nuclear loci have been unavailable due to the rarity of bone specimens with well-preserved ancient DNA (aDNA). Nevertheless, nuclear information often proves crucial for accurately recovering deep evolutionary relationships. Here, we use DNA extracted from fossil eggshell in conjunction with target enrichment and next-generation sequencing techniques to independently reconstruct the mitochondrial genome and recover nuclear loci from *Aepyornis* sp. We confirm that elephant birds are sister taxa to the kiwi (*Apteryx* spp.); however, our data suggests that, like neognaths, the notopalaeognathae (all palaeognaths excluding ostrich) underwent an explosive radiation between 64.2-54.2 mya—well after the break-up of Gondwana, and more rapidly than previously estimated from mitochondrial data alone. These results further support the idea that ratites convergently evolved flightlessness immediately following the K-Pg mass extinction event, favoring the dispersal hypothesis over a vicariant model. Our study reinforces the importance of including information from the nuclear genome of extinct taxa for understanding the evolutionary history of their modern relatives. With approximately 3% endogenous aDNA retrieved, avian eggshell can be a valuable substrate for recovering high quality aDNA, particularly from environments that are not typically conducive to aDNA preservation. We suggest that elephant bird whole genome recovery is ultimately achievable, and will provide future insights into the evolution, adaptation, and development of these enigmatic birds.

245

Ancient DNA preservation: cutting to the bone

Morten E Allentoft[1], Peter B Damgaard[1], Henrik B Hansen[1], Ashot Margaryan[1], Jørgen Hollesen[2], Michael Bunce[3], Eske Willerslev[1]

1. Centre for GeoGenetics, Natural History Museum, University of Copenhagen, Copenhagen, Denmark
2. The National Museum of Denmark, Copenhagen, Denmark
3. Trace and Environmental DNA (TrEnD) laboratory , Curtin University , Perth, WA, Australia

Poor preservation is a highly constraining factor in ancient DNA research. In the majority of ancient specimens, endogenous DNA molecules represent a minor fraction of the total amount of DNA, rendering shotgun sequencing inefficient for obtaining genomic data. It is therefore necessary to be extremely selective during the sampling process, ideally relying on published data and comparative analyses. Such data are, however, in short supply.

NGS data offer an excellent means to obtain detailed insigths into the molecular preservation of a given specimen. As a convenient "by-product" of shotgun sequencing, it is possible to estimate the endogenous DNA content, the average fragment lenght, the DNA decay rate and half-life, and the deamination damage fraction.

Using NGS data from hundreds of ancient skeletons we compare these signatures of molecular preservation in different skeletal elements, and in skeletons that differ in respect to age, preservation environment, and burial temperature. We support the findings with similar data from *in situ* burial experiments. While some degree of uncertainty persists, our results show that factors such as preservation state, burial temperature, and age can indeed provide grounds for predicting molecular preservation in ancient biological remains and the potential for genome-scale research.

246

Harvesting information from ultra-short ancient DNA sequences

Cesare de Filippo[1], **Matthias Meyer**[1], **Kay Prüfer**[1]

1. *Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany*

One of the characteristic features of ancient DNA is its degradation into short fragments. Advances in ancient DNA extraction and library preparation methods now make it possible to retrieve extremely short fragments, in principle improving access to highly degraded, ancient material. However, ultra-short sequences remain a challenge for analysis since it is not trivial to distinguish endogenous sequences from microbial contaminants, which typically constitute the vast majority of sequences recovered from ancient fossils.

To explore the utility of ultra-short sequences, we developed a method to estimate the proportion of spurious alignments to the human reference (hg19) genome and applied it to Neandertal samples of various ages with different proportions of endogenous DNA. The method is based on modifying the hg19 genome at random sites in non-repetitive, mappable regions. Sequence alignments overlapping mutated sites can then be classified as spurious or authentic alignments based on their sharing of the mutant or non-mutant state.

The proportion of spurious alignments decreases with increasing read length, depends on the relative abundance of microbial contaminant sequence, and is reduced by using only sequences with terminal C-to-T substitutions (i.e. showing evidence of deamination-induced base damage). Using this and other filters we define lower size cut-offs between 25 to 31 base-pairs (bp), depending on the specimen, while limiting the fraction of spurious alignments to less than 10%. When using only these short sequences in phylogenetic analyses, we observe no significant difference compared to using sequences of at least 35 bp, which is the size cut-off used in previous studies of archaic human DNA. By including shorter sequences, we considerably increase the amount of sequence information that can be recovered from highly degraded DNA. Our method may help to make samples available for genetic analyses that previously yielded too few or no informative DNA sequences.

247

Functional analysis of nine retrotranposons inserted in the promoter of a stress-response gene in Drosophila

Josefa González[1]

1. *Institute of Evolutionary Biology (CSIC-UPF), Barcelona, SPAIN, Spain*

We have performed a genome-wide screen for recent adaptive transposable element insertions in Drosophila melanogaster. While previous studies included only a subset of the transposable element insertions present in the reference genome, we have now analyzed all the euchromatic insertions in three out-of-Africa populations, North Carolina (US), Bari (Italy) and Stockholm (Sweden), and in one African population from the ancestral range of the species. We identified 41 candidate adaptive TEs with significantly different frequencies within and outside of Africa. One of the putatively adaptive insertions identified is a *roo* solo-LTR retrotransposon located in the 5'-UTR of *CG18446*, a candidate cold-stress response gene. We found that besides *FBti0019985,* there are another eight transposable elements inserted in the proximal promoter region of *CG18446*. All nine insertions are solo-LTRs that belong to the *roo* family. We found that different insertions have different molecular and functional consequences. The exact position where the transposable elements are inserted matters, as they all showed highly conserved sequences but only two of the analyzed insertions provided alternative transcription start sites, and only the *FBti0019985* insertion consistently affects *CG18446* expression. The phenotypic consequences of the different insertions also vary: only *FBti0019985* was associated with cold-stress tolerance. Interestingly, in the only previous report of transposable elements inserting repeatedly and independently in a promoter region in *D. melanogaster,* the insertions were also located upstream of a stress response gene. Our results suggest that functional validation of individual structural variants is needed to resolve the complexity of insertion clusters.

248

Endogenous L1 Retrotransposition in the Mammalian Primordial Germline and Early Embryo

Sandra R Richardson[1], **Patricia Gerdes**[1], **Daniel J Gerhardt**[1], **Francisco J Sanchez Luque**[1, 2], **Samuel Jesuadian**[1], **Marie-Jeanne H.C. Kempen**[1], **Gabriela-Oana Bodea**[1], **Patricia E Carreira**[1], **Adam D Ewing**[1], **Geoff J Faulkner**[1, 3]

1. *Mater Research Institute-University of Queensland, Woolloongabba, QLD, Australia*
2. *Pfizer-Universidad de Granada-Junta de Andalucia Centre for Genomics and Oncological Research (Genyo), Granada, Spain*
3. *School of Biomedical Sciences, University of Queensland, Brisbane, QLD, Australia*

Long Interspersed Element 1 (LINE-1 or L1) is a retrotransposon presently active in mammalian genomes. L1 insertions within and proximal to genes can impact gene expression in a variety of ways, and retrotransposition events are frequently associated with duplication, deletion, and rearrangement of genomic sequences. In order to exert an ongoing impact on genome evolution, new L1 insertions must occur in cells that will contribute their genetic material to subsequent generations—ie, within the germ lineage or in the pluripotent cells of the early embryo, prior to germline specification. Previous studies have suggested that the early embryo is a prominent milieu for L1 retrotransposition; however, systematic study of the frequency and developmental timing of heritable L1 retrotransposition events has been technically challenging. Here, we have adapted retrotransposon capture sequencing (RC-seq) to detect retrotransposon insertions in mouse genomes, and applied this technique to identify de novo heritable L1 insertions in multi-generation pedigrees of C57BL/6 mice. We identify 11 full-length Tf subfamily L1 insertions among 85 mouse genomes, providing a conservative estimate of 1 new inseriton per 8 mice. Using a PCR genotyping strategy to deduce developmental timing of these events, we find evidence consistent with L1 retrotransposition in the early embryo resulting in somatic and germline genetic mosaicism, as well as in early primordial germ cells (PGCs), giving rise to germline-restricted genetic mosaicism. We also identify L1 insertions attributable to later germline development. Furthermore, by exploiting 3' transductions carried by two de novo insertions, we identify progenitor L1 elements active in the early primordial germline and in the pluripotent cells of the early embryo. Our findings shed new light on the frequency and developmental origins of the ongoing retrotransposition events continuously shaping mammalian genomes.

249

The role of transposable elements for gene expression in Capsella hybrids and allopolyploids

Kim A Steige[1, 2], **Johan Reimegård**[3], **Carolin A Rebernig**[4], **Claudia Köhler**[4], **Douglas G Scofield**[5], **Tanja Slotte**[2]

1. *University of Cologne, Cologne, Germany*
2. *Stockholm University, Stockholm, Sweden*
3. *Science for Life Laboratory, Uppsala University, Uppsala, Sweden*

4. Swedish University of Agricultural Sciences, Uppsala, Sweden

5. EBC, Uppsala University, Uppsala, Sweden

The formation of allopolyploid species involves the merger of two genomes with separate evolutionary histories. In allopolyploids, genes derived from one progenitor species are often expressed at higher levels than those from the other progenitor. It has been suggested that this could be due to differences in transposable element (TE) content among progenitors, as silencing of TEs can affect expression of nearby genes. Here, we examine the role of TEs for expression biases in the widespread allotetraploid *Capsella bursa-pastoris* and in diploid F1 hybrids generated by crossing *Capsella orientalis* and *Capsella rubella*, two close relatives of the progenitors of *C. bursa-pastoris*. As *C. rubella* harbors more TEs than *C. orientalis*, we expect *C. orientalis* alleles to be expressed at higher levels if TE content is key for expression biases. To test this hypothesis, we quantified expression biases at approximately 5800 genes in flower buds and leaves, while correcting for read mapping biases using genomic data. While three of four *C. bursa-pastoris* accessions exhibited a shift toward higher relative expression of *C. orientalis* alleles, the fourth *C. bursa-pastoris* accession had the opposite direction of expression bias, as did diploid F1 hybrids. Associations between TE polymorphism and expression bias were weak, and the effect of TEs on expression bias was small. These results suggest that differences in TE content alone cannot fully explain expression biases in these species. Future studies should investigate the role of differences in TE silencing efficacy, as well as a broader set of other factors. Our results are important for a more general understanding of the role of TEs for *cis*-regulatory evolution in plants.

250

Retroelements and noncoding RNAs linked to centromere turnover during Macropodid chromosome evolution

Rachel O'Neill[1]

1. University of Connecticut, Storrs, CONNECTICUT, United States

Two paradoxes of centromere biology have confounded our understanding of the eukaryotic centromere, and consequently of chromosome evolution. The frist is that centromere function is highly conserved across eukaryotes, yet centromere-specific proteins that interact with nucleic acids diverge rapidly. The second is that while satellite DNA is found ubiquitously across eukaryotic centromeres, it is considered neither necessary nor sufficient for centromere formation. Major hurdles in understanding centromere evolution lie in the highly repetitive nature of most centromeric DNA and an inability to decouple centromere divergence from species evolution and stochastic processes such as genetic drift and molecular drive. Using comparative cytogenomics, we have identified specific small noncoding RNA sequences that are coincident with active centromere demarcation in a broad range of mammalian species spanning all Therian clades, affording the ability to assess the evolution of centromeres in the context of the transcriptional activity of nascent centromeric elements. As an exemplar, the macropodid species complex is typified by rapid chromosome evolution and convergence of karyotypes independent of ancestry. Notably, each of these karyotypic rearrangements involves centromeres and hybrids between even closely related species possess abnormalities delimited to centromeres and follow the observations of Haldane's Rule, indicative of strong, postzygotic reproductive isolation barriers. Our ChIP-seq, RIP-seq, RNA-seq, genome assembly and repeat analysis computational methods have established a testable model of centromere evolution in the context of rapid chromosome and species evolution. Our findings on satellites, retroelements and noncoding RNA elements will be presented in the context of conflict between nucleic acid binding proteins and centromeric DNA domains, and utlimately chromosome evolution.

251

Integration and fixation preferences of human and mouse endogenous retroviruses uncovered with Functional Data Analysis

Kateryna Makova[1], **Rebeca Campos-Sanchez**[1, 2], **Marzia Cremona**[1, 3], **Alessia Pini**[3], **Francesca Chiaromonte**[1]

1. Penn State University, University Park, PENNSYLVANIA, United States

2. Universidad de Costa Rica, San Jose, Costa Rica

3. Politecnico di Milano, Milano, Italy

Endogenous retroviruses (ERVs), the remnants of retroviral infections in the germ line, occupy ~8% and ~10% of the human and mouse genomes, respectively, and affect their structure, evolution, and function. Yet we still have a limited understanding of how the genomic landscape influences integration and fixation of ERVs. Here we conducted a genome-wide study of the most recently active ERVs in the human and mouse genome. We investigated 826 fixed and 1,065 *in vitro* HERV-Ks in human, and 1,624 fixed and 242 polymorphic ETns, as well as 3,964 fixed and 1,986 polymorphic IAPs, in mouse. We quantitated >40 human and mouse genomic features (e.g., non-B DNA structure, recombination rates, and histone modifications) in ±32 kb of these ERVs' integration sites and in control regions, and analyzed them using Functional Data Analysis (FDA) methodology. In one of the first applications of FDA in genomics, we identified genomic scales and locations at which these features display their influence, and how they work in concert, to provide signals essential for integration and fixation of ERVs. The investigation of ERVs of different evolutionary ages (young *in vitro* and polymorphic ERVs, older fixed ERVs) allowed us to disentangle integration vs. fixation preferences. As a result of these analyses, we built a comprehensive model explaining the uneven distribution of ERVs along the genome. We found that ERVs integrate in late-replicating AT-rich regions with abundant microsatellites, mirror repeats, and repressive histone marks. Regions favoring fixation are depleted of genes and evolutionarily conserved elements, and have low recombination rates, reflecting the effects of purifying selection and ectopic recombination removing ERVs from the genome. In addition to providing these biological insights, our study demonstrates the power of exploiting multiple scales and localization with FDA. These powerful techniques are expected to be applicable to many other genomic investigations.

252

The dynamic landscape of transposition across the speciation continuum of *Ficedula* flycatchers

Alexander Suh[1], **Linnéa Smeds**[1], **Hans Ellegren**[1]

1. Department of Evolutionary Biology (EBC), Uppsala University, Uppsala, Sweden

Birds have the smallest genomes among land vertebrates. Their genome sizes are roughly a third of the human genome due to massive genome compaction in their dinosaur ancestors, potentially resulting from the metabolic requirements of powered flight. Consequently, transposable elements are relatively scarce in avian genomes and it has been suggested that rarity of transposition explains the stability of genome size and chromosomal organization across extant birds. Here we show that avian genomes instead have a dynamic and diverse landscape of transposition-derived structural variation. We analyzed 201 re-sequenced genomes of six species of *Ficedula* flycatchers and a three-generation pedigree of eleven collared flycatchers for transposable element variation (TEV). Following read mapping and stringent filtering, we discovered >10,000

transposon presence/absence polymorphisms. These TEVs include many which are shared between multiple species, however, around two thirds of the TEVs are private alleles. In combination with our pedigree data, this suggests that transposition occurs relatively frequently in flycatchers. Most TEVs belong to eight different families of long terminal repeat (LTR) retrotransposons from the major groups of endogenous retroviruses (ERV); ERV1, ERV2, and ERV3. We further find that chromosomal recombination rate and density of sites under selection are predictors of TEV abundance. Altogether, we suggest that transposition is relatively frequent and diverse in birds, and that the overall scarcity of fixed transposable elements in avian genomes results from rare retention of TEVs due to low population frequencies.

253

CRISPR-Cas and origin of adaptive immunity from selfish genetic elements

Eugene Koonin[1]

1. National Institutes of Health, USA, Bethesda, MARYLAND, United States

The archaeal and bacterial CRISPR-Cas systems of adaptive immunity employ small guide RNAs derived from foreign or self DNA for adaptive immunity against viruses and plasmids and apparently, in some case, also for regulation of gene expression. The RNA-guided Cas nucleases comprise the new generation of genome editing tools and are often claimed to have ushered a revolution in genetic engineering. Comparative genomic analysis of the CRISPR-Cas loci identified multiple contributions of various mobile genetic elements to the evolution of prokaryotic adaptive immunity. The contributing mobile elements include: i) casposons, a newly discovered superfamily of archaeal and bacterial self-synthesizing transposons that gave rise to the adaptation modules of CRISPR-Cas and apparently CRISPR arrays themselves, ii) autonomous and non-autonomous transposons encoding the TnpB nuclease that gave rise to the RuvC-like nuclease domains of effector nucleases in type II and type V CRISPR-Cas systems, iii) toxin components of toxin-antitoxin systems that became ancestors of Cas2 subunit of the adaptation complex and the effector nucleases of type VI, iv) self-splicing introns which donated the HNH nuclease of the type II effectors. A parallel is drawn between the evolution of CRISPR-Cas, vertebrate adaptive immunity and the mechanism for DNA elimination during macronucleus maturation in ciliates. In each of these case, the molecular machinery for genome rearrangement evolved from unrelated transposons which reflects the inherent aptitude of mobile elements for these functions. Enzymes encoded by mobile elements, especially nucleases, appear to be “guns for hire” that utilized alternately for offense, defense and counter-defense by genetic parasites and their hosts.

254

Human migrations and megafaunal extinctions

Eske Willerslev[2, 1]

1. Department of Zoology, University of Cambrigde , Cambrigde , UK

2. The Natural History Museum of Denmark, København K, COPENHAGEN, Denmark

Research involving ancient DNA has experienced a true technological revolution in recent years through advances in the recovery of ancient DNA and, particularly, through applications of high-throughput sequencing. Formerly restricted to the analysis of only limited amounts of genetic information, ancient DNA studies have now progressed to whole-genome sequencing for an increasing number of ancient individuals and extinct species. In this talk I will provide an overview of recent findings done by my group. This concerns what we have learned on early peopling of the Americas, early peopling of Eurasia, and Australia as well the more recent human history of Europe and central Asia. I will also talk about some of our studies related to the population dynamics and extinction of the big bodied mammals (megafauna) around the end of the last Ice Age.

255

Insights into platypus population structure and history from whole-genome sequencing

Hilary C Martin[1, 2], **Elizabeth Batty**[1], **Julie Hussin**[1], **Portia Westall**[3], **Tasman Daish**[4], **Paolo Piazza**[1], **Rory Bowden**[1], **Margaret Hawkins**[5], **Tom Grant**[6], **Craig Moritz**[7], **Frank Grutzner**[4], **Jaime Gongora**[3], **Peter Donnelly**[1, 8]

1. Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, Oxfordshire, United Kingdom

2. Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, United Kingdom

3. Faculty of Veterinary Science, School of Life and Environmental Sciences, University of Sydney, Sydney, NSW, Australia

4. University of Adelaide, Adelaide, SA, Australia

5. Taronga Zoo, Sydney, NSW, Australia

6. University of New South Wales, Sydney, NSW, Australia

7. Research School of Biology and Centre for Biodiversity Analysis, Australian National University, Canberra, ACT, Australia

8. Department of Statistics, University of Oxford, Oxford, Oxfordshire, United Kingdom

The platypus has a remarkable combination of mammalian and reptilian characteristics and, as an egg-laying mammal, alongside the echidna, it occupies a unique place in the phylogenetic tree. Despite widespread interest in its unusual biology, much remains to be learned about its dispersal patterns, population structure, and recent evolutionary history. To address this, we sequenced the genomes of 57 platypuses from across the species’ range (eastern mainland Australia and Tasmania). Our results show very strong population structure, with our sampling locations corresponding to discrete populations between which there is no evidence for recent gene flow. We found that 31 of the 57 samples had at least a third-degree relative amongst other samples from the same river system, indicating that it is not uncommon for related individuals to remain in the same stretch of stream. Despite this, we see many individuals with little evidence of inbreeding, suggesting biological mechanisms to avoid mating with close relatives. Data from a family quartet allowed us to estimate the *de novo* mutation rate in the platypus at 1.2 - 9.6 × 10^{-8} bp/generation, one of the first direct estimates made in a non-model organism. Some patterns of similarity and differences between populations are not easy to reconcile with geography, suggesting historical migration patterns more complicated than predicted by simple isolation-by-distance models. Estimates of historical effective population sizes showed that Queensland populations underwent a strong bottleneck likely during the Last Glacial Maximum, when there is evidence for bottlenecks in other rainforest species in that region. Only the population in the Wet Tropics recovered, and the populations in central Queensland appear to be priorities for conservation due to small effective population size and low diversity. This study demonstrates the power of whole-genome re-sequencing of natural populations of an evolutionarily important species with a problematic reference genome

256

Convergence in the genomics of local adaptation to climate in conifers

Kathryn Hodgins[1], Sam Yeaman[2], Jason Holiday[3], Katie Lotterhos[4], Haktan Suren[3], Laura Gray[5], Simon Nadeau[6], Kristin Nurkowski[1, 7], Loren Rieseberg[7], Mike Whitlock[8], Sally Aitken[6]

1. Monash University, Clayton, VIC, Australia
2. University of Calgary, Calgary
3. Department of Forest Resources and Environmental Conservation, Virginia Polytechnic Institute and State University, Blacksburg
4. Department of Marine and Environmental Science, Northeastern, Nahant
5. Department of Renewable Resources, University of Alberta, Edmonton
6. Department of Forest and Conservation Sciences, University of British Columbia, Vancouver
7. Botany, University of British Columbia, Vancouver
8. Zoology, University of British Columbia, Vancouver

When confronted with a selective challenge, theoretical and empirical results reveal that closely related taxa often evolve similar phenotypes from the same genes. However, evolutionary convergence on the genetic level is thought to be less likely in more distantly related species, because of differences in genetic background and a lack of shared standing variation, although this has not been tested at the genome scale. Here, we provide the first population genomic study of convergent local adaptation to the same climatic gradients between two species diverged for more than 140 million years, lodgepole pine (*Pinus contorta*) and interior spruce (*Picea glauca*, *Picea engelmannii* and their hybrids). Using sequence capture approach that targeted the exome of both species, we used environment allele associations, and phenotype allele associations to identify candidate regions of the genome associated with local adaptation to climate. Our comparative analysis of these regions finds that adaptation to temperature shows polygenic signatures of convergence at the phenotypic and genomic level. This suggests that adaptation to climate is somewhat genetically constrained, with key genes, particularly transcription factors, playing non-redundant roles.

257

Patterns of diversification in two babbler species from Sundaland.

Emilie Cros[1], Frank E Rheindt[1]

1. National University of Singapore, Singapore, SINGAPORE

Pleistocene climate fluctuations are known to be an important driving force in biological differentiation. Despite Sundaland's high levels of biodiversity and the important impact Pleistocene climate fluctuations have had on Sundaic landmass connectivity and topography, studies of the impact of those fluctuations on differentiation in Sundaic species are scarce. Recent studies have shown complex patterns of bioacoustics and genetic variation among bird populations of different landmasses in Sundaland, suggesting that some of those populations may in fact no longer experience gene flow between one another despite recurrent land bridges connecting their ranges. Babblers (Timaliidae) are highly sedentary denizens of the undergrowth of Southeast Asian rainforests in which great regional bioacoustic variation has previously been reported. As such, they are an ideal model to study the mechanisms that have led to biological differentiation across Sundaland. Here, we use genome-wide sequence data to assess patterns of diversification and gene flow among populations from different parts of the Sundaic region. Pleistocene climate fluctuations may have affected population connectivity differently depending on species ecology. We therefore conducted a comparison between a forest-dependent babbler complex (*Cyanoderma erythropterum*) and an edge-tolerant babbler complex (*Mixornis gularis*).

258

Regulation of the mitochondrial transcriptome

Aleksandra Filipovska[1]

1. The University of Western Australia, Nedlands, WESTERN AUSTRALIA, Australia

The size and organization of the animal mitochondrial genome has been reduced and compacted significantly since its endosymbiosis from an α-proteobacterial ancestor. This compaction has necessitated the evolution of unique mechanisms to facilitate rapid changes in gene expression in response to the changing energy demands of the cell. The mitochondrial transcriptome encodes proteins that are subunits of the respiratory chain, responsible for most of the energy production required by the cell. Consequently the coordinated regulation of the mitochondrial transcriptome by the nucleus is of particular importance for the maintenance of cell health and energy metabolism. Over the last few years we have investigated the unusual features of mitochondrial RNAs and the RNA-binding proteins that control their production, maturation, translation and stabilization to understand the regulation of mitochondrial gene expression and its contribution to health and disease. We have characterised several important classes of sequence specific RNA-binding proteins in cells and mouse models of disease. We have established new tools and methods for massively parallel sequencing and analyses of RNase-accessible regions of mitochondrial RNAs to investigate the functions of these proteins in a high throughput manner *in vivo.* To date the regulation of mitochondrial RNA metabolism and its importance for ribosome biogenesis and energy metabolism are not clear. I will discuss the *in vivo* role of mitochondrial RNA regulation and its importance for mitochondrial biogenesis and energy metabolism.

259

Conventional and modern genetic approaches of cloning rust disease resistance in wheat

Sambasivam Periyannan[1], Burkhard Steuernagel[3, 2], Inmaculada Hernández-Pinzón[2], Kamil Witek[2], Matthew Rouse[4], Guotai Yu[3], Asyraf Hatta[3], Mick Ayliffe[1], Harbans Bariana[5], Jonathan Jones[2], Evans Lagudah[1], Brande Wulff[3, 2]

1. Commonwealth Scientific and Industrial Research Organization (CSIRO) Agriculture,, Canberra, ACT, Australia
2. The Sainsbury Laboratory, Norwich, UK
3. John Innes Centre, Norwich, UK
4. USDA-ARS Cereal Disease Laboratory and Department of Plant Pathology, University of Minnesota, Minnesota, USA
5. Plant Breeding Institute, University of Sydney, Cobbitty, NSW, Australia

Wheat is Australia's primary agricultural food crop and a globally important food crop which is constantly under threat due to the emergence of new fungal pathogen strains. For example, the stem rust strain Ug99 which first appeared in the Eastern parts of Africa has the potential to damage more than 60% of the wheat cultivars grown worldwide. As part of a global initiative (Durable Rust Resistance in Wheat) to counter the spread of Ug99, wheat rust research at CSIRO is focused on molecular genetic characterisation of stem rust resistance genes from wild relatives of wheat, which are effective against Ug99 and other virulent wheat stem rust strains.

Using induced mutations and arduous positional cloning, we were successful in isolating one of the first stem rust resistance genes from wheat, *Sr33*, which has also proven to be effective against Ug99. With the knowledge of *Sr33* and advancements in sequencing technologies, a rapid resistance gene cloning technique called Mutagenesis, Resistance gene enrichment and Sequencing (MutRenSeq) was developed in partnership with the John Innes Centre, UK. MutRenSeq has enabled cloning of two additional stem rust resistance genes, namely *Sr22* and *Sr45,* which are also effective against Ug99 and other virulent stem rust strains. This success has paved the way for the rapid isolation of new rust resistance genes, enabling the possibility of multiple transgene cassette deployment for durable management of rust diseases in wheat. In addition to the identification of rust resistance genes in wheat, the technology is currently being utilised for rapid identification of disease resistance genes in other agriculturally important crops such as barley, potato, soybean and rye.

260

'Fixation" of hybrid vigour

Elizabeth Dennis[1], Li Wang[1, 2], Ian Greaves[1], Jim Peacock[1]

1. CSIRO Plant Industry, Canberra, ACT, Australia

2. University of Technology, Sydney, NSW, Australia

A feature of hybrids is that as well as the F1 generation outperforming its parents the F2 generation loses much of the hybrid advantage and shows great heterogeneity in phenotype, particularly in biomass and flowering time. Among the F2 plants are some that have a biomass and flowering time phenotype very similar to the F1 hybrids. We selected a number of these F1 like plants and using recurrent selection in the F3, F4 and F5 generations we produced a number of independent pure breeding lines with a biomass similar to the F1 plants and increased seed yield. We called these lines "hybrid mimics" because although they are not hybrids they resemble them in phenotypes such as biomass, flowering time and seed yield. Genomic analysis of these hybrid mimics shows them to be essentially homozygous but with genomic segments from both parents. This result indicates that the large biomass phenotype of hybrids does not require heterozygosity. The hybrid phenotype must be brought about by interactions between these genomic segments both genetic and epigenetic. Gene expression analysis indicates that in the hybrid mimic lines many genes show the same level of expression as in the F1 hybrids. Pathways for auxin and for cell expansion are upregulated in all the hybrid mimic lines and in the F1 hybrids providing a potential mechanism for the large phenotype.

Hybrid mimic technology may be applicable to crops and provide a method of achieving hybrid level yields without the costs of hybrid seed production. They may also allow crops in which there is no hybrid seed technology to gain the yield advantages of hybrids. These benefits may be important in developing countries where the cost of hybrid seed prevents its use.

261

An educators journey

Dawn Gleeson[1]

1. School of Bioscience, University of Melbourne, Melbourne, VIC, Australia

What are the factors that influence the choice of a career? Is it genotype, environment or as we tell our students, most commonly an interaction between the two. This talk will examine one such career and the influences that steered that career into Genetics education. What factors make a career in Genetics education exhilarating? Having the pleasure of relating to students the exciting new discoveries across three decades. Adapting to the changes in the tertiary education system. Acknowledging the increase in the diversity and expectations of our students. Managing the new technologies in teaching. Training the students to be ready, not only for the changing employment opportunities, but also the impact of genetics on their life.

262

Understanding Evolutionary Rate Variation in Viruses

Sebastian Duchene[1]

1. University of Melbourne, Melbourne, VIC, Australia

Estimating the rate of molecular evolution over time is crucial for understanding the processes and forces that shape biological diversity. To this end, viruses are particularly useful study organisms because they evolve much more quickly than cellular organisms. For example, the rates of evolution of influenza viruses are up to six orders of magnitude higher than those of mitochondrial DNA in vertebrates. This is one of the main explanations for the large proportion of infectious diseases caused by viruses. Having high rates of evolution allows viruses to evade the immune response of their hosts and to infect different host species. Accurate estimates of rates of evolution are also necessary for inferring evolutionary timescales, which provide information about the emergence and long-term evolution of viruses. However, estimating rates of evolution is a challenging task that requires different statistical and computational tools. I will present a set of studies dealing with rates of evolution and the estimation of evolutionary timescales in viruses. Analyses of empirical data demonstrate the effect of natural selection and mutational saturation on viral rates of evolution. In particular, there is a time-dependen pattern in rate estimates, which is sustained across different groups of viruses. Finally, I will discuss computational tools to improve the accuracy and precision of estimates of evolutionary rates timescales in viruses and other microorganisms.

263

Avian Poxvirus Identified as Major Extinction Threat to Hawaiian Forest Birds

Rebecca Cann[1], Leonard Freed[2]

1. University of Hawaii at Manoa, John A. Burns School of Medicine, Kailua, HAWAII, United States

2. Biology, University of Hawaii at Manoa, Honolulu, Hawaii, USA

Much attention has been focused on climate warming and the dangers of an increase in prevalence of mosquito-borne pathogens around the world, including zoonosis shared between humans and animals. Avian malaria has increased at upper elevations in Hawaii, even at elevations too cool for malarial development in the mosquito vector. Three haplotypes of avian poxvirus occur in Hawaii, and a virulent one (canary pox) shares the same vector as avian malaria in Hawaii, so it too should increase in incidence, driven by the greater vector capacity of upper elevation populations of *Culex quinquefasciatus*. It is possible to test this prediction with Hawaiian birds at 1900m elevation, and we show that mosquito –transmitted poxvirus now effectively limits population size of species within the entire community. Although molecular methods greatly increase the precision by which conservation biologists can identify ongoing threats, action to halt or reverse the damage caused by introduced pathogens will require adopting diverse strategies, including the culling of alien species to reduce competition for resources and allow time for disease resistant genotypes to spread. Hawaiian birds are teaching us new insights into how evolutionary reversals in morphology and behavior can be tracked using accurate molecular taxonomies that nest haplotypes island by island, but these populations are, unfortunately, continuing to disappear before our eyes.

264

Genomics of parallel evolution and speciation during repeated adaptive radiations in cichlid fishes

Axel Meyer[1], Andreas Kautt[1], Gonzalo Machado-Schiffiano[1], Frederico Henning[1]

1. University of Konstanz, Konstanz, Germany

Genome-wide data reveal an often highly heterogeneous pattern of genomic divergence during speciation. Disentangling the dynamic effects of divergent selection and gene flow from the stochastic effects of a population's demographic history remains difficult. Cases of recent sympatric speciation seem promising in this regard, since selection had to be strong enough to overcome gene flow and may thus leave distinct signatures in the genome, while the confounding effect of genetic drift is usually assumed to be negligible. Midas cichlid fishes (Amphilophus sp.) inhabiting small and isolated crater lakes in Nicaragua form young (only < 2 - 22,000 years old) and monophyletic flocks of endemic species and no geographic barriers exist in these lakes. Thus, sympatric ecological speciation is the most likely mode of speciation. Moreover, Midas cichlids provide natural replicates of this process and several species seem to be at different stages of the speciation continuum. Based a comprehensive RADseq data set (of > 700 individuals) the source population and the crater lakes to (i) infer the demographic history of the speciation processes of small-scale radiations of Midas cichlids based on the joint site frequency spectrum and full-likelihood coalescent simulations and (ii) take this information into account when describing their differentiation at a genomic level. Overall, we find evidence for colonizations by small founder populations, speciation in sympatry, and document a highly heterogeneous landscape of genomic differentiation. Particular focus was on patterns of differentiation in previously identified QTL regions presumably underlying adaptive traits (such as body shape, coloration, dentition and lips) involved in ecological speciation in these species.

265

Convergent regulatory evolution and the origin of flightlessness in palaeognathous birds

Scott V. Edwards[1]

1. Harvard University, Cambridge, MA, United States

A major question in evolutionary biology, one posed in the mid-1970s by Allan Wilson, is whether genic or regulatory evolution underlies the diversity of phenotypes observed in nature. An additional question that Allan Wilson asked was whether convergent phenotypes are driven by convergence at the level of the genome. We have approached these questions by using comparative genomics to understand the genomic basis of flightlessness in palaeognathous birds, which include the flightless ratites (emu, ostrich, kiwi, etc) and the volant tinamous of the New World. In contrast to the early trees produced by Allan Wilson and others, recent phylogenetic work suggests that tinamous are embedded within the ratite radiation and that flight was likely lost multiple times within the group. We have produced 10 new high-quality palaoengath genomes and aligned these to 32 additional genomes from birds and non-avian reptiles in an easily searchable genome browser. We called ~1.5 million conserved non-exonic elements (CNEEs) in these genomes, of which ~284,000 were greater than 50 bp, and identified those that have undergone relaxation or acceleration in individual ratite lineages or convergently in multiple ratite lineages. We identified ~15,000 CNEEs undergoing acceleration at least one ratite lineage, a CNEE subset that is enriched for elements that have arisen since the avian ancestor, as well as significant numbers of CNEEs and coding regions that have undergone acceleration or adaptive evolution in multiple ratite lineages. We find that the genes nearest to convergently accelerating CNEEs are enriched for roles in development and that many of these show intriguing patterns of expression in developing chickens. Current work is focused on functionally examining the role of specific CNEEs in driving gene expression in chickens, emus and rheas. Overall our results suggest a strong role for non-coding regulatory evolution in the origin of flightlessness in palaeognathous birds.

266

Allan Wilson as I Knew Him

James A. Lake[1]

1. UCLA, Los Angeles, CA, United States

Allan Wilson was a complex but brilliant person. I don't recall when I first met him, but it was probably in about 1986 at a conference on evolutionary biology held during a heat wave in Northern Sweden. I still remember when Alan introduced the audience to PCR by sequencing a small region of about 30 nucleotides from 20 different kangaroo rats. It was immediately clear that restriction fragment analyses were a thing of the past and that his lab had begun a new gene sequencing era.

In 1989 Allan's lab published a now famous paper using PCR on '..the extinct marsupial wolf". That paper led Michael Crichton to publish "Jurassic Park", after a brief "sabbatical" in Allan's lab. Living in Los Angeles, it happened that one of our best friends was the Unit Production Manager for Jurassic Park and I learned a lot about the adventures that happened during its filming - much of which I'll tell you.

My wife and I also got to know Allan and his wife pretty well. We spent a week with him at a rented beach house in Pajaro Dunes. We played tennis, walked on the beach and talked scientific politics. He had a tremendous insight into scientific politics, and I'll talk about some of these. Among other things he helped me organize the first UCLA Winter Sloan Schools on Molecular Evolution. This school trained the first generation of molecular evolutionary biologists. Allan and I served on the initial Evolutionary Biology Board of the Sloan Foundation and we had great fun in shaping this new field, until Allan's death in 1991. Allan was a major supporter of the field of Molecular Evolution.

267

Sequencing our way towards understanding global eukaryotic biodiversity

W. Kelley Thomas[1]

1. University of New Hampshire - Hubbard Center for Genome Studies, Durham, NH, United States

One of the grand challenges is biology is to understand the patterns of evolutionary diversity and ecological roles for the vast unseen "creatures" that inhabit our planet. In the late 1980s the advent of the polymerase chain reaction (PCR) transformed molecular evolution. At the same time Allan Wilson's lab was investigating well known fundamental questions in evolutionary biology, including our own origins it also played a major role in developing PCR based applications in population genetics and systematics. Among the most impactful was a simple demonstration in 1989 that a single pair of common PCR primers could be used to amplify homologous regions of mitochondrial genomes from diverse animals. That methodology ultimately transformed population biology and gave rise to modern day meta-barcoding. In my group we have used these technologies to explore the patterns of diversity in the small eukaryotic phyla, and most recently applied those approaches to investigate the consequences of the Deep Water Horizon Oil Spill in the Gulf of Mexico. We now have the opportunity to apply those same approaches coupled to Next Generation Sequencing in an attempt to test for the existence of ecologically meaningful patterns of biogeographic structure among these small organisms long thought to be largely unstructured.

268

The evolution and changing ecology of the human microbiome

Christina Warinner[1]

1. University of Oklahoma, Oklahoma, United States

The newly appreciated importance of the human microbiome raises many questions as to its origin, evolution, and changing ecology. The application of advanced genomic and proteomic sequencing technologies to ancient human microbiomes, such as coprolites (paleofeces) and dental calculus (calcified plaque), as well as to contemporary microbiomes in traditional and industrialized societies, allows us to advance understanding of the evolutionary history of our microbial self and its impact on human health today.

269

Ancient DNA, archaeology and linguistics: on the chances, limitations and perils of working on the edge of time

Wolfgang Haak[1, 2]

1. Archaeogenetics, Max Planck Institute for the Science of Human History, Jena, THURINGIA, Germany

2. School of Biological Sciences, The University of Adelaide, Adelaide, South Australia, Australia

When population genetic research tackles questions of human dispersal, migration and demographic histories, the latter are often adopted from or inspired by research questions and debates in the neighboring fields of archaeology and anthropology. Ancient DNA studies are no exception and rely directly on samples from archaeologically defined contexts to study events in human prehistory. The time-travelling feature of ancient DNA, i.e. the ability to record data from before, during and after particular turning points in the human past, has helped to elucidate many of the long-standing debates in adjacent disciplines. A prime example is the Neolithic transition, which saw the expansion of early farmers rather than 'just' the idea of farming being spread from its Near Eastern origin. However, outside the field of genetics, archaeologists and historians do not always embrace these findings, and genetic results are often received with skepticism. Aside from over-simplified peopling and demographic scenarios, colleagues from the humanities caution the unsupervised use of simplistic classifications (culture = people = language) or feel deterred by 'deterministic' approaches (biologism). This talk will showcase a range of current misunderstandings between both fields in an attempt to bridge the divide that still lingers between archaeology and genetics.

270

Insights from Neandertal microbiota on the history of human health and disease

Laura S Weyrich[1], **Keith Dobney**[2], **Alan Cooper**[1]

1. University of Adelaide, Adelaide, SA, Australia

2. University of Aberdeen, Aberdeen, Scotland

Despite numerous descriptions about the interactions between Neandertals and anatomically modern humans, little is known about the diseases and microorganisms that were shared between these hominids. DNA sequencing of preserved dental plaque (calculus) from ancient hominid skeletons now provides a unique opportunity to examine the evolution of ancient diseases and commensal bacterial species (microbiota) through time. Exploring the past evolutionary history of microbiota is critical for modern human health, as alterations to these commensal species are now linked to many diseases, including obesity, diabetes, heart disease, and others. Using a shotgun sequencing approach, we obtained ancient bacterial DNA from 53 ancient dental calculus samples collected from chimpanzees, Neandertals, and ancient and modern humans. We reconstruct the first oral microbiota of an extinct species, and reveal nearly 200 bacterial species present in Neandertals, including some that are shared with modern humans today and are linked to tooth decay (*Streptococcus mutans*) and periodontal disease (*Porphyromonas*, *Tanerella*, and *Treponema* taxa). By comparing microbiota from Neandertal and a wide-range of ancient humans, we also reconstructed the evolutionary history of the human microbiota over the past 55,000 years. We observe a significant shift in ancient hunter-gatherer microbiota ~35,000 yBP in Europe and ~3,000 yBP in Africa, when the microbiota diverged from that shared amongst Neandertals, chimpanzees, and ancient humans. When ancient humans switched to agriculture, another significant change in microbiota was observed in both Europe and Africa. This change was demarked by an increase of Fusobacteria microorganisms linked to polymicrobial oral diseases, revealing why oral health significantly declined after ancient humans adopted agriculture. Together, this data from extinct and extant hominids over the last 55,000 years provides the first record of human microbiota evolution, and a means to understand why these bacterial communities are now linked to disease.

271

Genomic History of Upper Paleolithic Europeans

Qiaomei Fu[1], Cosimo Posth[2], Mateja Hajdinjak[3], Martin Petr[3], Janet Kelso[3], Nick Patterson[4], Svante Pääbo[3], David Reich[5], Johannes Krause[2]

1. Key Laboratory of Vertebrate Evolution and Human Origins of Chinese Academy of Sciences, Beijing, China

2. Max Planck Institute for the Science of Human History, Jena, THURINGIA, Germany

3. Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany

4. Broad Institute of MIT and Harvard, Cambridge Massachusetts, USA

5. Department of Genetics, Harvard Medical School, Boston, USA

Little is currently known about the genetic history of ancient Europeans before the advent of agriculture ~8,500 years ago. Here we have analysed genome-wide data from 51 modern human remains that span around 40,000 years of Eurasian prehistory. Over this time, the proportion of Neanderthal DNA decreased from 3–6% to around 2%, consistent with natural selection against Neanderthal variants in modern humans. Whereas the earliest modern humans in Europe did not contribute substantially to present-day Europeans, all individuals between ~37,000 and ~14,000 years ago descended from a single founder population which forms part of the ancestry of present-day Europeans. A ~35,000-year-old individual from northwest Europe represents an early branch of this founder population which was then displaced across a broad region, before reappearing in southwest Europe during the last ice age ~19,000 years ago. During the major warming period after ~14,000 years ago, a new genetic component related to present-day Near Easterners appears in Europe. These results document how population turnover and migration have been recurring themes of European pre-history.

272

The spread of farming across the Mediterranean to Iberia and its role in shaping ancient and modern European genomes

Cristina Valdiosera[1], Torsten Günther[2], Ricardo Rodriguez[3], Juan Carlos Vera[4], Eneko Iriarte[5], Rafael Martínez[6], Helena Malmström[2], Juan Luis Arsuaga[3], José Miguel Carretero[5], Anders Götherström[7], Mattias Jakobsson[2]

1. La Trobe University, Bundoora, ACT, Australia

2. Evolutionary Biology Center, Uppsala University, Uppsala, Sweden

3. Palaeontology, Centro Mixto UCM-ISCIII de Evolución y Comportamiento Humanos, Madrid, Spain

4. Historia, Geografía y Antropología, Universidad de Huelva, Huelva, Spain

5. Paleontología, Universidad de Burgos, Burgos, Spain

6. Prehistoria y Arqueología, Universidad de Granada, Granada, Spain

7. Archaeology, Stockholm University, Stockholm, Sweden

The transition to a farming lifestyle was one of the major episodes of innovation in the history of our species and it has been the subject of intense archaeological research for decades[1]. In the past few years, archaeogenetic studies have been crucial in resolving some of the longstanding questions about the Neolithisation of Europe[2-6]. Here, we analyse new genome sequence data from 13 early farmers from Spain and compare them to previously published modern day and ancient genomes from Europe, North Africa and the Near East. We show that the first farmers to arrive to the Iberian Peninsula during the Neolithic, followed a coastal Mediterranean route bringing farming practices with them. These Neolithic individuals show a similar genetic structure across the North, North East and South of Iberia with no evidence of north African influence. Furthermore, we observe a certain degree of genetic differentiation between Early Neolithic Iberian and Central European farmers. An indication of at least two founding populations of early Neolithic Europeans (one that arrived via the Mediterranean coast and the other via the Danube basin into Central Europe). Among all early European farmers the Iberian Neolithic groups show the highest genetic affinities to present-day Sardinians suggesting that the modern population of the island are relatively direct descendants of these early Mediterranean farmers. Later, Iberian Chalcolithic populations derive from the interbreeding between incoming farmers and native hunter-gatherers[3]. In turn, these Chaloclithic groups are closely related to modern day Basques whom appeared to be isolated since the Late Neolithic[3]. Finally, genetic similarities between Middle to Late Neolithic farmers from Ireland and Iberia potentially suggest the latter to be the origin of the Megalithic culture which spread along the Atlantic coast and later reached the British Isles and Scandinavia[7].

1. Childe, V. The Dawn of European Civilisation. (Kegan Paul, 1925).
2. Gamba, C. et al. Genome flux and stasis in a five millennium transect of European prehistory. Nature communications 5, 5257, doi:10.1038/ncomms6257 (2014).
3. Günther, T. et al. Ancient genomes link early farmers from Atapuerca in Spain to modern-day Basques. Proc. Natl. Acad. Sci. U. S. A. 112, 11917-11922, doi:10.1073/pnas.1509851112 (2015).
4. Lazaridis, I. et al. Ancient human genomes suggest three ancestral populations for present-day Europeans. Nature 513, 409-413, doi:10.1038/nature13673 (2014).
5. Skoglund, P. et al. Genomic diversity and admixture differs for Stone-Age Scandinavian foragers and farmers. Science 344, 747-750, doi:10.1126/science.1253448 (2014).
6. Omrak, A. et al. Genomic Evidence Establishes Anatolia as the Source of the European Neolithic Gene Pool. Curr. Biol. 26, 270-275, doi:10.1016/j.cub.2015.12.019 (2016).
7. Cassidy, L. M. et al. Neolithic and Bronze Age migration to Ireland and establishment of the insular Atlantic genome. Proc. Natl. Acad. Sci. U. S. A. 113, 368-373, doi:10.1073/pnas.1518445113 (2016).

273

The Neolithic Revolution developed among geographically adjacent but genetically distinct populations

Marcos Gallego Llorente[1], Sarah Connell[2], Eppie Jones[1], Deborah Merrett[3], Yeonsu Jeon[4], Anders Eriksson[1], Veronika Siska[1], Cristina Gamba[2], Christopher Meiklejohn[5], Robert Beyer[1], Sungwon Jeon[4], Yunsung Cho[4], Michi Hofreiter[6], Jong Bhak[4], Andrea Manica[1], Ron Pinhasi[2]

1. University of Cambridge, Cambridge, CAMBRIDGESHIRE, United Kingdom

2. School of Archaeology and Earth Institute, University College Dublin, Belfield, Dublin, Ireland

3. Department of Archaeology, Simon Fraser University, Burnaby, BC, Canada

4. The Genomics Institute, Ulsan National Institute of Science and Technology (UNIST), Ulsan, South Korea

5. Department of Anthropology, University of Winnipeg, Winnipeg, MB, Canada

6. Evolutionary Adaptive Genomics, Institute for Biochemistry and Biology, Department of Mathematics and Natural Sciences, University of Potsdam, Potsdam, Germany

The shift from hunter-gathering to food production, the so-called Neolithic Revolution, profoundly changed human societies. Whilst much is known about the mode of spread of people and domesticates into Europe during the Neolithic period, the origin of this cultural package in the Ancient Near East and Anatolia is poorly understood. By sequencing the whole genome (1.39x) of an early Neolithic woman from Ganj Dareh, in the Zagros Mountains of Iran, we show that the eastern part of the Ancient Near East was inhabited by a population genetically most similar to hunter-gatherers from the Caucasus but distinct from the Neolithic Anatolian people who later brought food production into Europe. Despite their key role in developing the Neolithic package, the inhabitants of Ganj Dareh made little direct genetic contribution to modern European populations, suggesting they were somewhat isolated from other populations in this region. Their high frequency of short runs of homozygosity, comparable to other early Neolithic farmers, suggests that they overwintered the Last Glacial Maximum in a climatically favourable area, where they may have received a genetic contribution from a population basal to modern Eurasians. Thus, the Neolithic package was developed by at least two genetically-distinct groups which coexisted next to each other, implying a degree of cultural yet little genetic exchange among them.

274

The genomics of adaptive radiation

Walter Salzburger[1]

1. University of Basel, Basel, BS, Switzerland

More than one and a half centuries after the publication of Charles R. Darwin's *The Origin of Species*, the identification of the processes governing the emergence of novel species remains a fundamental question to biology. Why is it that some groups have diversified in a seemingly explosive manner, while other lineages have remained unvaried over millions of years? What are the external factors and environmental conditions that promote diversification? And what is the molecular basis of adaptation, evolutionary innovation and diversification? Demonstrating particularly clear evidence of the power of natural selection, adaptive radiations emerge as outstanding systems for studying the mechanisms of evolution. The first wave of genomic investigation across major archetypal adaptive radiations – such as Darwin's finches, anole lizards, threespine stickleback fish, and cichlid fishes in the East African Great Lakes – is starting to shed light on the molecular basis of adaptive diversification. Here, I provide an overview of genomic studies on adaptive radiations, with a particular focus on cichlid fish.

275

Shuffling of modular enhancers by adaptive introgression generates convergence and novelty in butterfly patterns

Chris Jiggins[1]

1. University of Cambridge, Cambridge, UK, United Kingdom

An important goal in evolutionary biology is to understand the genetic changes underlying novel morphological structures. I discuss the origins of complex wing patterns found among neotropical *Heliconius* butterflies. Genome sequence data from 100s of individuals across two major radiations has identified narrow regions associated with distinct colour pattern elements. We hypothesise that these modules in non-coding sequence represent distinct *cis*-regulatory loci that control expression of just 3-4 key genes, including the transcription factor *optix* and the morphogen *WntA*, which in turn control pattern variation across *Heliconius*. Phylogenetic analysis of these elements demonstrated that they have distinct evolutionary histories and that novel adaptive morphological variation was created by shuffling these *cis*-regulatory modules through recombination between divergent lineages. In addition, recombination of modules into different combinations within species further contributes to diversity. Analysis of the timing of diversification supports the hypothesis of introgression moving regulatory modules between species, rather than shared ancestral variation, as divergence can be much younger at wing pattern loci relative to species divergence. I therefore argue that shuffling of existing enhancer elements both within and between species provides a mechanism for rapid diversification and generation of novel morphological combinations during adaptive radiation

276

The genetic basis for beak diversification and adaptive evolution in Darwin's finches

Sangeet Lamichhaney[1], **Jonas Berglund**[1], **Makus Sällman Almen**[1], **Fan Han**[1], **Matthew T. Webster**[1], **B. Rosemary Grant**[2], **Peter R. Grant**[2], **Leif Andersson**[1]

1. Department of Medical Biochemistry and Microbiology , Uppsala University, Uppsala, Sweden

2. Department of Ecology and Evolutionary Biology, Princeton University, Princeton, New Jersey, USA

Darwin's finches from the Galápagos are a celebrated study model regarding the processes of natural selection and adaptive evolution. Adaptation to the striking ecological differences throughout Galápagos have led to rapid diversification and speciation in these birds. This has resulted in remarkable diversity in their morphology, specifically the shape and size of the beaks. Long-term field studies in the past have documented that beaks in Darwin's finches correspond to specific feeding niche they occupy and evolve by natural selection in response to limiting food resources and interspecific competition. We have done extensive genomic characterization of the entire Darwin's finch radiation by whole-genome sequencing 180 birds that included all currently recognized species. Genome-wide comparisons among species with different beak shapes (blunt and pointed) and beak sizes (large, medium and small) identified candidate genes associated with beak morphology. The strongest association to beak shape was *ALX1*, a transcription factor involved in craniofacial development. Similarly *HMGA2,* a transcriptional regulating factor previously linked to human height and body size in other species, showed the strongest association to beak size. Interestingly, both loci were segregating in the medium ground finch, the species with considerable diversity in beak morphology. Genotyping these two loci in additional medium ground finches and comparison with individual measurements of beak shape and size confirmed the association and suggested additive effects of these loci, as heterozygotes showed intermediate beak types compared with the two homozygotes. In addition, *HMGA2* loci also had played a critical role in a documented character displacement episode in Darwin's finches when medium ground finches diverged from their competitor, the large ground finch, during a severe drought. In conclusion, we have provided evidence of two loci with major effects on beak morphology (shape and size), which probably were the most important trait during the adaptive radiation of Darwin's finches.

277

Repeated evolution of behavior via polygenic adaptation in Malawi cichlid fish

Ryan York[1], Chinar Patil[2], Kawther Abdilleh[2], Todd Streelman[2], Russ Fernald[1], Hunter Fraser[1]

1. Stanford University, Stanford, CA, United States

2. Georgia Institute of Technology, Atlanta, GA, United States

Identifying the proximate and ultimate causes that drive behavioral evolution is one of biology's great challenges. Central to this effort is resolving the nature of genes important for behavior. We leverage the substantial natural variation observed in a courtship behavior amongst the Malawi cichlid adaptive radiation to do so. Males of up to 200 Malawi cichlid species build mating nests (bowers) out of sand for the purpose of attracting females. Two basic types of bowers exist: "pits" (depressions) and "castles" (mounds). Phylogenetic analyses indicate that pits are the likely ancestral form from which castles have repeatedly evolved, possibly more than a dozen times. We assayed the genetic basis of bower building by sequencing the genomes of 10 castle-building and 12 pit-digging species from diverse genera. Analyses of genetic differentiation indicate that >6,000 SNPs are perfectly fixed between pit and castle species. Genes associated with these variants (~1,400) are related by function and are involved in key neural processes such as axon guidance and glutamate signaling. Since the majority of variants fall in non-coding regions we investigated the role of *cis*-regulatory divergence by performing RNA-seq on whole brain samples from F1 hybrids of a castle-builder and a pit-digger. We calculated differential allele-specific expression (diffASE) genome-wide between animals engaged in bower building and animals kept in isolation. We find hundreds of genes that possess both significant regulatory divergence and behaviorally driven expression changes and that are associated with fixed SNPs. These genes show significant overlap with the functional categories identified above, indicating that the repeated evolution of a bower-building has been driven by polygenic *cis*-regulatory adaptation. Our findings suggest a new and unexpected mechanism through which behaviors may evolve and provide grounds for future mechanistic studies of the genes involved.

278

Testing the role of selection and demography in driving a rapid postglacial radiation in the songbird genus *Junco*

Guillermo Friis[1], Angeles de Cara[2], Borja Mila[1]

1. Museo Nacional de Ciencias Naturales, Madrid, Spain

2. Lab d'Ecoanthropologie et Ethnobiologie, UMR 7206, Museum National d'Histoire Naturelle, CNRS, Paris, ILE DE FRANCE, France

Rapid evolutionary radiations likely result from the combined effects of selective pressures and demographic processes. The songbird genus *Junco* of North America includes several phenotypically divergent northern forms which have arisen within the last 10,000 years as a result of a rapid postglacial expansion across North America. These northern forms contrast with more genetically divergent ancestral southern forms that are geographically isolated, yet show moderate phenotypic divergence. In addition to the role of geographic and historical factors, the wide range of habitat types and the highly diversified patterns of plumage coloration suggest the role of multiple selective factors in driving lineage divergence. Here we combine whole-genome and genotyping-by-sequencing (GBS) data to reconstruct the evolutionary history of the genus and explore how genomic patterns of variation relate to demographic events and selective factors. We use MSMC (multiple sequentially Markovian coalescent) and G-Phocs (generalized phylogenetic coalescent sampler) to test the population-expansion and recent-divergence hypotheses in northern junco forms. MSMC revealed recent demographic expansions for all the northern junco forms, reinforcing the hypothesis of multiple lineage differentiation driven by a postglacial northward recolonization of North America. We also used Bayescan to calculate $F_{ST}$and posterior probabilities per SNP to infer selection-mediated divergence, and found no specific regions of high differentiation but rather a number of highly divergent variants scattered across the genome. This suggests the role of selection acting on numerous loci across the genome from the early stages of the speciation process. Our analyses show that juncos represent one of the fastest radiations documented in birds, with major roles for historical, demographic and selective factors.

279

High resolution populations genomics reveals an extremely complex dynamics of early speciation in Afrotropical malaria vectors.

Giordano Botta[1], Alistair Miles[2], Beniamino Caputo[1], Nick Harding[2], Alessandra della Torre[1], Dominic Kwiatkowski[2]

1. University of Rome La Sapienza, Roma, RM, Italy

2. Wellcome Trust Centre for Human genetics, University of Oxford, Oxford, Select State/Province, United Kingdom

Genomes of natural populations are continually exposed to a number of evolutionary forces, driving them towards either divergence, homogenisation, or a mosaicism of the two. Our ability to detect and describe genomic signatures left by these processes depends upon the level of resolution available in terms of loci, individuals and populations sequenced. The *Anopheles gambiae* complex is an established model system to study evolutionary dynamics between recently diverged species. The *Anopheles gambiae* 1000 Genomes Consortium made recently available an extraordinary genomic data resource consisting of 765 genomes sequenced at high coverage from 8 countries spanning Sub-Saharan Africa of the main malaria vectors *An. gambiae* and *An. coluzzii*. Here we describe at a very fine scale level complex dynamics of early speciation forging the genomes in contrasting way, leading to various levels of reproductive isolation and to the establishment of two new hybrid forms at the opposite edges of the species range. We virtually described all possible directions a geographically non-homogenised speciation process can take: i) local genomic divergence surrounded by remarkable homogeneity due to gene flow, ii) adaptive introgression events followed by the reestablishment of reproductive barriers, iii) reinforcement of reproductive barriers in sympatry after populations expansions, iv) origins of hybrid forms with local complete replacement of pure individuals. *An. coluzzii*experienced stronger overall selective pressure than *An. gambiae*, consistent with a recent speciation process driven by niche expansion in marginal habitats. Moreover, we identified candidate genes and described mutations under divergent selection between *An. gambiae* and *An. coluzzii* giving the first molecular support to previous hypothesis on mechanisms for the onset and maintenance of reproductive isolation necessary to species formation.

280

Patterns of nucleotide changes in human populations

Sankar Subramanian[1]

1. Griffith University, Nathan, QLD, Australia

A number of previous studies showed that heterozygosity declines with the increase in the geographic distance of human populations from East Africa. This relationship was also manifested in the observed higher number of polymorphisms in African compared to non-African populations. The cause for these trends was attributed to the effect of genetic drift resulting from serial bottlenecks (or founder effects) occurred during the range expansion of human populations. Although these studies reported about the quantitative difference in the number of polymorphisms, the effects of drift on the patterns of nucleotide change is unclear. Using large-scale data from whole genome and SNP array we show significant difference in the types of nucleotide change between global populations. We observed a much higher of AT to GC changes in African compared to non-African populations. Furthermore the magnitude of this difference negatively correlates with the geographic distance of the populations from East Africa. These results could be explained based on the combined effects of biased gene conversion and genetic drift.

300

Genetic analyses of five late Neandertal individuals

Mateja Hajdinjak[1], Martin Petr[1], Udo Stenzel[1], Hélène Rougier[2], Isabelle Crevecoeur[3], Patrick Semal[4], Marie Soressi[5], Sahra Talamo[6], Jean-Jacques Hublin[6], Ivan Gušić[7], Željko Kućan[7], Pavao Rudan[7], Liubov V. Golovanova[8], Vladimir B. Doronichev[8], Cosimo Posth[9], Johannes Krause[9], Petra Korlević[1], Sarah Nagel[1], Birgit Nickel[1], Kay Prüfer[1], Janet Kelso[1], Qiaomei Fu[10], Matthias Meyer[1], Svante Pääbo[1]

1. Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany

2. Department of Anthropology, California State University Northridge, Northridge, California, USA

3. Université de Bordeaux, CNRS, Bordeaux, France

4. Royal Belgian Institute of Natural Sciences, Brussels, Belgium

5. Faculty of Archaeology, Leiden University, Leiden, The Netherlands

6. Department of Human Evolution, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany

7. Croatian Academy of Sciences and Arts, Zagreb, Croatia

8. ANO Laboratory of Prehistory, St. Petersburg, Russia

9. Max Planck Institute for the Science of Human History, Jena, Germany

10. Key Laboratory of Vertebrate Evolution and Human Origins of Chinese Academy of Sciences, IVPP, Beijing, China

Comparisons of the Neandertal genome to present-day human genomes have revealed that ~1-2% of present-day genomes outside Africa come from Neandertals and it has been suggested that a major part of the admixture took place in the Levant between 47-65 kya [1, 2]. However, it has also been shown that a ~42,000-year-old modern human from Romania had a Neandertal ancestor four to six generations back in his family tree, indicating that the admixture between modern humans and Neandertals was not restricted to a single event in Near East [3]. To better understand late Neandertal populations and the interactions between Neandertals and modern humans we are investigating the genomes of European Neandertals from the time when they or their immediate ancestors could have met modern humans.
We identified five late Neandertal specimens – from the Troisième caverne of Goyet and Spy in Belgium, Les Cottés in France, Vindija Cave in Croatia and Mezmaiskaya Cave in Russia – where the fraction of endogenous sequences are between 6% and 64% after depleting microbial contamination through hypochlorite treatment. We have sequenced the nuclear genomes of these individuals to an average coverage between 1- and 2.7-fold. Present-day human DNA contamination varies between ~1% and ~2.5% for the nuclear and mitochondrial DNA sequences, respectively.
Based on the number of DNA fragments recovered from the X chromosome and the autosomes, we determined that the specimens from Goyet, Les Cottés and Vindija were females, whereas the Spy and Mezmaiskaya 2 specimens were males. We further use these genomes to determine population structure among late Neandertals and their relationships to the Neandertals that contributed DNA to present-day humans, as well as to determine whether there was gene flow from early modern humans into these late Neandertals.

1. [1] Green, R. E. et al. A draft sequence of the Neandertal genome. Science 328, 710-722, doi:10.1126/science.1188021 (2010)
2. [2] Prüfer, K. et al. The complete genome sequence of a Neanderthal from the Altai Mountains. Nature 505, 43-49, doi:10.1038/nature12886 (2014).
3. [3] Fu, Q. et al. An early modern human from Romania with a recent Neanderthal ancestor. Nature, doi:10.1038/nature14558 (2015)

301

Insights into South American population history, from ancient DNA from Tierra del Fuego.

Thiseas Christos Lamnidis[1], Zuzana Faltyskova[2, 3], Luca Pagani[2], Hannes Schroeder[4, 5], Susana Morano[6], Federico Abascal[7], Alfredo Prieto[6, 8], Yolanda Espinosa-Parrilla[9, 10], Carles Laluzela-Fox[9], Maru Mormina[2, 11], Yali Xue[7], Chris Tyler-Smith[7], Toomas Kivisild[2], Richard Durbin[7], Stephan Schiffels[1]

1. Department of Archaeogenetics, Max Planck Institute for the Science of Human History, Jena, Germany

2. Division of Biological Anthropology, University of Cambridge, Cambridge, UK

3. Research Department of Genetics, Evolution and Environment, University College London, London, UK

4. Centre of GeoGenetics, University of Copenhagen, Copenhagen, Denmark

5. Faculty of Archaeology, Leiden University, Leiden, the Netherlands

6. Institute of Patagonia, University of Magallanes, Punta Arenas, Chile

7. Wellcome Trust Sanger Institute, Cambridge, UK

8. Autonomous University of Barcelona, Barcelona, Spain

9. Institute of Evolutionary Biology, CSIC-Pompeu Fabra University, Barcelona, Spain

10. School of Medicine, University of Magallanes, Punta Arenas, Chile

11. Department of Applied Social Sciences, University of Winchester, Winchester, UK

According to genetic evidence, the ancestors of Native Americans entered the Americas through the Beringian land bridge circa 16,000 years ago(1). However, the exact arrival time and dispersal routes across the double continent and South America in particular remain unclear. Due to their unusual cranial morphology, Fuegians from the south of South America have been suggested to belong to a relic Paleoamerican population(2). Here we address this question using ancient DNA from five human samples from Tierra del Fuego dated between 200 and 4,600 years ago. We generated the full nuclear genomes of these samples, with average depth ranging between 2x and 10x, and analysed them jointly with modern sequences from public data sets. F3 and D statistics revealed no ancestral component in the Fuegians other than typical Native American, consistent with previous genetic studies(3), unlike the Karitiana population from Brazil that shows an Australasian component(4). We used Rarecoal(5) to construct a population model of 1000 Genome populations including post-colonial admixture events and discerned that the

Fuegians are more closely related to Colombians than to Peruvians, and more closely related to those than to Mexicans. We estimate that Fuegians diverged from the Colombian branch ~10,500 years ago, shortly after the Colombian/Peruvian divergence (~11,700 years ago). Our results demonstrate that rare variant analyses have the resolution required to make population history inferences between highly similar populations, like those of South America.

1. Llamas B, Fehren-Schmitz L, Valverde G, Soubrier J, Mallick S, et al. 2016. Ancient mitochondrial DNA provides high-resolution time scale of the peopling of the Americas. Science Advances 2
2. Lahr, MM. 1995. Patterns of modern human diversification: Implications for Amerindian origins. Am J Phys Anthropol 38
3. Raghavan M, Steinrücken M, Harris K, Schiffels S, Rasmussen S, et al. 2015. POPULATION GENETICS. Genomic evidence for the Pleistocene and recent population history of Native Americans. Science 349
4. Skoglund P, Mallick S, Bortolini MC, Chennagiri N, Hünemeier T, Petzl-Erler ML, Salzano FM, Patterson N, Reich D. 2015. Genetic evidence for two founding populations of the Americas. Nature 525
5. Schiffels S, Haak W, Paajanen P, Llamas B, Popescu E, et al. 2016. Iron Age and Anglo-Saxon genomes from East England reveal British migration history. Nat Commun 7

302

The preliminary report for the deep sequencing of the prehistoric Jomon genome from the Japanese archipelago

Hiroki OOTA[1], Takashi Gakuhari[1], Hiroki Shibata[2], Tadashi Imanishi[3], Ryan Schmidt[1], Martin Sikora[4], Thorfinn S Korneliussen[4], Hiromi Matsumae[1], Takehiro Sato[5], Takafumi Katsumura[1], Kae Koganebuchi[1], Takayuki Nishimura[6], Shigeki Nakagome[7], Shuhei Mano[8], Ryosuke Kimura[9], Kentaro Shimizu[10], Atsushi Tajima[5], Shoji Kawamura[11], Shigekazu Higuchi[2], Akira Yasukouchi[2], Tetsuaki Wakebe[6], Toshiyuki Tsurumoto[6], Motoyuki Ogawa[1], Tsunehiko Hanihara[1], Hajime Ishida[9], Yasuhiro Yamada[12], Hiroimi Shitara[13], Nobuo Shigehara[14], Eske Willerslev[4]

1. Kitasato University School of Medicine, Sagamihara City, KANAGAWA, Japan

2. Kyushu University, Fukuoka

3. Tokai University, Isehara, KANAGAWA, Japan

4. Centre for GeoGenetics Natural History Museum of Denmark, Oester Voldgade 5-7, DK-1350 Copenhagen, Denmark

5. University of Kanazawa, Kanazawa, Japan

6. University of Nagasaki, Nagasaki, Japan

7. University of Chicago, Chicago, USA

8. The institute of Statistical Mathematics, Tachikawa, Japan

9. University of Ryukyus, Naha, Okinawa, Japan

10. University of Zurich, Zurich, Switzerland

11. University of Tokyo, Kashiwa, Japan

12. National Museum of Japanese History, Sakura, Japan

13. University of Tokyo, Tokyo, Japan

14. Nara National Research Institute for Cultural Properties, Nara, Japan

After the late-Paleolithic period in the Japanese archipelago, the Jomon culture starts around 15,000 years ago, and the Yayoi culture took place of it around 3,000 years ago. The former is a culture by the indigenous people who have hunting-gathering life style, whereas the latter is a culture of large-scale rice cultivation that would be brought by immigrants (and their descendants) from the East Asian continent. Many of previous studies based on skeletal remains have described that the Jomon people have been morphologically homogeneous for more than 10,000 years. The dual structure model for peopling history of modern Japanese has been proposed; according to the model, the Jomon and the immigrants have gradually admixed since the Yayoi period, and the modern main-island Japanese have been formed at last. Our recent study based on computer simulation using genome-wide SNP data from modern Chinese and Hokkaido Ainu, however, has estimated that the admixture between the Jomon and the immigrants occurred 5~6,000 years ago that is twice older than the estimates based on the archaeological evidences, suggesting a possibility of gene flow from the East Asian continent before the Yayoi culture starts.

To resolve the discrepancy, we organized a research team of the Jomon genome sequencing, including physical anthropologists, physiological anthropologists, archaeologists, and researchers of statistics, genomics, and population genetics. We conducted prescreening for the Jomon specimens excavated from various sites by using the next generation sequencer (NGS). The states of preservation of DNAs from the specimens in the Japanese archipelago were commonly even worse than those in Europe and America, because of acid soils, and warm and high humid climate. But, a couple of specimens showed more than 1.0% mapping ratio. Here we present the results of preliminary analyses, and discuss about feasibility of the deep sequencing of Jomon.

303

Genome-wide allele frequency estimates from population level ultra-low coverage aDNA samples.

Luca Pagani[1, 2], Christiana L Scheib[2], Craig Cessford[2], John E Robb[2], Toomas Kivisild[2]

1. Estonian Biocentre, Tartu, Estonia

2. Department of Archaeology and Anthropology, University of Cambridge, Cambridge, United Kingdom

Recent improvements in DNA extraction techniques from ancient human remains have dramatically expanded the availability of starting material for population level genetic studies. However sequencing costs remain the main limiting factor shaping the balance between number of samples and sequencing depth in any such study design. Consequently, population level studies are often characterized by low or ultra-low coverage sequences, invariably affecting the quality of the obtained genotype calls.

Here we introduce a novel approach to obtain population level allele frequency estimates from pools of ~20 ultra-low (0.1-1x) coverage samples. Particularly we will focus on two British populations from the same geographic site, before and after the 14th century Plague epidemic. Furthermore we show through empirical simulations how reads coming from multiple individuals from the same population can be combined to form a “chimeric” genome, representative of an average good-quality individual from that population.

The population level allele frequencies hence estimated, can be used to detect selective sweeps occurred during the Plague epidemic and, compared with the modern GBR samples from the 1000 Genomes Project inform us on putative adaptive responses to pathogens.

304

Uncovering long-term evolutionary dynamics of hepatitis B virus using ancient DNA

Zoe Patterson Ross[1], Jennifer Klunk[2], Melanie Kuch[2], Sebastián Duchêne[1], Hendrik Poinar[2], Edward C. Holmes[1]

1. University of Sydney, Sydney, NEW SOUTH WALES, Australia

2. McMaster University, Hamilton, Ontario, Canada

Studying ancient pathogens is critical to improve our understanding of modern epidemics and the evolutionary processes that govern them. However, determining the authenticity of ancient DNA (aDNA) samples is difficult. Here, I will describe a hepatitis B virus recovered from a 16th century mummy. Statistical analysis of the damage patterns indicated an authentic pattern of degradation as expected for aDNA, however the phylogenetic analysis indicated that the virus was closely related to modern samples. Moreover, molecular clock analyses were inconclusive in estimating the age of this sample. A probable explanation for this paradox is that this close relationship is due to saturation by reverse mutation leading to loss of the phylogenetic signal over the time frame of several hundred years. This study presents a phylogenetic framework that may be used for determining the robustness of age estimates in aDNA work.

305

EAGER: Efficient Ancient Genome Reconstruction

Alexander Peltzer[2, 1], Günter Jäger[1], Alexander Herbig[2], Alexander Seitz[1], Christian Kniep, Johannes Krause[2], Kay Nieselt[1]

1. Integrative Transcriptomics, Eberhard Karls Universitaet , Tuebingen, Baden-Wuerttemberg, Germany

2. Max Planck Institute for the Science of Human History, Jena, THURINGIA, Germany

More and more research projects are investigating scientific questions using ancient DNA (aDNA). Many of the available methods and pipelines for the analysis of aDNA sequencing data are difficult in application and require complex configuration and manual assembly of analysis tools. Especially with modern sequencing technology at hand, larger and more datasets are created, which require methods for the efficient and scalable analysis of such kinds of data. To address these challenges, we introduce the EAGER pipeline.

EAGER provides state-of-the-art methods to perform quality control, mapping, authentication, contamination estimation and genotyping of NGS data in an accessible manner. Our pipeline incorporates several new methods for paired-end read merging, improved duplication removal and mapping that are specifically tailored to improve the analysis output for aDNA projects. Users are provided with a graphical user interface (GUI) to configure the pipeline, hiding much of the complexity of the analytical processes. The complete pipeline is distributed as a Docker image, thus there is no requirement to install all the underlying tools independently. All the required methods and tools are provided within a single image to the end user. To further increase the usability of the pipeline, users are provided with automatically generated extensive reports of their analysis runs. These include important analysis statistics in Excel compatible formats, making the assessment of whole-genome sequencing runs very easy.

We have successfully utilized the pipeline in several projects for both bacterial and human genetic data. EAGER can reconstruct the genome of an ancient human aDNA dataset of ~100GB in size in less than one week. The pipeline is provided to the public on GitHub and our webpage.

EAGER can provide a well-defined standard for aDNA analysis, specifically incorporating the needs of labs with limited bioinformatics resources, additionally minimizing both administrative and installation effort for users.

1. Peltzer, A., Jäger, G., Herbig, A., Seitz, A., Kniep, C., Krause, J., & Nieselt, K. (2016). EAGER: efficient ancient genome reconstruction. Genome Biology, 17(1), 60. doi:10.1186/s13059-016-0918-z
2. Andrews, S. (2010). FastQC: A quality control tool for high throughput sequence data. Reference Source.
3. Li, H., & Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics (Oxford, England), 26(5), 589–95. doi:10.1093/bioinformatics/btp698
4. Langmead, B., Trapnell, C., Pop, M., & Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol, 10(3), R25. doi:10.1186/gb-2009-10-3-r25
5. Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv Preprint arXiv:1303.3997.
6. Ginolhac, A., Rasmussen, M., Gilbert, M. T. P., Willerslev, E., & Orlando, L. (2011). mapDamage: testing for damage patterns in ancient DNA sequences. Bioinformatics (Oxford, England), 27(15), 2153–5. doi:10.1093/bioinformatics/btr347
7. Renaud, G., Slon, V., Duggan, A. T., & Kelso, J. (2015). Schmutzi: estimation of contamination and endogenous mitochondrial consensus calling for ancient DNA. Genome Biology, 16(1), 224. doi:10.1186/s13059-015-0776-0
8. McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., … DePristo, M. A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Research, 20(9), 1297–303. doi:10.1101/gr.107524.110
9. Korneliussen, T. S., Albrechtsen, A., & Nielsen, R. (2014). ANGSD: Analysis of Next Generation Sequencing Data. BMC Bioinformatics, 15(1), 356. doi:10.1186/s12859-014-0356-4
10. Schubert, M., Ermini, L., Der Sarkissian, C., Jonsson, H., Ginolhac, A., Schaefer, R., … Orlando, L. (2014). Characterization of ancient and modern genomes by SNP detection and phylogenomic and metagenomic analysis using PALEOMIX. Nat Protoc, 9(5), 1056–1082. doi:10.1038/nprot.2014.063

306

The mitochondrial genome of an archaic hominin from southwestern Germany

Cosimo Posth[1, 2], Christoph Wißing[3], Keiko Kitagawa[4], Kurt Wehrberger[5], Nicholas Conard[1, 6], Claus-Joachim Kind[7], Hervé Bocherens[3, 8], Johannes Krause[1, 2, 9]

1. Institute for Archaeological Sciences, University of Tübingen, Tübingen, Germany

2. Max Planck Institute for the Science of Human History, Jena, Germany

3. Department of Geosciences, Biogeology, University of Tübingen, Tübingen, Germany

4. Department of Prehistory, National Museum of Natural History, Paris, France

5. Ulmer Museum, Ulm, Germany

6. Department of Early Prehistory and Quaternary Ecology, University of Tübingen, Tübingen, Germany

7. State Office for Cultural Heritage Baden-Württemberg, Esslingen, Germany

8. Senckenberg Centre for Human Evolution and Palaeoenvironment, University of Tübingen, Tübingen, Germany

9. Senckenberg Centre for Human Evolution and Palaeoenvironment, University of Tübingen, Tübingen, Germany

In 1937 a right hominin femur shaft with archaic morphology was excavated from the cave of Hohlenstein-Stadel in the Swabian Jura of southwestern Germany. The specimen was discovered in a layer corresponding to the Middle Paleolithic. Attempts to directly date the femur were inconsistent and indicated that the bone may be out of range for radiocarbon dating. Here we present genetic analyses of the femur shaft in order to assess the age and phylogenetic position of this ancient hominin bone. Hybridization capture in combination with next generation sequencing were used to reconstruct the complete mitochondrial genome (mtDNA). A phylogenetic comparison with modern human, Denisovan and an extended dataset of Neandertal mtDNA sequences revealed a closer relationship of the femur's mtDNA to Neandertals. The Hohlenstein-Stadel mtDNA falls, however, basal to all other Neandertal individuals displaying the deepest divergence and a short phylogenetic branch length. Those results indicate an age for the hominin femur notably older than previously suggested. Using a Bayesian statistic framework we performed a molecular dating to identify the temporal range of the Hohlenstein-Stadel femur and provide important insights into the mitochondrial diversity of the Neandertal populations through the Late Pleistocene.

307

Signatures of archaic adaptive introgression in present-day human populations

Fernando Racimo[1], David Gokhman[2], Matteo Fumagalli[3], Davide Marnetto[4], Torben Hansen[5], Ida Moltke[6], Anders Albrechtsen[6], Liran Carmel[2], Emilia Huerta-Sánchez[7], Rasmus Nielsen[1, 8]

1. Department of Integrative Biology, University of California Berkeley, Berkeley, California, United States

2. Department of Genetics, The Hebrew University of Jerusalem, Jerusalem, Israel

3. Department of Genetics, Evolution, and Environment, University College London, London, UK

4. Department of Molecular Biotechnology and Health Sciences, University of Torino, Turin, Italy

5. The Novo Nordisk Foundation Center for Basic Metabolic Research, University of Copenhagen, Copenhagen, Denmark

6. The Bioinformatics Centre, Department of Biology, University of Copenhagen, Copenhagen, Denmark

7. School of Natural Sciences, University of California Merced, Merced, California, United States

8. Department of Statistics, University of California Berkeley, Berkeley, California, United States

Comparisons of DNA from archaic and modern humans show that these groups interbred, and in some cases received an evolutionary advantage from doing so – a process known as adaptive introgression (AI). However, introgression by itself changes both the haplotype structure and the distribution of allele frequencies in a genomic region, thus confounding traditional tests for detecting positive selection that do not model introgression. Here we explore models involving both introgression and positive selection to investigate the behavior of various statistics under AI. We find that the number and allelic frequencies of sites that are uniquely shared between archaic humans and specific present-day populations are particularly useful for detecting AI. We then examine the 1000 Genomes dataset to characterize the landscape of uniquely shared archaic alleles in human populations. Finally, we identify regions that were likely subject to adaptive introgression and discuss some of the most promising candidate genes located in these regions. One of these is the *TBX15/WARS2* region, which has been previously found to be under positive selection in Greenlandic Inuit and has also been associated with body fat distribution in humans. We show that an archaic haplotype was likely introduced into Eurasians by a population closely related to Denisovans, and was then subject to positive selection in a much larger geographic region than just Greenland. Furthermore, the introgressed SNPs are associated with changes in expression and methylation patterns of *WARS2* and *TBX15* in multiple tissues, suggesting the introgressed haplotype may have altered regulatory patterns in the region. Our study helps to elucidate the landscape of adaptive events possible through introgression from archaic hominids. It also helps us understand the consequences of these events on the phenotypic make-up of modern humans as they expanded around the globe.

308

FINGERPRINT: Computational filtering of targeted sequences from environmental contaminants

Michael S Rosenberg[1]

1. Arizona State University, Tempe, AZ, United States

DNA sequencing is often performed on mixed samples in NextGen laboratories, both intentionally as in meta-genomics or unintentionally as when the material of interest is mixed with environmental contaminants. The latter is particularly a problem in ancient DNA projects, where differential preservation makes avoidance of contamination difficult. The problem of contamination is substantially worse when the material of interest is a bacterium or virus which (a) cannot be readily separated or purified prior to sequencing and (b) is likely to resemble common environmental contaminants. A number of pre- and post-sequencing protocols have been developed to sort or filter NGS reads from mixed samples, but have proven to be inadequate for removing closely related contaminants from aDNA samples. FINGERPRINT is a new, simple bioinformatics approach based on targeted *k*-mer genome profiling that more readily filter NGS reads into desired and contaminant bins prior to sequence assembly. Data from an ancient tuberculosis sequencing project are used to illustrate the power and efficacy of the method.

310

The Neolithic in Northeast Asia in light of a 7,700 year-old genome

Veronika Siska[1], Eppie Ruth Jones[1], Tatiana Balueva[2], Jong Bhak[3], Daniel G. Bradley[4], Yunsung Cho[5], Anders Eriksson[1], Marcos Gallego Llorente[1], Michi Hofreiter[6], Sungwon Jeon[3], Hakmin Kim[3], Hyunho Kim[5], Kyusang Lee[3], Ron Pinhasi[7], Elizaveta Veselovskaya[2], Andrea Manica[1]

1. University of Cambridge, Cambridge, United Kingdom

2. Institute of Ethnology and Anthropology Russian Academy of Science, Moscow, Russia

3. The Genomics Institute, Ulsan National Institute of Science and Technology, Ulsan, Republic of Korea

4. Smurfit Institute of Genetics, Trinity College Dublin, Dublin, Ireland

5. Theragen BiO Institute, TheragenEtex, Suwon, Republic of Korea

6. Institute for Biochemistry and Biology, Faculty for Mathematics and Natural Sciences, University of Potsdam, Potsdam-Golm, Germany

7. School of Archaeology and Earth Institute, University College Dublin, Dublin, Ireland

Ancient genomes have revolutionised our understanding of Holocene prehistory in Western Eurasia, but Asia has received limited attention. Here we report genomic data from two individuals from an early Neolithic site, Devil's Gate (7742-7638 cal years before present) in East Asia, on the border between Russia and Korea. These hunter-gatherers are genetically most similar to geographically close populations from the Amur Basin, in particular the Ulchi, implying a high level of continuity in this region over most of the Holocene. Japanese and Koreans, who live further south, also showed genetic affinity to Devil's Gate; both these modern day populations were best described as mixes of a population close to Devil's Gate, likely the ancestral hunter-gatherers who inhabited that region, and modern populations from southern China and South-East Asia, possibly linked to the onset of the Neolithic around 8,000 years ago in the region. Thus, in contrast to Western Eurasia, East Asia experienced a higher degree of continuity, with little input in the northern regions and an integrations of incoming farmers and local hunter-gatherers further south.

311

Signs of mobility and migration in the megalithic graves of Western Sweden?

Emma M Svensson[1], Malou Blank Bäcklund[2], Arielle Munters[1], Luciana Simoes[1], Jan Storå[3], Mattias Jakobsson[1]

1. Uppsala University, Uppsala, UPPLAND, Sweden

2. Department of Historical Studies, Göteborgs Universitet, Göteborg, Sverige

3. The Osteo-archaeology Research Laboratory, Stockholm University, Stockholm, Sweden

During the Neolithic period in Scandinavia the Funnel Beaker complex gave way to the Battle Axe complex, which was later replaced by the more homogeneous Late Neolithic complex. This culture produced the finest flint work in Scandinavian Prehistory and the last megalithic tombs.

Immigration from various regions has been proposed as an explanation for the geographic distribution of megalithic burials in western Sweden, another alternative is that the tombs were used by populations from large areas. However, osteological and archaeological research suggests that the graves were used by local family groups. Earlier research suggests that 25% of the middle Neolithic population buried in the megalithic tombs were of non-local birth. However, in the Late Neolithic there was an increase in human mobility and about 60% of the buried individuals were non-locals. It is interesting to investigate the genetics over time in this area. Are there any traces of population shifts while still maintaining the same burial practices?

We use the gallery grave at Torbjörnstorp as a model site for investigation of mobility. Here the individuals analysed all date to around 1800 BC cal, the second half of the Late Neolithic period in Scandinavia. However, the megalithic graves have been used for successive burials over a long time and even though the skeletal remains are fairly well preserved, the bones have been moved to make way for new burials and the bone material is often fragmented. The aDNA analyses can in this case be important for disentangling the demography of the individuals buried here, and coupled to isotopic results this can be used for discussing mobility patterns.

Using the combined knowledge gained from archaeology, isotope analysis and ancient genomes we can address questions of migration and mobility in the late Scandinavian Neolithic.

312

The Aboriginal Heritage Project: Reconstructing the Genetic History of Aboriginal Australia with Ancient DNA

Ray Tobler[1], Wolfgang Haak[1, 2], Ali Abdullah-Highfold[3], Shane Agius[3], Isabel O'Laughlin[3], Amy O'Donohue[3], Fran Zilio[3], Lewis O'Brien[4], Leslie Williams[5], Keryn Walshe[3], Peter Sutton[3], John Mitchell[6], Chris Tyler-Smith[7], Emma Kowal[8], David Reich[9], Alan Cooper[1]

1. Australian Centre for Ancient DNA, University of Adelaide, Adelaide, SA, Australia

2. Max Planck Institute for the Science of Human History, Jena, Germany

3. South Australian Museum, Adelaide, South Australia, Australia

4. Education, Arts and Social Sciences, University of South Australia, Adelaide, South Australia, Australia

5. Department of Communities, Child Safety and Disability Services, Queensland Government, Brisbane, Queensland, Australia

6. La Trobe Universtiy, Melbourne, Victoria, Australia

7. Sanger Institute, Cambridge, UK

8. Deakin University, Melbourne, Victoria, Australia

9. Howard Hughes Medical Institute, Harvard University, Boston, Massachusetts, USA

We outline the **Aboriginal Heritage Project**: a collaboration between the Australian Centre of Ancient DNA (ACAD) and the South Australian Museum (SAM) that aims to reconstruct the genetic history of Aboriginal Australia. The project leverages the unparalleled collection of 5000+ hair samples curated by the SAM along with cultural, morphometric and genealogical data, which were collated by Joseph B. Birdsell and Norman B. Tindale during extensive anthropological expeditions across Australia between 1926 and 1963. The broad geographic sampling and unique combination of ancient DNA and deep genealogies contained in the SAM collection provides perhaps our best opportunity to understand Aboriginal Australian genetic history prior to European colonisation. We present our outreach activities, which crucially involve re-consenting the hair samples through in-depth consultation with Aboriginal families and communities, along with preliminary phylogeographic analyses which reveal that pre-colonial Aboriginal Australia was characterised by deeply structured populations dating back to the initial colonisation of the continent. Ultimately, we aim to create a reference map that allows current and future generations of Aboriginal Australians to retrace their ancestry - including the displaced Stolen Generations and their descendants - and illuminates this remarkable but still largely unknown chapter of human genetic history to the rest of world.

313

Positive selection on balanced standing variation

Philip Kleinert[1], Cesare de Filippo[1], Aida M Andres[1]

1. Department of Genetics, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany

In small populations, adaptation to the environment is likely limited by the availability of advantageous alleles. This is because in groups with small effective population size the low effective rate of new mutations and the low levels of segregating variation limit the probability of beneficial alleles to be present, at the right time and place, in the population. This affects positive selection both on new and on neutral standing variation, and both with monogenic and polygenic adaptation.

Variants under balancing selection hold the potential to mediate fast adaptations, if environmental change drives a shift from balancing to positive selection. This is because balanced alleles necessarily affect phenotype and fitness, and can be maintained by selection even in small populations.

Other non-neutral variants can also be maintained if they are tightly linked to sites under balancing selection. We discuss this model of positive selection on previously balanced alleles and show, using forward simulations, that it has high potential to mediate fast, local adaptation in human populations. We also introduce a statistic (DIFFSS) that jointly considers the site frequency spectra of two populations and has high power to identify positive selection on loci previously under balancing selection. We applied DIFFSS on the genomes of African and Eurasian human populations and, after combining these results with Approximate Bayesian Computations, identified dozens of loci with signatures of this mode of adaptation.

314

Hill-Robertson Interference Maintained by Red Queen Dynamics Favours the Evolution of Sex

Jack Da Silva[1]

1. University of Adelaide, Adelaide, SA, Australia

Although it is well established theoretically that selective interference among mutations coexisting on different genetic backgrounds (Hill-Robertson interference) favours meiotic recombination, genome-wide mean rates of mutation and strengths of selection appear too low to support selective interference as the mechanism favouring recombination in nature. A possible solution to this discrepancy between theory and observation is that selection is intermittently very strong due to the antagonistic coevolution between a host and its parasites. The Red Queen theory posits that such coevolution generates negative fitness epistasis among loci, which in turn generates negative linkages among alleles that favour recombination. However, Red Queen dynamics without epistasis may provide the ecological conditions that maintain strong and frequent selective interference. This hypothesis is developed here using recursion equations to simulate Hill-Robertson interference with Red Queen dynamics. A method is developed that allows the frequencies of haplotypes with an arbitrary number of loci to be calculated after recombination. Simulations show that an allele for recombination at a modifier locus is most strongly favoured when there are many selected loci and when the strength of selection, the mutation rate and population size are moderately large. Recombination is also most strongly favoured when a host and its parasites have similar evolutionary dynamics.

315

Detection of pathways affected by positive selection in primate lineages ancestral to humans

Josephine Daub[1, 2, 3], **Sébastien Moretti**[2, 4], **Iakov Davidov**[2, 4], **Laurent Excoffier**[1, 2], **Marc Robinson-Rechavi**[2, 4]

1. CMPG, Institute of Ecology and Evolution, University of Berne, Berne, Switzerland

2. SIB Swiss Institute of Bioinformatics, Lausanne, Switzerland

3. Institute of Evolutionary Biology (UPF-CSIC), Barcelona, Spain

4. Dept. of Ecology and Evolution, University of Lausanne, Lausanne, Switzerland

Gene set enrichment approaches have been increasingly successful in finding signals of recent polygenic selection in the human genome. In this study, we aim at detecting biological pathways affected by positive selection in more ancient human evolutionary history, that is in four branches of the primate tree that lead to modern humans. We tested all available protein coding gene trees of the Primates clade for signals of adaptation in the four branches mentioned above, using the likelihood-based branch site test of positive selection. The results of these locus-specific tests were then used as input for a gene set enrichment test, where whole pathways are globally scored for a signal of positive selection, instead of focusing only on outlier "significant" genes. We identified several pathways enriched for signals of positive selection, which are mainly involved in immune response, sensory perception, metabolism, and energy production. These pathway-level results were highly significant, at odds with an absence of any functional enrichment when only focusing on top scoring genes. Interestingly, several gene sets are found significant at multiple levels in the phylogeny, but in such cases different genes are responsible for the selection signal in the different branches, suggesting that the same function has been optimized in different ways at different times in primate evolution.

316

The impact of genetic architecture, demography and genetic load in natural selection inferences

Angeles de Cara[1], **Frederic Austerlitz**

1. Lab d'Ecoanthropologie et Ethnobiologie, UMR 7206, Museum National d'Histoire Naturelle, CNRS, Paris, ILE DE FRANCE, France

The relative importance of natural selection and genetic drift in maintaining genetic diversity is key to explain current diversity and future adaptive potential. By comparing allelic frequencies, expected and observed heterozygosity or haplotype homozygosity within and between populations, we can disentangle selection and demography in statistical inferences. Finding regions under selection is central to understanding the processes of adaptation and speciation. Recent developments include haplotype-based methods to infer selection within populations. We benchmark iHS, nSL and H12 in simulated data with selection on a polygenic trait, and show that these methods work best when the traits are mildly polygenic. We also test their power on known regions under selection using *Heliconius* butterflies. In a second step, we study the impact of demographic history, through a simulation study based on the history of Central African human populations.This is a good model since it is settled by agricultural and hunter-gatherer populations having experienced different adaptive histories and for which demographic histories have been inferred. We analyse genomic diversity by simulating the split of an ancestral human population into agriculturalists (AGR) and hunter-gatherers (HG), where the AGR population has been expanding during the last 20000 years. We consider several scenarios, in which a quantitative trait has been under selection since agriculture emerged 5000 years ago in the AGR population. Either all populations were in mutation-drift equilibrium, or they are under background selection together with selection on the quantitative trait on the AGR population only. We benchmark common statistics (Fst, dxy, iHS, nSL and H12) on the loci that code for the quantitative trait under selection in order to analyse the impact of demography and selection on their power and false positive rates. Lastly, we show the impact of differing demographic histories on the genetic load of both populations.

317

Temporal clustering of allele frequency trajectories in asexual population under neutral equilibrium versus positive selection

Yuseob Kim[1], **Yeongseon Park**[1]

1. Ewha Womans University, Seodaemun-gu, SEOUL, South Korea

Recent advances in DNA sequencing allows us to observe the evolution of populations in real time by tracking the change in the frequencies of variant nucleotides through time, particularly for microbes and viruses that reproduce asexually. To understand the causes and patterns of evolutionary changes through DNA sequence data, it is important to define and use an appropriate parameter to summarize the evolutionary dynamics of a population. Temporal clustering of fixations, occurring when multiple alleles at different loci are fixed together due to various evolutionary forces, can be quantified by the index of dispersion (IOD = variance/mean). In the previous study, IODs of asexual populations under neutral equilibrium or under recurrent positive selection were shown to be sufficiently larger than one, indicating that temporal clustering of fixation events occurs on wide parameter space for asexually evolving systems. In this study, we ran simulations and recorded times at which mutant alleles reach an intermediate threshold frequency. The temporal clustering of these events is also quantified by IOD. We obtained IODs for ten different threshold frequencies, under both neutral and selection models. There is a qualitative difference in neutral vs. selection models in the relationship between the threshold frequency and IOD: increase in IOD with increasing threshold frequency is much steeper in the neutral model than in selection models. This result may allow us to assess the presence of recurrent positive selection in asexually-evolving populations with genomic data. We also elaborate on theoretical explanation for this result in terms of genetic hitchhiking and clonal interference.

318

Sherpas share genetic variations with Tibetans for high-altitude adaptation

Sushil Bhandari[1], Chaoying Cui[2], Xiaoming Zhang[1], Yangla X[2], Lan Liu[2], Ouzhuluobu X[2], Baimakangzhuo X[2], Gonggalanzi X[2], Caijuan Bai[2], Bianba X[2], Yi Peng[1], Hui Zhang[1], Kun Xiang[1], Hong Shi[1], Shimin Liu[3], Gengdeng X[3], Tianyi Wu[3], Xuebin Qi[1], Bing Su[1]

1. Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, YUNNAN, China

2. High Altitude Medical Research Center, School of Medicine, Tibetan University, Lhasa, Tibet Autonomous Region, China

3. National Key Laboratory of High Altitude Medicine, High Altitude Medical Research Institute, Xining, Qinghai, China

Sherpas, a highlander population living in Khumbu region of Nepal, are well known for their superior climbing ability in Himalayas. But the genetic basis of their adaptation in highland region remains largely unknown. Here, we collected DNA samples of 582 Sherpas from Nepal and Tibet of China, and we measured their hemoglobin levels and degrees of blood oxygen saturation. We conducted genotyping of 32 sequence variants of three genetic loci, including EPAS1, EGLN1 and TED, which have been shown involved in high altitude adaptation of Tibetans. We found similar allele frequencies of all tested variants in Sherpas as compared with Tibetans, and most of the variants showed significant association with hemoglobin levels, but not with degrees of blood oxygen saturation. We propose that the shared sequence variants between Sherpas and Tibetans indicate a shared genetic basis for high altitude adaptation, consistent with the proposal that Sherpas are recently derived population from Tibetans and they inherited the adaptive variants for high altitude adaptation from their Tibetan ancestors.

319

Genome wide association studies with insecticide resistance lead us to selective sweep loci and beyond.

Charles Robin[1], Paul Battlay[1], Llewellyn Green[1], Joshua Schmidt[1], Alexandre Fournier-Level[1]

1. University of Melbourne, Parkville, Vic, Australia

Insecticide resistance evolution is a classic microevolutionary model typified by extreme selection pressure on natural populations. Insecticide resistance loci therefore rank as among the most likely of all insect loci to exhibit the footprints of selective sweeps. Indeed scans of the *Drosophila melanogaster* genome uncover loci that are the molecular targets of insecticides and genes known to detoxify insecticides. We have performed Genome wide association studies in Drosophila using insecticides with different modes of action and they confirm that major sweep loci are indeed associated with insecticide resistance. The GWAS also uncover other loci and provide us with an opportunity to search for polygenic signatures of selection. These efforts are aided by contrasting the genetic architecture of old insecticides and new insecticides that have not been around long enough to drive sweeps. Among the approaches we have explored are systems approaches that integrate polymorphism, trancriptome, and phenotypic data that reveal cohesive pathways perturbed by multiple xenobiotic compounds.

320

Analyzing the “snapshots” of directional selection spreading over space from incomplete selective sweeps in African Drosophila melanogaster

Thi Ha My Vy[1], Yuseob Kim[2]

1. Interdisciplinary Program for EcoCreative , Ewha Womans University, Seoul, South Korea

2. Department of Life Science, Ewha Womans University, Seoul, South Korea

Interplay between the spatial pattern of selective environment, the mode of directional selection, and the pattern of migration is expected to determine how beneficial alleles propagate over geographic regions. With the availability of whole-genome sequencing data from hundreds of wild-derived individuals in African populations of *Drosophila melanogaster* and the method of detecting loci under ongoing directional selection, it is now feasible to investigate the general mode and spatial patterns of positive directional selection. We scanned for signatures of incomplete selective sweeps in Rwanda and Zambia samples of *D. melanogaster* using our recently developed composite likelihood ratio (CLR) method and a haplotype homozygosity method (nS_L test). By choosing only significant signals detected by either one or both tests and re-examining the local topology of genealogical trees, we selected 46 loci with clear patterns of incomplete sweep for further analysis. The geographical distribution of the putatively beneficial haplotype at each locus was then obtained from a genealogical tree constructed for all individuals from 11 populations across Africa. We observed distinct spatial distributions of beneficial haplotype across loci, suggesting the operation of different modes of positive selection. To explain this range of results, simulations were performed under the island model of two subpopulations with selective pressure that vary in space (local vs. global selection) and time (constant vs. diminishing selection). More than half of the loci appear to be under simple selection with constant selective pressure. However, there are also many loci compatible with diminishing selection, for example due to heterozygous advantage. We also found a few loci under incomplete soft selective sweeps. One of them is characterized by a complex haplotype distribution that can only be explained by very high adaptive mutation rate and possibly heterozygous advantage.

321

The evolution of genomic content over generations of inbreeding maize lines

Kyria Roessler[1]

1. UCI, Newport Beach, CALIFORNIA, United States

Inbreeding depression has severe consequences to the genomic make-up of an organism, such as a reduction in genome size and purging of deleterious mutations. The purpose of this study was to track these genomic changes during the transition from outcrossing to self-fertilization. Since transposable elements (TEs) are deleterious, we expected that they would be purged and possibly cause the decline in genome size. To investigate this, we first used flow cytometry on the first six generations in 11 selfing landraces of maize (*Zea mays*). We found that three of the maize lines had a significant decrease in genome size, up to 20%, yet there was no significant change in the other lines. Next, to establish the cause for this reduction, we performed whole-genome sequencing on S1 and S6 generations, including all three with a genome size reduction. The sequenced reads were mapped to annotated genes, exemplar transposable elements, knob repeats and rDNA. We found in all the lines that there was indeed a loss in TE content. In addition, landraces that did not show a decrease in genome size contained more knob repeats and these knob repeats increased over generations. These results show there are changes in genomic content including TEs as a repercussion of selfing.

322

Loss of self-incompatibility in the allotetraploid *Arabidopsis kamchatica* by degradation of the male component

Chow Lih Yew[1], **Takashi Tsuchimatsu**[2], **Rie Shimizu-Inatsugi**[1], **Kentaro Shimizu**[1]

1. Department of Evolutionary Biology and Environmental Studies, University of Zürich, Zürich, Switzerland

2. Department of Biology, Faculty of Science, Chiba University, Chiba, Japan

The genetic basis of the recurrent evolutionary transition from outcrossing to selfing has been a major focus in evolutionary biology. The relationship between selfing and polyploidy has been debated for years. Polyploids, which are commonly found in plants, are suggested to self more frequently than their diploid relatives, although the underlying mechanism is still largely unknown.

The transition from outcrossing to selfing typically occurs through the breakdown of the self-incompatibility (SI) system. Sporophytic SI system prevents self-pollen tube growth via interactions between the male component, *S-locus cysteine-rich protein* (*SCR*) and the female component, *S-locus receptor kinase* (*SRK*) of the same *S*-haplogroups. While theories predict that mutations in male component, SCR, are more likely to be fixed as they have increased opportunities for outcrossing, empirical evidence is still lacking.

Here, we study the loss of SI in *Arabidopsis kamchatica* to reveal the molecular mechanisms underlying the evolution of self-compatibility in polyploids. *A. kamchatica* is a self-compatible allotetraploid species, originated through allopolyploidization between two predominantly outcrossing diploid species, *Arabidopsis halleri* and *Arabidopsis lyrata*. We applied high-throughput sequencing approach to isolate *SCR* genes from anther cDNAs, which has not been successful through conventional PCR method due to the short and highly polymorphic sequences among *SCR* genes of different *S*-haplogroups. Mutations were detected in *SCR* genes of *A. kamchatica*. Thus, transgenic method was developed to transform functionally restored *SCR* gene into *A. kamchatica* bearing functional female components. SI was recovered in transformed *A. kamchatica*, indicating that the degradation of SCR is primarily responsible for the loss of SI in *A. kamchatica*. This is in agreement with theoretical predictions that mutations in male components are more advantageous than mutations in female components. Moreover, dominance hierarchy among different *S*-haplogroups may be involved in the evolutionary loss of SI in allotetraploid *A. kamchatica*.

323

An experimental test of the emergence of Cellular Complexity by Constructive Neutral Evolution

Samuel Elley[1]

1. University of Canterbury, Christchurch, CANTERBURY, New Zealand

An experimental test of the emergence of Cellular Complexity by Constructive Neutral Evolution

Sam Elley, Alicia Sook-Wei Lai, Nicole E. Wheeler, Anthony M. Poole

Biomolecular Interaction Centre, School of Biological Sciences, University of Canterbury, Christchurch, New Zealand

Small, asexual populations are known to accumulate sublethal mutations over successive generations through a process called Muller's Ratchet. Under such conditions, sublethal mutations can become fixed, which can lead to fitness decline. We subjected lines of *E. coli* to single-cell bottlenecks and found that RNA slippage-type 'editing' can evolve under conditions favouring genetic drift. This observation fits the model of Covello and Gray, which proposes that the emergence of RNA editing is a product of constructive neutral evolution (Covello & Gray, 1993). This model proposes that complex cellular systems can originate by non-adaptive evolution.

We are now screening for the emergence of a broader range of cellular systems by constructive neutral evolution. Our lab group has evolved a lineage of *E.coli* that has been subjected to multiple population bottlenecks (which favours genetic drift) and has subsequently accumulated numerous sub-lethal mutations. We have subjected this line to further evolution under bottleneck relief (i.e. where natural selection can operate efficiently). We have analysed genomic data and screened for the emergence of altered substrate utilisation capacities using phenotype arrays, comparing data collected from the original ancestral strains, both pre-bottleneck and following the bottleneck, as well as data following bottleneck relief. In addition, we used a newly-described method (delta-bitscore) to assess the functional severity of mutations and mapped these onto cellular pathways (Wheeler, Barquist, Ashari Ghomi, Kingsley, & Gardner, 2015). We present the results of these analyses, and report on the impact of neutral evolutionary processes in shaping biomolecular systems.

1. Covello, P. S. & Gray, M. W. On the evolution of RNA editing. Trends Genet. 9, 265–8 (1993).
2. Wheeler, N. E., Barquist, L., Ashari Ghomi, F., Kingsley, R. A. & Gardner, P. P. A Profile-Based Method for Measuring the Impact of Genetic Variation. bioRxiv (Cold Spring Harbor Labs Journals, 2015).

324

Emergence of RNA editing in an evolution experiment

Alicia Lai[1, 2], **Anthony Poole**[1, 2]

1. Biomolecular Interaction Centre, University of Canterbury, Christchurch, New Zealand

2. School of Biological Sciences, University of Canterbury, Christchurch, New Zealand

RNA editing, the correction of genomic errors at the level of messenger RNA, has evolved multiple times independently, but it remains unclear exactly how this process has evolved. A model for the emergence of RNA editing proposed that RNA editing activity pre-exists but there is no substrate for editing activity to act upon. Subsequently, mutation results in appearance of editable nucleotide sites, which may be fixed by genetic drift. Fixation results in RNA editing becoming indispensable for protein production from affected genes[1]. We sought to test this model by asking whether slippage-type editing can evolve under experimental conditions designed to maximize the impact of genetic drift.

We previously showed that, in the bacterial endosymbiont *Buchnera*, RNA polymerase slips at long poly(A/T) tracts, leading to stochastic incorporation or removal of As or Us in the nascent messenger RNA. This results in a heterogeneous population of mRNAs. RNA polymerase slippage was thus shown to correct natural frameshift mutations in mRNA, but the stochasticity of correction suggested there was reduced expression efficiency, as only some mRNAs carried corrected reading frames[2].

In an evolution experiment using *Escherichia coli*, we subjected 10 lines to daily single-cell bottlenecks. Following 50 days of bottlenecking, one line showed an observable reduction in growth rate. Genome sequencing revealed the emergence of 38 frameshift mutations that appear to require slippage-type editing for gene expression. We present data showing that slippage-type editing indeed rescues frameshift mutations and that protein production is reduced, consistent with this type of mutation being slightly deleterious. Our results support the hypothesis that, under conditions favouring genetic drift, editing readily emerges. To our knowledge, this is the first experimental demonstration of the evolutionary drivers for the emergence of RNA editing.

1. [1] Covello, P.S. & Gray, M. W. (1993) On the evolution of RNA editing. Trends Genet., 9(8), 265-268.
2. [2] Tamas, I. et al. (2008) Endosymbiont gene functions impaired and rescued by polymerase infidelity at poly(A) tracts. Proc. Natl. Acad. Sci., 105, 14934-9.

325

New insight into human evolution from character compatibility analysis of the mitochondrial genomes of three primate taxa

Shaun Lehmann[1], Lars Jermiin[2], Michael Ott[2], Simon Easteal[1]

1. Australian National University, Acton, ACT, Australia

2. CSIRO, Canberra, ACT, Australia

Character compatibility analysis can improve the reliability of phylogenetic inference by removing 'noisy' sites in a sequence alignment. Two sites are 'compatible' if the same phylogeny can explain their character state distributions across the alignment. Incompatibility indicates homoplasy arising from back or parallel mutation. Sites are assigned compatibility scores proportional to their compatibility with other sites. Iterative removal of low compatibility sites and recalculation of compatibility scores identifies 'cliques' of sites that are perfectly inter-compatible. A clique with substantially more sites than all others is likely to indicate a true phylogeny. We evaluated the usefulness of this approach using three primate mitochondrial genome sequence alignments: 128 genomes from the four subspecies of the common chimpanzee, *Pan troglodytes*; 50 genomes from genera of the gibbon family Hylobatidae; and 87 diverse human (*Homo sapiens*) genomes. Sites were iteratively removed using the software package 'Shuffle' (Jermiin unpublished) and ML and MP phylogenies were estimated using the 'ape' and 'phangorn' packages in R. We predicted that for these closely related sequences compatibility analysis would provide greater confidence in well-separated branches and resolve uncertainty about poorly separated branches. As predicted, confidence in the main branches of the *P. troglodytes* tree, including those separating subspecies, increased; and previously uncertain gibbon lineages were resolved to be consistent with morphological and biogeographical evidence. Analysis of human sequences gave a very different result. Support for the major internal branches that separate deep African lineages collapsed, resulting in a 'star' phylogeny with all major lineages radiating from a single point. This unexpected result may reflect an relatively high proportion of rare variants in the human population that cause homoplasy at sites with pre-existing variation. We are currently investigating whether the pattern of variation in humans is better explained by population expansion or by directional natural selection.

326

The Origin of Life; solvable problems.

David Penny[1]

1. Massey University, Palmerston North, MANAWATUE, New Zealand

The Origin of Life was not originally considered a problem before about 1690; now it is considered a major (and difficult) issue. There are very many components to the question, but some can be addressed as relatively straightforward issues. One of these could be the temperatures at which the origin of life arose. Perhaps the favoured temperature is 'black smokers' at the bottom of the ocean (which might be at high temperatures). However, the standard equation for Gibbs Free Energy is DG = DH -TDS, where DG is thought to be the controlling factor, and DH might be a kinetic component, and DS the entropy (order) component. This equation could be interpreted as favouring a much lower temperature origin because of the high cost of temperature as the entropy increases. Certainly, folding of RNA is favoured at lower temperatures. Another question might be the numbers of nucleotides – this can be expressed as 'why 4'? Does having only 2 nucleotides fold too many ways for a given sequence? And 'Black membranes' form spontaneously, and it would be an interesting question to measure the temperatures at which simpler molecules form. The length of the amino acid code is a triplet. Why three? Possibly adding three nucleotides at a time might lead to improved accuracy – and thus to an increase in the 'Eigen limit' of the length of the RNA being coded for. Thus the potential ways of folding, membrane formation, and measurements of accuracy of replication are all important. The main issue is that they are all subject to experimental testing!

327

Experimental evidence that translation initiation in bacteria was invaded by a selfish genetic element.

Alannah Rickerby[1, 2], Ryan Catchpole[3, 1, 2], Stinus Lindgreen[1, 2], Katherine Donovan[1, 2], Brigitta Kurenbach[2], Jack Heinemann[2], Anthony Poole[1, 2]

1. Biomolecular Interaction Centre, University of Canterbury, Christchurch, New Zealand

2. School of Biological Sciences, University of Canterbury, Christchurch, New Zealand

3. Institut Pasteur, Paris, France

The core machinery for protein synthesis is universal to cellular life. However, idiosyncrasies exist that differentiate translation across Archaea, Bacteria and Eukarya. An example is found in bacteria, mitochondria and chloroplasts, where a formyl group is added to methionine prior to translation initiation. This formyl group is removed from the nascent polypeptide by peptide deformylase before protein production is complete, and appears to have no clear function. Despite this, it is essential to bacterial translation: interrupting formylation is deleterious. A well-conserved process that, if lost, leads to a severe phenotype is usually associated with important function. Our previous work indicates formylation and deformylation likely evolved from an ancient, plasmid-transmitted, toxin-antitoxin system capable of post-segregational killing (PSK). PSK systems get their name because of how they act: the antitoxin is more labile than the toxin, so segregating daughter cells that do not inherit the gene-pair die through action of the toxin. This creates an 'addiction', because cells can no longer lose the genes. We predicted that the formyl group is toxic, and that removal negates this toxicity. A line devoid of the formylase and deformylase genes was produced. While initially very unfit, after 1,500 generations of evolution, we found that growth rates of knockouts were identical to the wild-type lineages. We showed that there were mutational changes to genes involved in translation which enabled the cells to adapt to formylation loss. Moreover, introducing the genes on a plasmid elicited a PSK phenotype, as per our model. To reproduce initial adaptation to formylation, we reintroduced the genes on the *E. coli* chromosome, and performed a further 3,000-generation evolution experiment. Our results indicate that, despite our lines not requiring formylation, addiction to the gene-pair has reasserted itself, with formylation appearing essential. Our results suggest that formylation invaded and spread via addiction, with these genes becoming 'essential' as a result of PSK.

328

Exploring the origin of DNA through synthetic biology: is ribonucleotide reduction essential for deoxyribonucleotide synthesis *in vivo*?

Nellie Sibaeva[1], Dayong Si[2], Nobuyuki Horinouchi[2], Alicia Lai[1], Ryan Catchpole[3], Makoto Hibi[2], Koichi Tanaka[2], Jun Ogawa[2], Anthony Poole[1]

1. University of Canterbury, Christchurch, CANTERBURY, New Zealand

2. Kyoto University, Kyoto, Japan

3. Institut Pasteur, Paris, France

One focus of synthetic biology is the assessment of minimal gene sets required for cellular life. However, in early evolution, pathways in operation may no longer be used in modern cells. Fundamentally rewiring pathways of conserved central metabolism is the next frontier for this field, and is essential to assessing their origins. Our group is using synthetic biology to test a hypothesis about the emergence of DNA genomes. Ribonucleotide reduction is universally used in biology for *de novo* synthesis of deoxyribonucleotides. The chemical complexity of this reaction suggests that transition to DNA genomes may have occurred relatively late, possibly after the primary lineages of modern life began to diverge. However, ongoing interdomain horizontal gene transfer obscures the evolutionary history of ribonucleotide reductases (**RNR**s). However, an alternative, chemically simpler, pathway for the synthesis of deoxyribonucleotides may predate ribonucleotide reduction. In this pathway, production of deoxyribose is catalysed by deoxyriboaldolase (**DERA**). This pathway is ubiquitous, but naturally runs in the catabolic direction. Our goal is to establish the operation of the DERA pathway in the synthetic direction *in vivo,* aiming for complete functional replacement of ribonucleotide reduction by reverse DERA. If achieved, this will be the only free-living cell that produces its deoxyribonucleotides *de novo* without RNRs, bolstering the plausibility of an earlier transition to DNA genomes by simpler chemistry. Our group has generated knockouts of all RNRs in *Escherichia coli*, completely abolishing ribonucleotide reduction activity, resulting in deoxyribonucleoside (dNS) auxotrophic strains. We are screening for conditions where the DERA pathway permits cells to synthesise dNS in the absence of RNRs. This research highlights the value of synthetic biology for experimentally testing hypotheses on the origin of life.

329

Time-reversible vs. Lie-Markov phylogenetic models in the presence of inhomogeneous substitution processes.

Michael Woodhams[1], Jeremy Sumner[1], Barbara Holland[1], Michael Charleston[1]

1. University of Tasmania, Hobart, TAS, Australia

We present simulation results showing that Lie-Markov DNA evolution models outperform time-reversible models when inferring phylogenetic trees for sequence data that has been generated under a inhomogeneous process.

The Lie-Markov models form a hierarchy of DNA models which are ``multiplicatively closed'', meaning if we take any two Markov matrices generated by a Lie-Markov model, their matrix product is also generated by the same model.

This property is useful for analysing non-homogeneous DNA evolution, where model parameters can vary between branches.

Without this closure property, non-homogeneous analysis, where the fitted model is allowed to vary parameters across the tree, is mathematically inconsistent.

Many time-reversible (TR) DNA models currently in use fail to be multiplicatively closed (e.g. HKY, GTR), and are hence not Lie-Markov.

Our conclusions are obtained by comparing our hierarchy of 37 Lie-Markov models which enable differing transition/transversion rates, to a comparable hierarchy of 28 time-reversible models.

330

Chloroplast Phylogenomic Inference of Green Algae Relationships

Bojian Zhong[1], LinHua Suan[1], David Penny[2]

1. Nanjing Normal University, Nanjing, JIANGSU, China

2. Institute of Fundamental Sciences, Massey University, Palmerston North, New Zealand

The green algal phylum Chlorophyta has six diverse classes, but the phylogenetic relationship of the classes within Chlorophyta remains uncertain. In order to better understand the ancient Chlorophyta evolution, we have applied a site pattern sorting method to study compositional heterogeneity and the model fit in the green algal chloroplast genomic data. We show that the fastest-evolving sites are significantly correlated with among-site compositional heterogeneity, and these sites have a much poorer fit to the evolutionary model. Our phylogenomic analyses suggest that the class Chlorophyceae is a monophyletic group, and the classes Ulvophyceae, Trebouxiophyceae and Prasinophyceae are non-monophyletic groups. Our proposed phylogenetic tree of Chlorophyta will offer new insights to investigate ancient green algae evolution, and our analytical framework will provide a useful approach for evaluating and mitigating the potential errors of phylogenomic inferences.

331

Developmental Genomics of Sponges

Marcin Adamski[1], Sofia Fortunato[2, 3], Mary Laplante[3], Jon Bråte[4], Anthony Leon[3], Ilja Borisenko[3], Genevieve Aisbett [1], Rachel Rathjen[1], Meg McFetridge[1], Maja Adamska[1]

1. Australian National University, Acton, ACT, Australia

2. ARC Centre of Excellence for Coral Reef Studies, James Cook University, Townsville, QLD, Australia

3. Sars Centre, University of Bergen, Bergen, Hordaland, Norway

4. University of Oslo, Oslo, Norway

Sponges are likely to be the earliest branching animal lineage, making them key models in studies aimed at understanding evolutionary history of animal genomes. From the morphological and developmental perspectives, sponges combine features of single-cell eukaryotic organisms and the complex multicellular animals (Eumetazoa). Analysis of the first sequenced sponge genome, the demosponge *Amphimedon queenslandica*, demonstrated a limited number of homologues of genes involved in eumetazoan development, suggesting a gradual assembly of the complex eumetazoan developmental toolkit.

However, sponges are a diverse phylum, composed of four distinct lineages (Demospongiae, Hexactinellida, Calcarea and Homoscleromorpha). Taking advantage of the accessibility of next generation sequencing technologies, we have sequenced genomes of five calcisponges: two calcaroneans (*Sycon ciliatum* and *Leucosolenia complicata*) and three calcineans (*Clathrina lacunosa*, *C. laminoclathrata* and *C. coriacea*), as well as a demosponge only distantly related to *Amphimedon* (*Halisarca dujardini*). For some of these species, we have also generated an extensive collection of transcriptome datasets representing embryonic and postembryonic development and regeneration.

Comparisons of gene content, with emphasis on developmental regulatory genes, demonstrated unexpected complexity and diversity of developmental toolkits among sponges. In particular, calcaronean and calcinean developmental regulatory gene families are significantly more complex than demosponge gene families. Overall, it appears that significant gene loss occurred independently in the calcisponge and demosponge lineages. These gene losses have been followed by gene family expansions occurring independently in the calcaronean and calcinean lineages, and, to a lesser extent, in the lineage leading to *Halisarca*. Strikingly, conservation of gene sequence correlates with similarity of gene expression profiles across the sponge species.

Overall, usage of developmental regulatory genes demonstrated deep conservation of body plan patterning and regeneration mechanisms between sponges and the eumetazoans. At the same time, the detailed picture – numbers of genes within individual families – revealed spectacular diversity among sponge species.

332

Investigating neuropeptide function during cnidarian development

Nagayasu Nakanishi[1], Mark Q Martindale[2]

1. NASA Astrobiology Institute / Whitney Laboratory for Marine Bioscience, University of Florida, St. Augustine, FLORIDA, United States

2. Whitney Laboratory for Marine Bioscience, University of Florida, St. Augustine, Florida, USA

Neuronal signaling controls a variety of behavioral (e.g. motor) and developmental (e.g. metamorphic) processes essential to multicellular life; however, the ancestral neuronal signaling mechanisms and their function remain ill-defined. We are investigating the molecular basis of neural function in the starlet sea anemone *Nematostella vectensis*—a member of an early-branching animal group Cnidaria (e.g. corals, sea anemones and jellyfishes). Here we report that genes encoding Antho-RFamides and GLWamides, the neuropeptides conserved across Cnidaria and Bilateria (e.g. vertebrates, insects and worms), are differentially expressed during *N. vectensis* development. At gastrulation, Antho-RFamide-expressing neurons and GLWamide-expressing neurons begin to emerge in the ectoderm. During planula larval development, both neuronal populations develop in the pharyngeal ectoderm, and GLWamide neurons, but not Antho-RFamide neurons, arise in the endoderm. GLWamide neurons, but not Antho-RFamide neurons, are asymmetrically distributed in the developing pharynx along an axis perpendicular to the oral-aboral axis. At metamorphosis, Antho-RFamide and GLWamide expression largely disappears from the ectoderm of the body column, and both populations of neurons develop in the ectoderm of growing oral tentacles. Antho-RFamide-expressing cells, but not GLWamide neurons, are strongly enriched at the tip of each tentacle. Differential expression of Antho-RFamides and GLWamides in development appears to suggest that these neuropeptides mediate distinct behavioral and/or developmental processes in *N. vectensis*. We are currently taking a CRISPR-Cas9-mediated genome editing approach to determine specific functions of Antho-RFamides and GLWamides during *N. vectensis* development.

333

The generation of somatic cell nuclear transfer embryos with different oocyte mtDNA haplotypes

Kanokwan Srirattana[1, 2], Justin C. St. John[1, 2, 3]

1. Centre for Genetic Diseases, Hudson Institute of Medical Research, 27-31 Wright Street, Clayton, VIC 3168, Australia

2. Department of Molecular and Translational Science, Monash University, 27-31 Wright Street, Clayton, VIC, 3168, Australia

3. Corresponding author , Email: justin.stjohn@hudson.org.au

We set out to determine the effects of mtDNA haplotype on the developmental potential of cattle oocytes using somatic cell nuclear transfer (SCNT) to generate embryos. To ensure that only one mtDNA haplotype was transmitted, mtDNA was depleted from donor fibroblasts prior SCNT by treatment with 2', 3'-dideoxycytidine. SCNT embryos were produced using *Bos taurus* (Holstein) donor fibroblasts and recipient oocytes from *Bos taurus* (Angus) and *Bos indicus* (Sahiwal) cattle. Firstly, donor cells from depleted and non-depleted (control) cells were fused to enucleated *Bos taurus* recipient oocytes. After chemical activation, embryos were *in vitro* cultured for 7 days until they reached the blastocyst stage. We found that blastocysts derived from depleted cells harboured only oocyte mtDNA whilst the co-existence of donor cell and oocyte mtDNA was found in blastocysts derived from non-depleted cells. Therefore, SCNT embryos produced from depleted cells possessed only one mtDNA population. Moreover, mtDNA copy number in blastocysts derived from depleted cells was not significantly different when compared with blastocysts from non-depleted cells. Consequently, depleting mtDNA from donor cells had no negative effect on mtDNA copy number. Secondly, we generated embryos using Holstein depleted cells and the more genetically divergent *Bos indicus* oocytes. Sahiwal cattle were assigned to one of 2 sub-haplotypes (A and B) based on their mtDNA sequences. Blastocyst rates for embryos derived from *Bos indicus* oocytes (3.5%) were significantly lower than those from *Bos taurus* oocytes (35.0%). Interestingly, only embryos derived from group B developed to the blastocyst stage and produced one live offspring, which inherited only oocyte mtDNA. MtDNA copy number for Sahiwal oocytes from group A was significantly lower than for *Bos*

taurus oocytes, which likely accounted for their developmental failure. In conclusion, oocytes with different mtDNA backgrounds exhibited different developmental potentials for the generation of SCNT embryos.

334

Identification of the locus responsible for polydactyly in horses using whole-genome re-sequencing

Brandon D Velie[1], Kim Jäderkvist Fegraeus[1], Ove Wattle[2], Gabriella Lindgren[1]

1. Department of Animal Breeding & Genetics, Swedish University of Agricultural Sciences, Uppsala, UPPSALA, Sweden

2. Department of Clinical Sciences, Swedish University of Agricultural Sciences, Uppsala, Sweden

Polydactyly, a genetic defect that results in an increased number of digits, has historically been observed in a range of species. In cats, polydactyly is considered autosomal dominant and results from point mutations in a control region which effect the expression of the developmental gene Sonic Hedgehog. However, in cattle, polydactyly is believed to be of multifactorial inheritance with no key mutation known at this time. While the heredity of polydactyly has also been explored in chickens and domestic pigs, little is known about the underlying molecular basis for polydactyly in horses.

A miniature Shetland Pony suspected of expressing polydactyly was brought into the local veterinary clinic for examination. Polydactyly was verified through computer tomography and DNA samples from the animal and its family were taken. Preliminary analyses of this family suggest that polydactyly has a recessive mode of inheritance. Through whole-genome re-sequencing of 5 individuals from this miniature Shetland Pony family (two affected and three suspected carriers) and 3 individuals from a family of Warmblood horses (one affected and two suspected carriers) we aim to confirm this mode of inheritance and identify the causative locus. Sequencing was performed by Illumina Hi-seq technology (PE 2+100 bp) with a resulting average depth of 10X for each individual. All reads have been aligned to the genome reference assembly using BWA and additional samples from polydactyly cases are being collected to validate our results. At the meeting, we will present the results of the sequence analysis.

335

Investigating the genetic diversity of the Arabian Oryx from Oman

Qais Abdullah Suliman AlRawahi[1, 2], Mehar Khatkar[1], Helen Senn[3], Jaime Gongora[1]

1. Faculty of Veterinary Science, University of Sydney, Sydney, NSW, Australia

2. Office for Conservation of Environment, Diwan Of Royal Court, Muscat, Oman

3. WildGenes Laboratory, Royal Zoological Society of Scotland, Edinburgh , EH12 6TS, UK

The Arabian oryx (*Oryx leucoryx*) is an endangered ungulate which became extinct in the wild in 1972 but conservation plans have been implemented to save this species. This includes a program in Oman involving breeding and reintroduction into the wild using a few captive populations from zoos and private collections around the world. However, there is limited knowledge of the actual genetic diversity of the breeding populations present in three centers from Oman itself. Understanding population genetic parameters has been identified as a fundamental aspect of better informed management plans. To address this, we are investigating the mitochondrial and nuclear DNA in the breeding populations of the Omani Al Wusta Wildlife Reserve. Here we present preliminary mtDNA control region data from 89 individuals from three groups: originally from Oman (O-Oman); United Arab Emirates (UAE); and populations which are mixed as a result of preliminary breeding between these two groups. We identified 14 variable sites which resulted in 7 haplotypes out of 18 being found which have previously been identified globally in the Arabian Oryx founder population for the Omani conservation initiative from which the animals were sourced. The mixed herd shared all seven of the haplotypes between the Omani and UAE groups. The results indicate a genetic variability slightly below the observed average among the other Arabian Oryx populations. AMOVA analyses indicate a low level of population differentiaton among the three populations (F_{ST}=0.104, $P<0.001$) with restriction of gene flow between the O-Omani and UAE herds. Our findings improve our knowledge of the current status of the genetic diversity of the Oryx and support the future strategy of translocation and genetic management of reintroduced populations.

336

Connectivity among New Zealand coastal marine communities

Vanessa Arranz Martinez[2, 1], Rachel Fewster[3], Shane Lavery[2, 1]

1. Institut of Marine Science, University of Auckland, Auckland, New Zealand

2. School of Biological Sciences, University of Auckland , Auckland, New Zealand

3. Statistics, University of Auckland , Auckland, New Zealand

The levels of genetic connectivity among populations of marine organisms around the New Zealand coast are a key parameter for conservation. They can be crucial in providing adequate management of exploited species, marine protection for threatened species and an understanding of the inter-dependence of marine protected areas. Our aim is to address this issue by taking a community perspective of connectivity rather than simply a species-by-species view. By simultaneously examining the genetic population structure of a diverse range of coastal species we can ascertain if there are common locations where connectivity is either restricted or promoted.

As a first step in this project, we examined connectivity among populations of two New Zealand endemic intertidal species with differing larval dispersal potentials. We used mitochondrial cytochrome oxidase I (COI) and nuclear ITS DNA sequences to investigate the population genetic and phylogeographic structure of the cat´s eye snail (*Lunella smaragdus*) and half crab (*Petrolisthes elongatus*) in order to determine: (1) if they exhibit geographic restrictions to genetic connectivity and if so, where and (2) if they follow one of the common patterns of connectivity previously displayed by other species.

These two species differ significantly in both their dispersal abilities and their patterns of connectivity. Both animals show some degree of genetic discontinuity between the North and South Islands, concordant with other taxonomically and ecologically different species. However they also show some idiosyncratic patterns of gene flow. The data from these species are now being compared with previously published genetic datasets to ascertain if common dispersal barriers are present around New Zealand.

337

Planktonic Relationships of Rottnest Island – A 5-year metabarcoding survey

Tina Berry[1], Megan Coghlan[1], Anthony Richardson[2], Claire Davies[3], Michael Stat[1], Oliver Berry[4], Nicole White[1], Michael Bunce[1]

1. *TrEnD Lab, Department of Environment and Agriculture, Curtin University, Bentley, WA, Australia*
2. *CSIRO Oceans and Atmosphere, Brisbane*
3. *CSIRO Oceans and Atmosphere, Hobart*
4. *CSIRO Oceans and Atmosphere, Perth*

Background and significance: Zooplankton consists of mostly microscopic animals at the mercy of oceanic currents. These animals, together with phytoplankton, form the base of the marine food web. They may be holoplankton, which spend their entire lives as plankton, or meroplankton, which are temporary residents. Meroplankton are typically larval stages of larger animals such as fish or crustaceans. The Integrated Marine Observing System (IMOS) has been collecting plankton samples for morphological analysis for over five years. However taxonomic identification via morphological examination does have its difficulties. The taxa being examined must be in good condition and the identification of the larval stages of both holoplankton and meroplankton are problematic for even a skilled taxonomist. This study uses 55 plankton samples collected from Rottnest Island, in Western Australia over 5 years. The aim of the project is to use molecular methods to identify and map changes in taxa over time, and in response to seasonal and temperature variations.

Basic Methodologies: This research applies environmental metabarcoding to characterise the species composition contained within zooplankton communities. Eight primer sets were used, targeting fish, copepods, crustaceans and other zooplankton. The resulting sequences are to be examined taxonomically and then examined as OTUs for an insight into the genetic diversity of the area.

Major Findings: This poster presents data from several million sequences across the 8 primer sets. From only two sets, 60 taxa of fish and 25 crustacean taxa were identified, many to a species level. Data is presented across the 2011 marine heatwave event to explore biotic patterns associated with this thermal anomaly. This ongoing study will, in the near future, be expanded across the Australian IMOS stations and analysed together with other abiotic data to give a broad picture of the planktonic communities and the value of metabarcoding in marine ecosystem monitoring.

338

Sequence capture arrays for studying immunogenetic variation in non-model species

Amanda Y Chong[1], Leah Clissold[1], Helen E Everett[2], Bryan Charleston[3], Federica Di Palma[1]

1. *The Genome Analysis Centre, Norwich, NORFOLK, United Kingdom*
2. *Virology Department, Animal and Plant Health Agency, Addlestone, SURREY, United Kingdom*
3. *The Pirbright Institute, Pirbright Laboratory, Pirbright, SURREY, United Kingdom*

Diversity at immune loci such as the Major Histocompatibility Region (MHC) is often used as a proxy for determining potential immunological health of a population and the chance that a given population will be able to adapt to the emergence of novel pathogens or disease threats. However, efforts to characterise sequence diversity at these loci have been hampered by a need for prior sequence data. Furthermore, despite efforts to generate genomic sequence data for a wide variety of vertebrate species (e.g. the G10K and B10K efforts), highly complex regions such as the MHC and immunoglobulin loci remain difficult to assemble.

We are combining targeted sequence capture arrays with a combination of long and short read sequencing approaches to improve contiguity over these genomic regions for existing genomes and extending the use of these technologies to characterise the MHC of species without a reference genome. Here we will discuss the use of sequence capture arrays for conservation immunogenetics, and their potential for use in cross-species immunogenetic studies. We will present pilot studies for determination of MHC haplotypes in the well characterised domestic pig model, and discuss the application to endangered mustelids.

339

Marker choice for zooplankton DNA metabarcoding studies

Laurence J Clarke[2, 1], Jason M Beard[3], Kerrie M Swadling[1, 3], Bruce Deagle[2, 1]

1. *Antarctic Climate & Ecosystems Cooperative Research Centre, University of Tasmania, Hobart, Tasmania, Australia*
2. *Australian Antarctic Division, Kingston, Tasmania, Australia*
3. *Institute for Marine and Antarctic Studies, University of Tasmania, Hobart, Tasmania, Australia*

DNA metabarcoding studies of diverse assemblages like zooplankton typically target the conserved nuclear 18S rDNA, sacrificing taxonomic resolution to take advantage of its broad taxonomic coverage. Mitochondrial COI offers better taxonomic resolution and can take advantage of growing DNA barcode databases; however, several studies have demonstrated the taxonomic bias inherent in many COI markers due to the lack of conserved primer-binding sites. PCR-bias is also likely to reduce the utility of COI markers for estimating relative abundance from high-throughput sequencing (HTS) data. To avoid taxonomic bias but retain taxonomic resolution, mitochondrial 12S and 16S rDNA have been proposed as alternative metabarcoding markers. We used HTS to compare the performance of 18S, COI and 16S metabarcoding markers for characterizing zooplankton assemblages. In contrast to previous studies that recommend 18S, we find degenerate COI primers provide the best taxonomic coverage, taxonomic resolution and agreement between morphology- and DNA-based identifications. All markers revealed similar patterns of beta-diversity, although different taxa were identified as the greatest contributors to these patterns for 18S compared to the mitochondrial markers and the morphology-based analysis. Interestingly, all markers showed potential for retrieving semi-quantitative abundance data; however, 18S provided the strongest relationship between biomass and number of HTS reads. Our results show the taxonomic coverage and resolution provided by degenerate COI primers, combined with a comparatively well-developed reference sequence database, make them valuable metabarcoding markers for biodiversity assessment.

340

Archaeological forensics: estimating the provenance of ivory from two 17th and 18th century shipwrecks.

Megan Coghlan[1], Jeremy Green[2], Dáithí Conall Murray[1], Alicia Grealy[1], Carlo Pacioni[4, 3], Michael Bunce[1]

1. *TrEnD Lab, Department of Environment and Agriculture, Curtin University, Bentley, WA, Australia*
2. *Department of Maritime Archaeology, Western Australian Museum, Perth, WA, Australia*
3. *CSIRO, Perth, WA, Australia*

4. Murdoch University, Perth, WA, Australia

Background and Significance: European trade of African elephant ivory was gaining in momentum during the 16th and 17th centuries with several nations setting up trading ports in Africa. One of these was the Netherlands whose Dutch East India Company (VOC) loaded cargo at the Cape of Good Hope, and for a short time, Mozambique. Two vessels, the *Vergulde Draeck* in 1656, and the *Zeewijk* in 1727 both wrecked on the coast of Western Australia, of which both were carrying elephant tusks as unrecorded cargo. To determine whether the tusks were illegal or official trade, this study aims to use mitochondrial DNA analysis to provenance the tusks to an approximate geographical origin using reference data from modern African elephants. A West African origin may indicate illegal trade, whereas an East African origin would indicate official trade.

Basic Methodologies: Twenty-four tusks from the two shipwrecks were sampled. A 316 bp fragment of the mitochondrial control region was sequenced and compared to 280 modern African elephant reference sequences of known provenance using BEAST and Arlequin. Haplotype diversity, and whether geographic origin could be estimated compared to the modern reference data was determined.

Major Findings: Twenty-three shipwreck samples were assigned to an approximate geographical origin based on modern reference data. Control region mtDNA analysis showed that all of the shipwreck samples matched most closely with haplotypes seen in forest elephants. The geographical origins were most likely within Western, West-central, South-central, and North-central Africa. This result indicates that the tusks were illegal trade as none of the samples matched closely with East African haplotypes. New haplotypes were identified that are no longer present in modern elephants. Habitat destruction and extensive hunting of African elephants for their ivory over the last 400 years is the likely cause to this observed decline in haplotype diversity.

341

Counting with DNA: can species proportional biomass be estimated using high-throughput amplicon sequencing?

Bruce E Deagle[1, 2], Laurence J Clarke[1, 2], Austen C Thomas[3, 4], Andrew W Trites[4], Karen J Westwood[1]

1. Australian Antarctic Division, Kingston, Tasmania, Australia

2. Antarctic Climate & Ecosystems Cooperative Research Centre, University of Tasmania, Hobart, Tasmania, Australia

3. Smith-Root Inc., Vancouver, Washington, USA

4. Marine Mammal Research Unit, University of British Columbia, Vancouver, British Columbia, Canada

DNA metabarcoding using high-throughput amplicon sequencing is a powerful tool for characterizing species assemblages. However, using DNA metabarcoding to quantify relative species abundances is currently limited by both biological and technical biases that influence sequence read counts. To assess the problem and explore whether correction factors can improve quantitative estimates, we examined sets of known species mixtures using Illumina amplicon sequencing. In the first study, we characterised mixtures of single-celled marine protists containing either just diatom species or including a broader range of protists from several different phyla. Species-specific biases in recovery of 18S rDNA barcodes compared to cell counts were apparent, and were particularly strong for some diatoms. In mixes covering broad taxonomic groups, corrections based on cell volumes improved quantitative estimates of cell numbers, but this did not apply to diatoms. In the second study, we examined mixtures of tissue from prey species consumed by harbour seals. The goal was to improve diet estimates obtained through DNA-based analysis of faeces by accounting for differential recoverability of prey DNA. Our approach was to sequence mtDNA amplicons recovered from 50/50 mixtures of target fish species and a control species to generate relative correction factors (RCFs). These RCFs measured how under- or over-represented a particular species was relative to the control. In our experiment RCFs remained consistent regardless of the co-occurring species, but RCFs did vary with input ratio (i.e. biases changed when mixes were not 50/50). Still, 50/50 RCFs applied to DNA sequence counts from various mixtures of the target species significantly improved relative abundance estimates of proportions. Overall, our results indicate that correction factors applied to metabarcoding datasets improve estimates of relative species abundance in simple communities.

342

Clarifying the phylogeny and phylogeography of two commonly traded cockatoo species and the development of a wildlife forensic toolbox to identify illegal trade in these species

Kyle M Ewart[1, 2], Rebecca Johnson[1], Nathan Lo[2], Greta Frankham[1], Rob Ogden[3], Leo Joseph[4]

1. Australian Museum Research Institute, Sydney, NSW

2. School of Biological Sciences, The University of Sydney, Sydney, NSW, Australia

3. TRACE Wildlife Forensic Network, Edinburgh, UK

4. Australian National Wildlife Collection, CSIRO National Facilities and Collections, Canberra, ACT, Australia

Wildlife crime continues to be one of the prominent black market activities, and is particularly concerning for parrot species. Australian parrots and cockatoos, in particular the red-tailed black cockatoo (Calyptorhynchus banksii) and the Major Mitchell cockatoo (Lophochroa leadbeateri), are commonly found in the illegal pet trade in places like South East Asia. These two species are both variously listed under State and Federal conservation legislation, and yet little phylogenetic and phylogeographic information is available for these iconic cockatoos endemic to Australia.

The aim of this research is to establish the phylogeny and phylogeography of these two commonly traded cockatoo species and to develop a multifunctional suite of SNPs appropriate for wildlife forensic application based on whole genome sequencing, and single nucleotide polymorphism (SNP) markers. SNPs are the most abundant genetic variants and have immense potential to provide accurate and extensive analyses. This project will involve using SNPs, whole genomes and cutting-edge analytical methods to obtain high-resolution estimates of species boundaries, effective population size, and genetic structure and diversity for both cockatoo species. Further, the current subspecies divisions within each of these two cockatoo species is largely based on morphology across their extensive Australian distribution. An extensive genetic analysis may demonstrate whether the currently recognised subspecies are accurate or whether some subspecies should in fact be elevated to separate species.

A subset of SNPs will be subsequently validated for wildlife forensic purposes, to conduct individualization, phylogeographic location, clutch ID and progeny testing. Being able to identify source populations will allow enforcement and compliance resources to be directed towards these at risk populations, thus mitigating this key threatening process as well as and potentially identifying smuggling route and individuals involved to improve prosecution outcomes.

343

Species detection at extremely low densities: the role of environmental DNA surveys

Elise Furlan[1], Richard Duncan[1], Jonah Yick[2], Chris Wisniewsk[2], Dianne Gleeson[1]

1. University of Canberra, Bruce, ACT, Australia

2. Inland Fisheries Service, New Norfolk, Tasmania, Australia

Environmental DNA (eDNA) surveys involve detecting species from the trace amounts of DNA they shed into their environment. Water, soil or air can be sampled and analysed using species-specific markers to infer the presence of a target species in the environment. eDNA surveys have been shown to be highly sensitive and, in many instances, provide improved detection capabilities at a lower cost than traditional sampling methods (Jerde et al. 2011; Dejean et al. 2012). The potential application of eDNA for detecting particularly low-density populations, such as recent incursions or confirming eradications, is widely promoted (Dejean et al. 2011; Rees et al. 2014). In this study, we investigate the capacity for eDNA to detect an invasive freshwater fish present at extremely low densities. We focus on the species-specific detection of European carp, *Cyprinus carpio*, from Lake Sorell in Tasmania. The large (~51.6 km^2) lake contains the last known population of *C. carpio* in the state and has been the focus of a control program to contain and eradicate the invader. We present the results of our eDNA survey to detect the remaining *C. carpio* individuals and outline some of the factors influencing eDNA detection sensitivities. We indicate the conditions under which eDNA surveys may provide the greatest benefit for detecting species at extremely low densities.

Dejean T, Valentini A, Duparc A, Pellier-Cuit S, Pompanon F, Taberlet P, Miaud C (2011) PLoS ONE, 6:e23398.

Dejean T, Valentini A, Miquel C, Taberlet P, Bellemain E, Miaud C (2012) Journal of Applied Ecology, 49:953-959.

Jerde CL, Mahon AR, Chadderton WL, Lodge DM (2011) Conservation Letters, 4:150-157.

Rees HC, Maddison BC, Middleditch DJ, Patmore JRM, Gough KC (2014) Journal of Applied Ecology, 51:1450-1459.

344

Impact of dams on neutral and adaptive genetic diversity in an Australian freshwater fish

Bertrand B Gauffre[1, 2], Alexandra A Pavlova[2], Rhys R Coleman[3], Minami M Sasaki[2], Joanne J Kearns[4], Jarod Lyon[4], Luciano L Beheregaray[5], Paul P Sunnucks[2]

1. Centre d'études biologiques de Chizé (CNRS - ULR), INRA - French national institute for agronomic research, Beauvoir sur niort, France

2. School of Biological sciences, Monash University, Melbourne, Victoria, Australia

3. Melbourne Water, Melbourne, Victoria, Australia

4. Department of Sustainability and Environment, Melbourne, Victoria, Australia

5. School of Biological sciences, Flinders University, Adelaide, South Australia, Australia

Freshwater ecosystems are particularly vulnerable to habitat loss and fragmentation, given the linear nature of rivers and streams. Many freshwater systems are affected worldwide, most notably by widespread construction of dams to meet increasing water demand for growing human populations. Small, fragmented populations of wildlife are vulnerable to stochastic events and loss of genetic diversity through drift, with an increased risk of inbreeding and the fixation of deleterious alleles. In addition, loss of genetic diversity reduces the adaptive potential of populations to respond to future environmental changes. By comparing microsatellite and hundreds of genome-wide markers, we investigated the impact of dams on neutral and adaptive genetic diversity and genetic structure of an endemic freshwater fish of Australia, *Gadopsis marmoratus*, in the Yarra River system (Victoria, Australia). A space-for-time design was implemented, with several populations sampled upstream and downstream in five streams systems of different size and with barriers of different known ages (from 45 to 120 years old). Results indicated a slight effect of barriers on genetic diversity in the smaller streams only. No effect of barrier age was detected. Indeed, catchment area and the presence of large reservoir upstream of barrier appear to prevent loss of genetic diversity upstream. Isolation-by-distance was more supported than Isolation-by-barrier in the whole study system. Our results, combined with computer simulations, provide general insights into the timing of barrier effects on different classes of molecular marker in freshwater systems.

345

Optimising the sampling design of landscape genetic analyses using resistance surfaces

Bernd Gruber[1], Niko Balkenhol[2]

1. Institute for Applied Ecology, University of Canberra, Canberra, ACT, Australia

2. Wildlife Sciences, University of Goettingen, Goettingen, Germany

To be able to predict the spatial dynamics of species, knowledge on the ability and frequency of individual dispersal events are a fundamental requirement. Recent advances in genetic techniques have enabled ecologists to estimate the effect of landscape features such as roads, rivers and unsuitable habitat on dispersal as the needed resources of such studies are decreasing steadily. Least-cost path analysis is the most common approach to study the effect of landscape features on population structure. The basic idea is simple – various landscape models represented by distance matrices that incorporate the effect of landscape features on dispersal are compared to genetic distance matrices. We present an approach that allows us to predict the likely performance of an intended study design by simply studying the geometric properties of the sampling design. We demonstrate the application of the approach employing a simulation study and a real-world example using a well studied possum meta-population. Finally we present the implementation of the approach within the R-package PopGenReport. We demonsrate that by using approach the sampling design of an intended study can be optimised, which leads to more precise estimates of the effect of landscape features on dispersal, a better abilityto predict the spatial dynamics of a species and ultimately it will allow us to conserve threatened populations more efficiently using landscape-genetic approaches.

346

Exploiting conserved genomic elements to create a universal amplicon resource for use in amniote species with no genome assembly

Jessica F Gurr[1], Claire Wade[1]

1. University of Sydney, Sydney, NSW, Australia

Despite the increasing affordability of next-generation sequencing technologies, there is still no appreciable quantity of sequence data available for many non-model species. The deficiency of genomic resources in wild species has provided the impetus to create a versatile and cost-effective amplicon-based sequencing tool. By exploiting highly conserved genomic elements we have designed a suite of universal nuclear markers that are

transportable across amniote orders. This presents a simple method for next-generation sequencing and comparative genomics research in the absence of a reference genome.

Genomic elements with a high degree of similarity between species, or evolutionarily conserved elements, occur throughout the genomes of vertebrates. We utilised these conserved elements to develop an automated BLAST and PRIMER3-based pipeline for the identification of shared priming regions. After filtering we flagged >49,000 regions suitable for primer placement in eutherians, >13,000 in therians (eutherian + metatherian), and >3,000 in amniotes. Primers were designed to amplify unique in placental, marsupial, monotreme, and bird species. The utility of 30 of these candidate primers was established using a long-range PCR approach, resulting in 57% to 89% PCR success in 7 representative amniote species (*Canis familiaris, Equus caballus, Bos taurus, Mus musculus, Monodelphis domestica, Ornithorhynchus anatinus,* and *Gallus gallus*). We aim to establish the variability of these markers within and between lineages, therefore we will be conducting deep sequencing of a selection of pooled markers across 12 wild and domestic species using the MiSeq platform (Illumina, Inc.).

The capture of orthologous genome-wide sequence data in evolutionarily diverged species presents a valuable resource, eliminating the requirement for laborious species-specific marker development in non-model organisms, and facilitating phylogenetic inference. This is particularly beneficial for species conservation strategies, where economical approaches are crucial. This tool has applications in areas as diverse as systematics, SNP discovery, diversity screening, kinship studies, and individualisation.

347

Comparative transcriptome analyses of eight endangered amphibians in Ryukyu Archipelago: expression patterns and repertoire of immunity related genes

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

1. Hiroshima University, Higashi-Hiroshima, Hiroshima, Japan

2. Hokkaido University, Sapporo, Hokkaido, Japan

3. The University for Advanced Studies, Hayama, Kanawaga, Japan

For animals, immunity is an essential function to survive and a fundamental factor to adapt various environments. Evolution of immune-related genes is thus an important factor for diversification of organisms and its genomes. Because of their evolutionary history as the earliest terrestrial tetrapods and adaptation to both aquatic and semi-aquatic life through metamorphosis, amphibians is one of the key organisms for understanding evolution of genes related to acquired and natural immunity in vertebrates. Moreover, their immunity-related genes have been paid attention in the context of conservation genetics due to emergence of amphibian fungal pathogen. We therefore, sequenced skin and spleen transcriptomes of eight endangered amphibians (7 anurans: *Odorrana splendida*, *O. ishikawae*, *O. amamiensis*, *O. narina*, *O. supranarina*, *Babina subaspera*, and *B. holsti* and a urodele: *Echinotriton andersoni*) in Ryukyu archipelago in Japan and compared expression patterns and repertoire of the genes. As the results, the highly expressed genes were similar between species, and multiple antimicrobial peptides were identified in skin transcriptomes of each species.

348

Patterns of gene expression of natural hybrids and parental species of *Cottus* in different seasons

Elzbieta Iwaszkiewicz[1], Stefan Dennenmoser[1], Arne W Nolte[2]

1. Max Planck Institute for Evolutionary Biology, Plön, Germany

2. Carl von Ossietzky University, Oldenburg, Germany

A hybrid lineage of *Cottus* fish has recently (less than 200 years) invaded vast habitats within the lower River Rhine basin, previously uninhabited by their parental species *Cottus perifretum and Cottus rhenanus*. Despite the lack of geographical and intrinsic barriers to reproduction, invasive hybrid fish appear to be isolated from both parental species. This suggests that adaptation to different habitats drives divergence, although, the nature of this is not known. Previous analyses of differential gene expression identified genes that are likely to contribute to adaptive changes in hybrid *Cottus*. However, it is not clear how tissue specific these patterns were and how variation in the environment affected the results. Here, we used microarray analyses and RNASeq to identify differentially expressed genes in fin and liver tissues collected from controlled laboratory environments and in different seasons (winter/summer). We found that overall gene expression patterns were specific to lineages, with parental species being the most different from each other and hybrids showing intermediate phenotypes. Moreover, certain patterns of gene expression are consistently differentiating the hybrid lineage from both parental species. It is likely that the latter constitute fixed traits of the hybrid lineage. These traits may be most relevant to understand the rapid divergence between parents and hybrids and can give insight on the onset of hybrid speciation in nature.

349

Approaches to restoration of *Phragmites australis* in the salinizing wetlands of the Gippsland Lakes, southeastern Australia

Elizabeth A James[1], Gareth D Holmes[1], Nathan E Hall[2], Anthony R Gendall[2], Paul I Boon[3]

1. Royal Botanic Gardens Victoria, South Yarra, VICTORIA, Australia

2. La Trobe University, Bundoora, Vic, Australia

3. Victoria University, Footscray, Vic, Australia

Phragmites australis is a frequent component of temperate inland and coastal wetlands. Ongoing environmental changes have resulted in the decline of this species in many areas and invasive expansion in others. In south-eastern Australia, the Gippsland Lakes system (GLS) is an extensive coastal waterway system that has experienced increasing salinization since the late 19th century. Increasing salinity is thought to have contributed to the loss of fringing *Phragmites* reed beds leading to increased shoreline erosion. A major goal of restoration in this waterway is to address the effect of salinity by planting a genetically-diverse range of salt-tolerant *Phragmites* lineages. This has prompted an interest in examining the genetic structure of *Phragmites* across the GLS and the variation in salinity tolerance among lineages.

We used microsatellites to investigate population structure of *Phragmites* across the GLS and transcriptomics to identify differential gene expression in response to salinity among *Phragmites* clones. We used an RNA-Seq approach to identify culm-expressed genes

in *Phragmites* associated with exposure to saline water. Six clonal lineages were obtained from areas of low or high salinity across the GLS and grown in pot trials. Paired-samples of the clones from each site were irrigated with either fresh-, or highly saline water. We sequenced transcriptomes from the culms of each of these twelve samples allowing an analysis of differential gene expression.

Among key findings, *Phragmites* formed a single genetic cluster across the GLS consistent with high levels of genetic connectivity facilitated by wind dispersal. Several genes were differentially expressed in clones from highly saline sites when irrigated with saline water relative to clones from low salinity sites. Our data suggest local adaptation of certain clones to salinity and provide scope to develop restoration protocols designed to address the impacts of increasing salinity.

350

Maximizing environmental DNA capture and extraction in the marine environment.

Gert-Jan Jeunen[1], Hamish Spencer[1], Miles Lamare[1], Michael Knapp[1], Neil Gemmell[1]

1. University of Otago, Dunedin, OTAGO, New Zealand

Environmental DNA (eDNA) has recently emerged as a promising tool for monitoring biodiversity in aquatic systems at both the species and ecosystem level. As eDNA studies become commoner, the spectrum of methodological procedures used to capture and extract DNA from environmental samples is broadening. Inconsistent use of different protocols could result in different DNA yields and detection probabilities, hindering the comparability of studies over space and time. All workflows currently in use to analyze aquatic eDNA consist of three steps: (i) sampling a volume of water from the area of interest; (ii) a DNA capture step, concentrating the volume of the sample and; (iii) the extraction and purification of eDNA. We conducted a literature review encompassing all published aquatic eDNA studies and found a significant variation among studies in the methods used at each of these stages. In fact, we found more protocols described than there were papers published on aquatic eDNA. We also investigated the difference in DNA yield for the most commonly used methodological procedures and some recommended protocols from related fields. We included five variations on the filtration capture method, and seven different extraction protocols. Our results show a marked difference between capture methods, with both pore size and filter material used having a significant effect on DNA yield. Water type (eutrophic vs. oligotrophic) had no effect. Our data identify a clear need for the standardization of methods used in aquatic eDNA studies and we suggest an optimal methodological procedure to capture and extract marine environmental DNA.

351

Soaring on the wings of giants - rapid evolution of island gigantism in extinct New Zealand birds of prey

Michael Knapp[1], Stefan Prost[2], James Haile[3], Jessica Thomas[4, 5], Olga Kardailsky[1], Paul Scofield[6], Michael Bunce[7], Tom Gilbert[5]

1. University of Otago, Dunedin, OTAGO, New Zealand

2. University of California, Berkeley, Berkeley

3. University of Oxford, Oxford

4. Bangor University, Bangor

5. University of Copenhagen, Copenhagen

6. Canterbury Museum, Christchurch

7. Curtin University, Perth

Evolution on islands has produced extraordinary forms of adaptations including flightless birds, such as the New Zealand Kakapo, and dog sized elephants, such as the extinct Sicilian Pygmy Elephant. Among the most noticeable environmental adaptations is island gigantism[1]. It describes the phenomenon of island species that are significantly larger than their closest mainland relatives. New Zealand has given rise to a number of such island giants, including birds as well as insects.

Here we use complete mitochondrial genome data to study the evolution of two extinct New Zealand island giants: Haast's eagle, the largest raptor in the world, and Eyles harrier, one of the largest Harriers in the world. Our comparative studies show that both species are closely related to much smaller Australasian species, confirming results from an earlier study on Haast's eagle[2] and providing new evidence of a broader pattern in New Zealand birds of prey. Recent divergence times with their respective closest relatives in both cases suggest a rapid evolution of island gigantism, most likely in adaptation to the unique, mammalian predator free New Zealand ecosystem.

The study provides a foundation for investigations into the functional genomic basis of island gigantism and highlights the potential of state-of-the-art ancient DNA and high throughput sequencing technology for evolutionary studies of extinct organisms.

References:

1. Meiri, S., Cooper, N. and A. Purvis (2008). *The island rule: made to be broken?* Proceedings of the Royal Society B-Biological Sciences. 275:141-8.

1. Bunce, M., Szulkin, M., Lerner, H.R.L., Barnes, I., Shapiro, B., Cooper, A. and R.N. Holdaway (2005). *Ancient DNA provides new insights into the evolutionary history of New Zealand's extinct giant eagle.* PLoS Biology. 3:44-6.

352

The N_e to N ratio is dependent on census population size

Mads Fristrup MFS Schou [1], Volker VL Loeschcke[1], Jesper Smærup JSB Bechsgaard[1], Christian CS Christian Schlötterer[2], Torsten Nygaard TNK Kristensen[3]

1. Department of Bioscience, Aarhus University, Aarhus, Denmark

2. Institut für Populationsgenetik, The University of Vienna , Vienna, Austria

3. Department of Chemistry and Bioscience, Aalborg University, Aalborg, Denmark

Natural populations are finite in size and thus subjected to genetic drift. Genetic drift changes the genetic variance structure and decreases genetic variation within populations. However, selection may counteract or reinforce the effects of genetic drift, and environmental conditions determine not only the strength and direction of selection but also the loci targeted by selection. We performed a large-scale experimental evolution study using 42 replicate populations of *Drosophila melanogaster* subjected to three different rates of genetic drift, by rearing them at population sizes 10 (N10), 50 (N50) or 500 (N500), and two different ecological relevant thermal regimes, one stable across generations and one increasing in minimum, maximum and average temperature. RAD sequencing of pooled samples at five time points across 20 generations revealed a higher rate of loss of genetic diversity in small compared to large populations. Given the hypothesis that effective population size (N_e) is a fraction of the census size (N), due to deviations from Fisher-Wright's idealized population, and constant across different N, we estimated N_e/N. N10 populations had N_e/N ratio close to 1, with this ratio decreasing with increasing population size (N50 and N500), showing significantly slower loss of genetic diversity in small populations than expected. This pattern is potentially due to directional selection in the large populations, strong associative overdominance during the sustained bottleneck of the small populations or a decrease in the variance in reproductive output in the small populations. The loss of genetic diversity was generally not affected by the stressful ramping thermal regime. In conclusion effects of demography on the loss of genetic diversity deviate from the typical expectations of small populations, and interact with environmental stress in shaping the genetic variance structure across chromosomes.

353

Population genetics and relatedness of Afro-Malagasy *Otomops* (Chiroptera: Molossidae) at lineage and colony level

Taryn M Ralph[1], Peter J Taylor[2, 1], Jennifer M Lamb Alport[1]

1. School of Life Sciences, University of KwaZulu-Natal, Westville, KwaZulu-Natal, South Africa

2. Ecology and Resource Management, University of Venda, Thohoyandou, Limpopo, South Africa

We examined population genetics and kinship in 10 colony-based groupings of Afro-Malagasy *Otomops* from three species: near-threatened *O. martiensseni* from eastern South Africa (6 colonies); newly-described *O. harrisoni* from northeast Africa (3 colonies); and *O. madagascariensis* from Madagascar (1 colony, least concern). Our aim was to compare genetic structure based on 6 nuclear microsatellites with that based on mitochondrial cytochrome *b* and D-loop sequence data, at both species and colony level. Further, we aimed to shed light on social structure in *Otomops* by analysing gene flow, migration, relatedness and kinship among and within colonies. Three major lineages were identified in analyses of nuclear and mitochondrial datasets, separated by significant ($p < 0.01$) pairwise F_{ST} values, consistent with *O. martiensseni*, *O. harrisoni* and *O. madagascariensis*. PairwiseF_{ST} and mean relatedness values between colonies from the same species lineage were lower than those between lineages. 70% of individuals sampled were part of either parent/offspring, full-sibling or half-sibling relationships within geographically-based species level lineages, whereas no kinship was observed across lineages. Two parent/offspring dyads were identified within colonies belonging to the *O. martiensseni* lineage and one within the *O. harrisoni* lineage, whereas no parent/offspring dyads were established between colonies. Full-sibling and half-sibling pairs were observed both within and between colonies within their respective lineages. However, most kinship within lineages took the form of half-sibling relationships, reinforcing the suggestion that *Otomops* engages in extra-colony mating. Our results suggest that individuals do not exhibit strict colony faithfulness, and that gene flow is maintained through extra-colony mating. We find little evidence to support the presence of a social system based on either female or male philopatry in Afro-Malagasy *Otomops*.

354

Selection acting on major histocompatibility complex class I genes in Japanese Ranidae frogs

Quintin Lau[1], Takeshi Igawa[2], Shohei Komaki[3], Yoko Satta[1]

1. Department of Evolutionary Studies of Biosystems, Sokendai (The Graduate University for Advanced Studies), Hayama, Kanagawa, Japan

2. Global Career Design Center, Hiroshima University, Higashi-Hiroshima, Hiroshima, Japan

3. Graduate School of International Development and Cooperation, Hiroshima University, Higashi-Hiroshima, Hiroshima, Japan

The major histocompatibility complex (MHC) is an important component of adaptive immunity in all jawed vertebrates, and understanding the evolutionary mechanisms that have shaped these genes in amphibians, one of the earliest terrestrial tetrapods, is important. We characterised MHC class I variation in three common Japanese *Rana* species (*Rana japonica*, *Rana ornativentris*, and *Rana tagoi tagoi*) and identified a total of 60 variants from 21 individuals. We also found evolutionary signatures of gene duplication, recombination and balancing selection (including trans-species polymorphism), all of which drive increased MHC diversity. A unique feature of MHC class I from these three Ranidae species includes low synonymous differences per site (d_S) within species, attributed to a more recent diversification of these sequences or recent gene duplication. The resulting higherd_N /d_S ratio relative to other anurans studied could be related to stronger selection pressure at peptide binding sites. This is one of the first studies to investigate MHC in Japanese amphibians, and permits further exploration of the polygenetic factors associated with resistance to infectious diseases.

355

Gene duplication and relax from selective constraints of *Gcyc* genes create high flower diversity in Didymocarpoideae (Gesneriaceae)

Jing-Yi Lu[1], Kuan-Ting Hsin[2], Chun-Neng Wang[1, 2]

1. Department of Life Science, National Taiwan University, Taipei, Taiwan

2. Institute of ecology and evolutionary biology, National Taiwan University, Taipei, Taiwan

Floral bilateral symmetry is the most important trend for flower shape diversity in 200,000 angiosperm species. Zygomorphy and reversal back to actinomorphy independently evolved multiple times in angiosperm evolution. The genetic basis and the role of selection, however, remain lots of unknown. Yet the developmental mechanism which controls the floral bilateral symmetry relies on *CYCLOIDEA* (*CYC*), a member of TCP gene family. Meanwhile, *Gcyc* genes, *CYC*-like genes in Gesneriaceae, duplicated multiple times in evolutionary history. Members of Gesneriaceae consist of mainly zygomorphic flowers with only a few exceptions, providing a great natural experience opportunity for revealing the evolution patterns of symmetry evolution. Didymocarpoideae has the most number of actinomorphic species among Gesneriad subfamilies. Three duplication events lead to multiple copies of *Gcyc* ranging from two to four in each species. In this study, we detected the selective pressures of duplication *Gcyc* lineages, revealing a relaxation signals after duplication but without lineage-specific patterns. Expression of *Gcyc* genes of two zygomorphic and actinomorphic species indicates a species-specific pattern, which two zygomorphic species utilize different copies to retain

bilateral symmetry. All *Gcyc* genes in actinomorphic species absent in corolla, stamens, and gynoecium which implies ventralization. Based on the relaxation of duplicated *Gcyc* genes and species-specific expression patterns, we suggested a possible "evolutionary flexibility" after *Gcyc* genes duplication which leads to diverse floral symmetry in Didymocarpoideae. This finding, therefore, supports the observation that lineages with floral bilateral symmetry have been able to generate more floral shape diversity.

356

Phylogeny and innate immune gene diversity within the Australian parrot genus *Neophema*

Caitlin Morrison[1], Catherine Grueber[2, 3], Greta Frankham[1], Carolyn Hogg[2, 4], Rebecca Johnson[1]

1. Australian Museum, Sydney, NSW, Australia

2. Faculty of Veterinary Science, University of Sydney, Sydney, NSW

3. San Diego Zoo Global, San Diego, CA, USA

4. Zoo and Aquarium Association Australasia, Mosman, NSW

Neophema is a genus of small Australian parrots otherwise known as the grass parakeets. It contains six species, including several popular aviary birds, as well as one of Australia's most endangered birds, the orange-bellied parrot (*Neophema chrysogaster*, OBP). This project will investigate diversity within *Neophema* through shotgun sequencing and assembly of whole mitochondrial genomes, as well as sequencing of innate immunity genes the toll-like receptors (TLRs). Diversity will be assessed across the genus as well as between individual birds. The project will generate a resolved phylogeny of *Neophema*, which contains two suspected subgenera as well as two sets of subspecies. To date no molecular analysis has included all six species or confirmed the assignment of subspecies; additionally, previous molecular phylogenies have contradicted the current division into two subgenera. Resolution of a complete phylogeny including all six species in the genus will assist in development of management priorities for the orange-bellied parrot. Additionally, a robust phylogeny will assist in the assessment of diversity of toll-like receptor (TLR) genes between *Neophema* species, and in the context of the disease Psittacine beak and feather disease (PBFD). PBFD affects many species of parrots worldwide, and is one of the major threats to the persistence of the wild population of orange-bellied parrots, which currently contains fewer than 70 individuals. Understanding species and individual-level differences in TLRs as well as disease prevalence and resolution may produce information of relevance to disease management, especially for the critically endangered OBP. We are working in collaboration with the Department of Primary Industries, Parks, Water and the Environment (DPIPWE) Tasmania, and the results of this work will be applied directly in management decisions in real time.

357

Molecular Archaeology: Recycling the rubbish

Dáithí Conall Murray[1], Joe Dortch[2], James Haile[1], Michael Bunce[1]

1. TrEnD Lab, Bentley, WA, Australia

2. The University of Western Australia, Perth, WA

Archaeological insights have traditionally been restricted by the preservation of organic material and the presence of diagnostic morphological features. However, in the case of non-woody plants few visible remains are preserved while faunal identification is often skewed towards easy-to-identify taxa. This gap in the ecological record ultimately leads to an imbalance in the analysis of plant and animal material. This issue is particularly acute in Australia where, for instance, non-woody plants had many important and varied uses in Aboriginal societies.

Southwest Western Australia is a biodiversity "hotspot," rich in archaeologically and culturally significant Aboriginal sites (e.g. Devil's Lair). An absence of long pollen records in this region has limited archaeological and palaeo-environmental inferences. Additionally, typical of Australian archaeological sites, more than 80% of the bone fragments previously recovered from past excavations at a number of sites in the area had no diagnostic features at all. With these issues in mind, four archaeological deposits were revisited with the goal of applying ancient sedimentary DNA (*seda*DNA) analysis and a novel bulk-bone metabarcoding (BBM) technique across stratigraphical layers to analyse ecological shifts against the backdrop of episodic human occupation and changing climate.

The approach adopted in this project to studying past and present ecological shifts is a useful adjunct to traditional ecosystem monitoring regimes. This poster highlights key considerations when embarking on HTS projects, offers suggestions for experimental design and applies the methodologies proposed in one of only a handful of difficult to characterise biodiversity hotspots worldwide. The data generated to date suggests that the use of *seda*DNA and BBM techniques will add significantly to our understanding of other archaeological sites across Australia and worldwide.

358

Support for the evolutionary speed hypothesis from intraspecific population genetic data in the non-biting midge *Chironomus riparius*

Ann-Marie Oppold[1, 2], Joao AM Pedrosa[3, 4], Miklos Balint[2], Joao B Diogo[2], Julia Ilkova[5], Joao LT Pestana[3], Markus Pfenninger[1, 2]

1. Molecular Ecology Group, Goethe Universität Frankfurt, Frankfurt Am Main, DEUTSCHLAND, Germany

2. Senckenberg Biodiversity and Climate Research Centre, Frankfurt am Main, Germany

3. Departamento de Biologia and CESAM, Universidade de Aveiro, Aveiro, Portugal

4. Conservation Genetics Group, Senckenberg Research Institute, Gelnhausen, Germany

5. Institute of Biodiversity and Ecosystem research, Bulgarian Academy of Sciences, Sofia, Bulgaria

The evolutionary speed hypothesis (ESH) proposes a causal mechanism for the latitudinal diversity gradient. The central idea of the ESH is that warmer temperatures lead to shorter generation times and increased mutation rates. On an absolute time scale, both should lead to an acceleration of selection and drift.

Based on the ESH, we developed predictions regarding the distribution of intraspecific genetic diversity: Populations of ectothermic species with more generations per year due to warmer ambient temperatures should be more differentiated from each other, accumulate more mutations and show evidence for increased mutation rates compared to populations in colder regions. We used the multivoltine insect species *Chironomus riparius* to test these predictions with COI sequence data and found that populations from warmer regions are indeed significantly more differentiated and have significantly more derived haplotypes than populations from colder regions. We also found a significant correlation of the annual mean temperature with the population mutation parameter θ, that serves as a proxy for the per generation mutation rate under certain

assumptions. This pattern could be corroborated with two nuclear loci. Overall, our results support the ESH and indicate that the thermal regime experienced may be crucially driving the evolution of ectotherms and may thus ultimately govern their speciation rate.

1. Oppold, A. M., Pedrosa, J. A., Balint, M., Diogo, J. B., Ilkova, J., Pestana, J. L., and Pfenninger, M. (2016): Support for the evolutionary speed hypothesis from intraspecific population genetic data in the non-biting midge Chironomus riparius. Proc Biol Sci 283 (1825).

359

Multiple genetic mechanisms contribute to visual sensitivity variation in the Labridae

Genevieve A. C. Phillips[1], Karen L. Carleton[2], Justin Marshall[1]

1. Visual Ecology Lab, The Queensland Brain Institute, The University of Queensland, Brisbane, Queensland, Australia

2. Department of Biology, University of Maryland, College Park, Maryland, USA

Understanding how coral reef fish view their environment is central to learning more about their behavioural ecology. We hypothesise that visual systems have evolved to be well adapted to the feeding ecology of the animal. For example zooplanktivores are likely to have highly acute short-distance vision and ultraviolet (UV) vision for better discrimination of small, colourless animals against a highly UV background. In contrast, herbivorous fishes are likely to have poorer visual acuity, but good colour discrimination at longer wavelengths (brown/green). To test these hypotheses, we investigated the visual ecology of 27 species within the coral reef fish family Labridae: a large, polyphyletic group of predatory wrasses and herbivorous parrotfish. RNA-Seq identified the visual opsin genes expressed within the retina, and their relative expression levels. Six of the seven teleost opsin classes were found in the labrids: *RH1* (rods); *SWS1*; *SWS2B*; *RH2A*; *RH2B*; *LWS* (cones). We then estimated the spectral sensitivities of the labrids (λ_{max}), by identifying amino acid changes at key tuning sites within the visual pigment sequences. In some species, this was compared to microspectrophotometric (MSP) recordings that identify the wavelength of peak absorption of an individual photoreceptor. We also investigated the anatomical properties of the eye, including the distribution of cone photoreceptors in a sub-sample of species. Our results suggest the labrids have extremely variable visual systems. A few species express UV-sensitive opsins (zooplanktivores), while the majority do not. In particular, the Scarines (herbivorous parrotfish) and Cheilines (predatory wrasses) have retinas dominated by long-wavelength-sensitive opsins, with many species having *LWS* gene duplications. Estimates of λ_{max} using opsin sequences broadly agree with MSP recordings. We conclude that there are strong correlations between labrid visual opsin expression, feeding ecology, and anatomical measurements of the eye that support our initial hypotheses.

360

Replicated island invasions of an invasive species (*Rattus rattus).*

Florian Pichlmueller[1], James C Russell[1]

1. University of Auckland, Auckland, AUCKLAND, New Zealand

The New Zealand archipelago is under constant invasion pressure by a great variety of introduced species, with one of the most notorious island pests being the black rat (*Rattus rattus*). Conservation managers particularly need to know if individuals detected after an eradication are survivors or re-invaders, particularly when islands exist in a meta-population with ongoing reinvasion.

Goat Island (<10ha) is an island off the east coast of New Zealand, which is separated from the mainland by a small 50m water channel at low tide. Black rats first invaded the island before the 1970s and despite repeated annual eradication attempts, rats are constantly redetected on the island.

Using molecular tools this study aims to determine which proportion of these individuals are survivors or swimming re-invaders. Tissue samples from the island and the adjacent mainland have been collected during the years 2011-2013, with eradications carried out on the island in 2012&2013. For comparison we included the microsatellite dataset collected in 2005 by Russell et al. We assessed the genetic diversity and population structure using 14 polymorphic microsatellite loci.

We combined Bayesian clustering approaches (STRUCTURE) with conventional F-statistics to assess the gene flow and structure between putative source and sink populations. Contrary to our expectations, the pairwise F_{ST} estimates show significant genetic differentiation between the sampling sites. The clustering results suggest that all sampled individuals belong to three different gene pools (K=3), with two distinct clusters being found for the island population. Interestingly, the island individuals share no ancestry with the distinct mainland population, despite the close proximity to the mainland.

This might be explained by the water barrier clearly restricting gene flow and acting as dispersal barrier. However it is more likely an incumbent advantage of the resident island population resisting new mainland invaders, until its own density is suppressed by annual control programmes.

361

Naive predators and toxic prey: A population genetic analysis of yellow-spotted goannas (*Varanus panoptes*) across Australia

Kimberly Pinch[1], Thomas Madsen[1], Denis O'Meally[2], Beata Ujvari[1]

1. Centre for Integrative Ecology, School of Life and Environmental Sciences, Deakin University, Waurn Ponds, VIC, Australia

2. Faculty of Veterinary Science, University of Sydney, Sydney, NSW, Australia

The introduction of the toxic cane toad (*Rhinella marina*) to central Queensland in 1935, and its subsequent spread across northern Australia, has caused massive declines in native naïve predator numbers such as yellow-spotted goannas (*Varanus panoptes*). Attempting to feed on the toxic toads leads to cardiac arrest and ultimately death of these large carnivorous lizards. A long-term ongoing study (initiated in 2001) of yellow-spotted goannas conducted on the Adelaide River Floodplain in the NT has shown population declines of more than 90% coinciding with arrival of the cane toads in 2005. Surprisingly, population genetic analyses of the TLR4 gene as well as 8 microsatellite markers do not show any loss of genetic diversity of the few surviving yellow-spotted goannas 10 years after the decline. Consequently, the next stage of the study will focus on elucidating whether the observed high genetic diversity in post decline goannas is caused by the extremely high pre-decline genetic diversity and/or by migration of goannas into the study area from adjacent habitats. In addition the study will investigate the evolutionary and molecular mechanisms that allow the yellow-spotted goanna populations persisting in Queensland to co-exist with cane toads for >70 years.

362

Comparative analysis of the thermal stress response in two intertidal *Neritid* snails

Peter J Prentis[1], Shorash Amin[2], Chris Collett[2], Daniel Broszczak[2], Edward K Gilding[3], Ana Pavasovic[2]

1. School of Earth, Environmental and Biological Sciences, Science and Engineering Faculty, Queensland University of Technology, Brisbane, QLD, Australia

2. School of Biomedical Science, Faculty of Health, Queensland University of Technology, Brisbane, QLD, Australia

3. Institute for Molecular Biosciences, University of Queensland, Brisbane, Old, Australia

Nerita melanotragus and *N. albicilla* are widespread intertidal molluscs, distributed across a number of temporally and spatially fluctuating environmental gradients, including abrupt changes in temperature over a tidal cycle. The species differ in their ecology and geographic ranges, as *N. melanotragus*is a mid-littoral species ranging from temperate to subtropical climes, while *N. albicilla* is a low-littoral species found in tropical and subtropical regions. Colonisation of different areas in the intertidal zone means that *N. melanotragus* (mid-littoral) is likely to sustain longer periods of temperature stress than *N. albicilla* (low-littoral). Consequently, these two species present an interesting case to examine differences in their gene expression patterns in response to the same temperature conditions. In this experiment, nine individual samples from each of *N. melanotragus* and *N. albicilla* were randomly allocated into three treatments (14 °C, 22 °C and 31 °C) with three replicates in each treatment. Animals were euthanized, RNA extracted and each individual was sequenced on an Illumina Hiseq. Sequences from each species were assembled and gene expression patterns were determined across the treatment temperatures. The two species had highly divergent patterns of gene expression under treatment conditions. Few differentially expressed genes (~20) were observed in *Nerita albicilla*, and these were dominated by molecular chaperones. More differentially expressed genes (~100) were observed in *N. melanotragus,* but no dominant class of genes was observed. This data suggested that *N. albicilla* had a more significant stress response to temperature and this supports the idea that low-littoral species undergo thermal stress at lower temperatures than mid-littoral species.

364

Evolutionary study of bacteriorhodopsin using genomes and metagenomics

Yoshimoto Saito[1], Katsuhiko Mineta[1], Takashi Gojobori[1]

1. Computational Bioscience Research Center, King Abdullah University of Science and Techonology, Thwal, MAKKAH, Saudi Arabia

Bacteriorhodopsins (BRs) are known as an alternative system of the photosynthesis to produce energy from sunlight in prokaryotes. A lot of BR sequences have been reported so far, however, information regarding their evolutionary history is still less. To analyze their evolution in details, we performed the comprehensive survey of BR genes using about 39,000 prokaryotes complete genomes registered in RefSeq database, and surprisingly, found only about 220 species possessing BR genes in their genomes. BRs are only found in Halobacteria species in archaea. Most of Halobacteria species have two or three BRs in their genomes. On the other hand, BRs were detected in a wide range of species in eu-bacteria while they have only one copy of BR gene in their genomes in general.

Phylogenetic analysis of BRs showed seven large clusters in the resultant tree. Four of them were composed of archaeal BRs while the remaining three composed of eu-bacterial BRs (E1-E3). Each eu-bacterial cluster contains branchings between largely different taxonomic groups. These results suggest that BRs have moved from archaea to eubacteria and then prevail into a wide range of eu-bacteria by lots of horizontal gene transfer [HGT] events. Next we investigated distributions of BRs using metagenomes from various environments. BRs were found in shallow aquatic environments in general. Archaeal BRs were only detected in high salinity and shallow water environments. Although eu-bacterial E1 and E2 BRs were found in most of shallow aquatic environments, E2 BRs were well detected in fresh water environments. E3 BRs were not found from any metagenomes tested, however, the BRs were found in genomes of strains isolated from plant bodies or rhizosphere. These results suggest that BRs might be related to environmental adaptation in prokaryotes.

366

Museum occurrence data predicts genetic diversity in a species-rich clade of desert lizards

Sonal Singhal[1], Daniel Rabosky[1]

1. University of Michigan, Ann Arbor, MI, United States

One of the fundamental characteristics of a species is the amount of genetic variation segregating in its population, which can impact several aspects of a species' biology, including the phenotypic variability it exhibits and its response to selection. Predicting how much variation we should see in a given species is relatively simple under neutral and equilibrium conditions; genetic diversity should scale positively and linearly as a function of census population size and a species' mutation rate. Indeed, many species appear to have levels of variation that correspond with ecological approximations of their census population sizes. That said, despite this being a simple prediction, many studies have found no correlation between aspects of species ecology and geography and their genetic diversity. And, those studies that do find a correlation all show a puzzling pattern: species exhibit a much narrower range of genetic diversity (i.e., a narrower range of effective population sizes) than one would expect given their range of census population sizes. We revisit this puzzle by investigating patterns of genetic diversity across ~80 species in a diverse group of Australian lizards. Applying a phylogenetically-informed approach to genomic data, we find that occcurrence data from museum databases is the best predictor of levels of genetic diversity. However, our proxy for census population size shows a thousand-fold difference across our sampled species, yet, our corresponding estimate for effective population size shows nearly two orders of magnitude less variation across the same species. We explore some of the factors that might help explain this discrepancy.

367

Towards marine ecosystem based management through metabarcoding eDNA

Michael Stat[1], Tina Berry[1], Jeffery John[1], Joseph DiBattista[1], Euan Harvey[1], Michael Bunce[1]

1. Curtin University, Bentley, Perth, WA, Australia

Conservation of the marine ecosystem and effective management of its resources requires the ability to characterise the biodiversity present, and to monitor for changes in composition through time. Auditing environmental DNA (eDNA) present in the ocean may provide a non-destructive and highly sensitive means of detecting changes in marine communities.

DNA isolated from a variety of substrates such as seawater is collectively referred to as eDNA. When combined with next generation sequencing (NGS) and metabarcoding, eDNA can provide a wealth of information for studies of biodiversity, food web dynamics, diet analysis and invasive

species monitoring. Metabarcoding eDNA has become feasible only because it is now possible to simultaneously sequence millions of copies of DNA from complex multi-species environmental samples.

Our research using eDNA isolated from seawater focuses on developing metabarcoding approaches to assess marine biodiversity. By utilising a suite of primer sets targeting different taxa and genes, our analyses enables the characterisation of the entire tree of life – from prokaryotes to mammals. To further understand the sensitivity of metabarcoding eDNA from seawater, we compare the diversity of fish recovered using this approach to the widely adopted BRUV (Baited Remote Underwater Video) methodology.

Lastly, this presentation will highlight the challenges associated with eDNA metabarcoding - false positive/negatives and contamination, detection limits, quantitativeness, DNA movement and longevity of eDNA.

368

Is there genomic support for assisted migration of *Eucalyptus marginata* provenances?

Megan A Supple[1], Jason Bragg[1], Rose Andrew[2], Adrienne Nicotra[1], Linda Broadhurst[3], Margaret Byrne[4], Justin O Borevitz[1]

1. The Australian National University, Acton, ACT, Australia

2. University of New England, Armidale, NSW, Australia

3. CSIRO, Canberra, ACT, Australia

4. Department of Parks and Wildlife, Kensington, WA, Australia

Eucalyptus marginata faces many threats across its native range in southwestern Australia, including potential impacts from climate change. Ongoing reforestation efforts provide an opportunity to establish resilient forests by selecting seeds that are adapted to predicted future climates, as well as current conditions. We use genotyping-by-sequencing (GBS) methods to genotype hundreds of trees from transects across a rainfall gradient. Models of isolation by distance and isolation by environment provide valuable information on the processes that drive patterns of genomic variation across the landscape, with important implications for seed sourcing. Weak isolation by distance indicates that provenancing does not have to be strictly local. Environmental association suggests that assisted migration of seeds over longer distances might be necessary given future climate change scenarios.

369

Sea turtle embryos may be able to adapt to climate change through molecular responses to thermal stress

Jamie Tedeschi[1], Jason Kennington[1], Joseph Tomkins[1], Oliver Berry[2], Scott Whiting[3], Mark Meekan[4]

1. Centre for Evolutionary Biology, University of Western Australia, Crawley, WA, Australia

2. Oceans and Atmosphere Flagship, CSIRO, Floreat, WA, Australia

3. Marine Science Program, Department of Parks and Wildlife, Kensington, WA, Australia

4. Australian Institute of Marine Science, Crawley, WA, Australia

The capacity of species to respond adaptively to warming temperatures will be key to their survival in the Anthropocene. The embryos of egg-laying species such as sea turtles have limited behavioural means for avoiding high nest temperatures, and responses at the physiological level may be critical to coping with predicted global temperature increases. Using the loggerhead sea turtle (*Caretta caretta*) as a model, we used quantitative PCR to characterise variation in the expression response of heat shock genes (*hsp60*, *hsp70*, and *hsp90*; molecular chaperones involved in cellular stress response) to an acute non-lethal heat shock. We show significant variation in gene expression at the clutch and population levels for some, but not all *hsp* genes. Using pedigree information, we estimated heritabilities of the expression response of *hsp* genes to heat shock and demonstrated both maternal and additive genetic effects. This is the first evidence that the heat shock response is heritable in sea turtles and operates at the embryonic stage in any reptile. The presence of heritable variation in the expression of key thermotolerance genes is necessary for sea turtles to adapt at a molecular level to warming incubation environments.

370

An aukward story of evolution and extinction - the disappearance of the Great Auk

Jessica E Thomas[1, 2], James Haile[3], André ER Soares[4], Beth Shapiro[4], Michael Hofreiter[5], John R Stewart[6], Gary R Carvalho[1], M Thomas P Gilbert[2], Michael Knapp[7]

1. Bangor University, Bangor

2. University of Copenhagen, Copenhagen

3. University of Oxford, Oxford

4. University of California, Santa Cruz, Santa Cruz

5. University of Potsdam, Postdam

6. Bournemouth University, Bournemouth

7. University of Otago, Dunedin

The Great Auk, *Pinguinus impennis*, was once found in large numbers across the North Atlantic. However, in June 1844 the last two individuals ever reliably seen were killed, and another species was lost forever. While the Great Auk's extinction was probably caused by overhunting, it is not known whether its' disappearance was facilitated by stress due to increasingly warm and less favourable habitats. Ancient DNA research provides a powerful tool that enables us to learn from the past and to evaluate and investigate factors that influence species' extinction. This evolutionary time travel allows us to explore species' response to environmental change by inferring population dynamics and correlating these with environmental changes. As one of few cold-adapted marine bird species in the Northern Hemisphere to have gone extinct in the Holocene, the Great Auk lends itself well to be a model species to investigate and understand extinction risk from environmental change, and therefore allow us to help endangered species of today. To investigate the relative importance of environmental change and human hunting in the extinction of the Great Auk, we used the latest ancient DNA extraction, capture enrichment and high-throughput sequencing techniques to sequence complete mitochondrial genomes from Auks from across their range. Interestingly, the results show significant genetic diversity and no evidence of phylogeographic structure or a recent bottleneck. This suggests that genetically diverse populations were migrating and mixing throughout their large range, a finding very much in contrast to expectations of a species under threat of extinction. This insight into the genetic history of the Great Auk only adds to the mystery of the bird and its demise and raises the question whether intensive hunting alone could have driven the species to extinction.

371

The role of immunoglobulins in the Tasmanian Devil Facial Tumour Disease

Beata Ujvari[1], Thomas Madsen[1], Katherine Belov[2]

1. Centre for Integrative Ecology, Deakin University, Waurn Ponds, Vic, Australia

2. University of Sydney, Sydney, NSW, Australia

The Tasmanian devil immune system has recently been under scrutiny because of the emergence of a contagious cancer, the Tasmanian Devil Facial Tumour Disease (DFTD), which has decimated devil numbers. In the current study therefore we investigate the role of immunoglobulins in DFTD status, by (i) providing a comprehensive description of the Tasmanian devil immunoglobulin variable regions, and by (ii) investigating how IgG and IgM expression dynamics determines DFTD prevalence. First we show that heavy chain variable (VH) and light chain variable (VL) repertoires are similar to those described in other marsupial taxa: VL diversity is high, but VH diversity is restricted and belongs only to clan III. Phylogenetic analyses revealing highly complex and ancient devil VL gene segments, with some lineages predating the separation of marsupials and eutherians. We suggest that, similar to other studied marsupial species, the complex VL segment repertoire compensates for the limited VH diversity in Tasmanian devils.

Second we show that immunoglobulin M (IgM) and G (IgG) expression levels as well as IgM/IgG ratios decrease with increasing devil age. Neither age, sex, IgM nor IgG expression levels affect devil DFTD status in our analyses. However, devils with increased IgM relative to IgG expression levels show significantly lower DFTD prevalence. Our results therefore suggest that IgM/IgG ratios may play an important role in determining devil susceptibility to DFTD. Our study further support the fact that the Tasmanian devil's immune system is competent and that their susceptibility to Devil Facial Tumour Disease is the result of immune evasion by the tumour and not because of inadequacies in the Tasmanian devil immune system.

372

Metabarcoding of Bacterial Pathogens in a Rodent Pest: Which Organ?

Petra Villette[1], Patrick Giraudoux[1], Jean-François Cosson[2], Eve Afonso[1]

1. UMR 6249 Chrono-Environnement, Université Bourgogne Franche-Comté, Besancon, Franche-Comté, France

2. Centre de Biologie pour la Gestion des Populations (CBGP) France , French National Institute for Agricultural Research , Montpellier, France

Identifying bacterial pathogens in rodents is of critical concern in both public health and wildlife ecology; zoonotic bacteria pose significant human health risks globally, and have the potential to drive population dynamics in their rodent hosts. High-throughput sequencing technologies now allow for rapid and cost-effective surveys of multiple pathogens in rodent populations, but it is currently unclear if the host organ chosen for screening influences the number and identity of bacteria detected. We used 16s rRNA metabarcoding to identify bacterial pathogens in the heart, liver, lungs, kidneys and spleen of 13 water voles (*Arvicola terrestris*) collected from two populations in Franche-Comté, France to determine if bacterial assemblages within organs are similar, if all five organs are necessary to detect all of the bacteria present in an individual animal, and if differences between the two host population's bacterial assemblages can be detected by each organ. We detected 25 bacteria representing 17 genera; average bacterial richness for each organ ranged from 1.5 ± 0.4 (mean ± standard error) to 2.5 ± 0.4 bacteria/organ and did not differ significantly between organs (Kruskal-Wallis test, $\chi_{24} = 4.70$, $p = 0.300$). The average bacterial richness when organ assemblages were pooled within animals was 5.4 ± 0.7 bacteria/animal, and rarefaction analysis indicates that all five organs must be included to obtain this. Organ type does not, however, influence bacterial assemblage composition in a systematic or predictable way (PERMANOVA, 999 permutations, pseudo-$F_{4,51}$=1.34, p=0.12). Ordination and PERMANOVA analysis indicates that differences in pooled, liver, and lung assemblages map to host populations, but heart, kidney and spleen assemblages do not differ between host populations. Our results demonstrate that the number of organs sampled influences the power to detect bacterial pathogens and host-population trends in bacterial assemblage compostion. These results can inform sampling decisions in public health and wildlife ecology.

373

The geographical distribution of grey wolves in China

Ya-Ping Ma[2, 1], Lu Wang[2], Peter Savolainen[3], Ya-Ping Zhang[2, 1], Guo-Dong Wang[1]

1. State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, China

2. State Key Laboratory for Conservation and Utilization of Bio-resources in Yunnan, Yunnan University, Kunming, China

3. Department of Gene Technology, KTH-Royal Institute of Technology, Science for Life Laboratory, Solna, Sweden

The grey wolf, *Canis lupus*, is one of the most widely distributed terrestrial mammal. It has generally been reported to be distributed throughout Europe, Asia and North America, from 15°N latitude in North America and 12°N latitude in India to beyond the Arctic Circle, but has been considered to be absent from Africa and the southern part of East Asia. However, recent articles report that the Egyptian jackal (*Canis aureus lupaster*, Hemprich and Ehrenberg 1833) should be reclassified as the African wolf *Canis lupus lupaster*. Similarly, there exist misconceptions in the western literature about the distribution of wolf in China. In the present study, we surveyed rich literature in Chinese concerning past and present distribution of wolves to synthesize a comprehensive picture of wolf distribution in China, and to make this significant information available to an international audience. The results show that the wolf is represented in 26/30 provinces across continental China between 1981 and the present. Wolves are still frequently observed all across China, as exemplified by sightings in 2011 reported from the south Chinese province Yunnan, and in 2000 reported from the two southernmost provinces Guangdong and Guangxi. We also made a survey of wolf skins in three major Zoological Museums. We found, e.g., wolf skins obtained as far south as Zhejiang and Fujian in 1974, and southern Yunnan in 1985. Numbers of wolves in China have decreased during the last 50 years, and large populations now remain only in the northwestern and northeastern parts, and in Inner Mongolia and Tibet. However, also in these regions numbers are relatively small with, e.g., 2000 wolves in Inner Mongolia reported in the 1990s. Fortunately, we have here shown that wolves are still present across all parts of the Chinese mainland, including the most southern parts of China.

374

Black-cockatoos and molecular ecology approaches for conservation

Nicole E White[1], Christine Groom[2], Megan Coghlan[1], Michael Bunce[1]

1. Curtin University, Bentley, WESTERN AUSTRALIA, Australia

2. University of Western Australia, Perth, Western Australia, Australia

The conservation of highly mobile species, such as black-cockatoos (*Calyptorhynchus* spp.), presents challenges for the agencies responsible for their management. Very few ecosystems and/or habitat remain that haven't been affected by anthropogenic activity. More often than not, this results in species decline, but there are exceptions. Since European settlement, a large portion of the southwest corner of Western Australia has been cleared and modified for agriculture (~95,800 km^2). Here we present the molecular toolkit used to investigate; (1) the relationship of cockatoos; (2) fragmented habitat and the resulting consequences of genetic diversity for the endangered white-tailed black-cockatoo; (3) deterrents for illegal trade and poaching; (4) preservation and augmentation of significant nesting habitat; and (5) adaption of dietary choices in today's modified ecosystems. When the genetic data is combined, it provides another approach to understand species and ecosystem management in an ever-changing landscape.

375

Transcriptome sequencing of *Crucihimalaya himalaica* (Brassicaceae) reveals how *Arabidopsis* close relative adapt to the Qinghai-Tibet Plateau

Ticao Zhang[1], Qin Qiao[2], Qia Wang[1]

1. Kunming Institute of Botany, Chinese Academy of Sciences, Panlong, YUNNAN, China

2. School of Agriculture, Yunnan University, Kunming, Yunnan, China

The extreme environment of the Qinghai-Tibet Plateau (QTP) provides an ideal natural laboratory for studies on adaptive evolution. Few genome/transcriptome-based studies have been conducted on how plants adapt to the environments of the QTP compared with numerous studies on vertebrates. Crucihimalaya himalaica is a close relative of Arabidopsis with a typical QTP distribution, and is hoped to be a new model system to study speciation and ecological adaptation in extreme environments. In this study, we generated a de novo transcriptome sequence of C. himalaica, with a total of 49,438 unigenes. In a comparison with five related species, we identified 10,487 orthogroups shared by all six species and 4,286 orthogroups that contained a putative single-copy gene. Further analysis identified 487 extremely significantly positively selected genes (PSGs) in the C. himalaica transcriptome. Theses PSGs were enriched in functions related to specific adaptation traits, such as responses to radiation, DNA repair, nitrogen metabolism, and membrane stabilization. These functions are probably responsible for the adaptation of C. himalaica to the high radiation, depleted soil and low temperature environments on the QTP. Our findings indicate that C. himalaica has evolved complex strategies to adapt to the extreme environments on the QTP and provide novel insights into the genetic mechanisms of highland adaptation in plants.

376

The genetic architecture of cold tolerance plasticity in the *Drosophila Genetic Reference Panel*

Michael Ørsted[1], Palle Duun Rohde[2], Torsten Nygaard Kristensen[1]

1. Section of Biology and Environmental Science, Department of Chemistry and Bioscience, Aalborg University, Aalborg, Denmark

2. Center for Quantitative Genetics and Genomics, Department of Molecular Biology and Genetics, Aarhus University, Tjele, Denmark

Developments in molecular biology and bioinformatics have enabled researchers to associate individual genetic markers such as single nucleotide polymorphisms, SNPs, with phenotypic variation. Traditional genome-wide-association-studies (GWAS) enables identification of individual genetic markers associated with variation in a phenotypic trait, but for complex traits such as responses to environmental stress, this single variant method is highly underpowered. Instead, statistical methods utilizing the collective action of multiple genetic markers within a biological pathway might have more power to detect biological causality. We used 166 inbred lines of the *Drosophila* Genetic Reference Panel (DGRP) to investigate genotype-by-environment-interactions (GxE) across five developmental temperatures ranging from 17 to 29°C. We assessed cold tolerance of adult flies from each developmental temperature using the measure critical thermal minimum (CT_{min}). Variation in the plasticity of cold tolerance in the DGRP were then determined as the norm of reaction of individual lines across the five developmental temperatures. By grouping genetic markers (approximately 2 million) by biological pathways we aim at identifying sets of genetic markers associated with the norm of reaction to detect genetic fingerprints of plastic and canalized genotypes. Additionally, we associate genetic variation with variation in basal cold tolerance in flies from each developmental temperature to investigate whether the same genetic architecture can explain variation in cold tolerance across developmental temperatures. Cold tolerance differed significantly among the developmental temperatures with CT_{min} spanning from 1.5 to 11.4°C in flies developing at 17 and 29°C, respectively. Large variation in CT_{min} between the lines within each developmental temperature indicates a strong genetic component of cold tolerance. We found significant variation in the slope of reaction curves, thus genetic variation for plasticity of cold tolerance within the DGRP. Preliminary association analyses have identified several regions of the genome that explain differences in cold tolerance within temperatures and in the plasticity of cold tolerance.

377

DNA Sequence Evolution Simulation and Phylogenetic Reconstruction Using Pen and Paper

Nicole T Perna[1], Jeremy D Glasner[1]

1. University of Wisconsin - Madison, Madison, WI, United States

Tree-thinking is an approach to emphasizing evolution as a fundamental unifying concept in biology backed by a growing body of educational research. Many curricular resources related to phylogenetics are available for secondary school and undergraduate biology instruction. Most focus on interpretation of trees, or phylogenetic reconstruction using morphological or molecular characters. Few active-learning exercises link phylogeny to the process of mutation, and fewer yet address the role of data simulation in evolutionary biology. The exercise presented here fills this gap, while not requiring computational resources sadly still limited in many U.S. schools.

We developed a 4-part pen and paper DNA sequence evolution simulation and phylogenetic reconstruction lesson. Part 1 involves instruction in tree interpretation and includes a short formative assessment tool. In Part 2, students are given a 5-taxon tree topology with nodes calibrated in millions of years, and a 10 bp ancestral DNA sequence. They use two polyhedral dice (one 10-sided, one 4-sided) to simulate sequences at each descendant node under a model where each lineage experienced one fixed mutation per million years. In part 3, students are asked to reconstruct the tree using simulated sequences without any algorithmic instruction. Part 4 involves instruction about distance matrices and application of UPGMA.

An instructor's guide and student handouts are freely available (Creative Commons License). These include teaching tips and readings to engage students to consider how modeling and simulation can be used to test hypotheses about evolutionary processes and examine the performance of analysis methods. The modularity of the lesson accommodates a variety of class formats and educational levels. We field tested the exercise in undergraduate courses and professional development workshops aimed at secondary school teachers in Wisconsin, USA. It is amenable to implementation elsewhere.

378

Re-imagining the Hardy-Weinberg Law

Von Bing Yap[1]

1. National University of Singapore, Singapore, SINGAPORE

The Hardy-Weinberg Law is about allelic independence, and may be stated as follows: In a closed random-mating population without selection, a progeny's alleles are independently sampled from the respective parental gene pools. Population genetics textbook proofs of this fundamental result have hardly changed since its discovery in 1908. The standard argument, whcih applies Mendel's First Law on mating types, presents significant algebraic challenge for multiple alleles; even for the biallelic case, it takes quite a bit of concentration to follow. It is observed that the Law is a sort of counter-example to Simpson's paradox. This key observation inspires a new simple argument that retains its elegance for any number of alleles. Thus, it should be incorporated in standard textbooks. Furthermore, it is found that random mating together with certain fertility selection can result in allelic independence, which may have important implication on the interpretation of real data: allelic independence may not mean no selection. Also obtained is a mathematical characterisation of fertility selection coefficients that work a given population to yield allelic independence.

379

Molecular characterization of MHC class IIB genes of sympatric Neotropical cichlids in crater lakes of Nicaragua

Marta Barluenga[1], **Melinda J Hofmann**[1], **Christophe Eizaguirre, Seraina Bracamonte**

1. Museo Nacional de Ciencias Naturales, CSIC, Madrid, Spain

The Major Histocompatibility Complex (MHC) is a key component of the adaptive immune system of all vertebrates and consists of the most polymorphic genes known to date. However, due to this complexity, MHC has not yet been characterized in many species including any Neotropical cichlid fish. Neotropical crater lake cichlids are ideal models to study evolutionary processes as they display one of the most convincing examples of sympatric and repeated parallel radiation events within and among isolated crater lakes. Here we characterize the genes of MHC class II beta chain of the Midas (*Amphilophus)* cichlid species complex including *Amphilophus citrinellus, A. labiatus, A. xiloaensis* and *A. amarillo* from five lakes in Nicaragua. We designed 19 new specific primers anchored in a stepwise fashion in order to detect all alleles present. We obtained 856 genomic DNA (gDNA) sequences from thirteen individuals and 756 additional sequences from complementary DNA (cDNA) of seven of those individuals. We identified 69 distinct alleles with up to 24 alleles per individual. We also found considerable intron length variation and mismatches of alleles detected in cDNA and gDNA.. Lastly, we created a model of protein structure homology for each allele and identified their key structural components. Overall, *Amphilophus* cichlids have one of the most diverse repertoires of MHC class II B genes known which could serve as a powerful tool to elucidate the process of divergent radiations, colonization and speciation in sympatry

380

Birth of a new gene: the problem of self-tolerance versus autoimmune reaction

Cemalettin Bekpen[1], **Diethard Tautz**[1]

1. Max Planck Institute of Evoltionary Biology, Ploen, SCHLESWIG-HOLSTEIN, Germany

In vertebrates, the adaptive Immune system has an extraordinary potential for generating receptors that sense and neutralize any foreign antigens entering the body. Efficient recognition of the foreign antigens depends on the regulation in Thymus tissue where T cells are selected positively and negatively. Promiscuous gene expression of Tissue specific antigens(TSA), required for negative selection, is introduced to T cells by mature mTEC cells mostly through the control of *Aire* gene function in Thymus. Any failure within the function of the *Aire* gene results in the loss of sense and non-sense separation and thereby autoimmunity due to improper representation of TSA in Thymus.

De-novo evolved genes usually bring novel expression pattern with newly evolved genetic content to specific tissues and therefore new protein products within certain cell types. These novel protein products or peptides, which are presented by the major histocompatibility complex on the cell surface, have to be introduced to the immune system to avoid autoimmune reaction. Otherwise any tissue having a newly emerged peptide, represented on their cell surface, will be considered as foreign antigen and tissue/cells that are having such peptide will be attacked and destroyed by the immune system. Therefore, we propose that *de-novo* evolved protein-coding genes should also be expressed within the Thymus to generate self-tolerance.

Based on the analysis from Thymus RNA seq data along with 9 other tissues within the phylogeny of Mus genus, we provide evidence that Thymus plays very critical role in the evolution of metazoan by controlling birth of a new gene in a specific tissue. Our results indicate a primary role for the Thymus controlling expression of all protein-coding genes within both annotated Genic and Non-Genic regions. The mechanisms may also be relevant for hybrid incompatibility effects between species and sub-species and thus also of relevance for speciation processes.

381

A phylogenomic study of speciation dynamics in a large radiation of Australian skinks

Jason Bragg[1, 2], **Sally Potter**[1], **Mozes Blom**[1], **Ana Silva**[1], **Craig Moritz**[1]

1. CBA and RSB, ANU, Acton, ACT, Australia

2. Royal Botanic Garden, Sydney, NSW, Australia

With the increasing availability of large-scale phylogenies and data on distributions of species, there is intense interest in discovering what factors and processes influence speciation-extinction dynamics over time. This endeavor is especially interesting for the highly endemic Australian biota, in

which major radiations have accompanied interaction with the Asian fauna from the early Miocene, and dramatic changes in Australian biomes from the mid Miocene to the present. We will present initial results from phylogenomic analyses of diversification for the Eugongylus group skinks of Australia. This species-rich group covers the entire continent, with high diversity in eastern forests and less so across the extensive arid zone. Our macro-evolutionary examination of this clade is augmented by detailed phylogeographic analyses of selected lineages, which point to a high rate of phenotypically cryptic divergence in the Eugongylus group. This poses the question of whether the usual starting point in macroevolutionary analyses – taxonomically recognized, morphologically-differentiated, species – leads to a biased perception of speciation processes and dynamics.

382

Gene regulatory networks underlying the diversity of cichlid fish vision

Karen L Carleton[1], Sri Pratima Nandamuri[1], Brendan Joyce[1], Matt Conte[1], Thomas D Kocher[1]

1. University of Maryland, College Park, MD, United States

Cichlid fishes are among the most diverse and speciose vertebrate clades. Their bright mating colors are matched by equally diverse visual sensitivities. Species vary in the wavelengths of light they detect and that variation can drive cichlid speciation. Cichlid visual sensitivities arise from the expression of different subsets of seven cone opsin genes. We are using genomic approaches to characterize the gene regulatory networks controlling opsin expression in Lake Malawi cichlids. At least 6 QTL affecting opsin gene expression have been identified in multiple F2 crosses between species with varying expression. One of these QTL is a putative cis regulatory locus located in or near the SWS1 opsin gene. The other QTL are not near opsin genes, and represent trans-acting loci. Using network analysis of retinal transcriptomes and association mapping across species, we have identified transcription factors in several of these QTL that control expression of the SWS2A (blue sensitive) and LWS (red sensitive) opsins. Insertions and deletions in the cis-regulatory regions of these transcription factors are highly correlated with opsin expression. The transcription factors are on different chromosomes so that opsin expression is not correlated in F2 progeny. However, expression is correlated among species, suggesting that particular combinations of opsins might be optimal and so selected. Further, these indels are sorting through the entire Malawi flock and are likely ancestral polymorphisms within the lake. These causative loci suggest a role for ancestral insertion/deletion polymorphisms in creating cichlid diversity.

383

Compositionally heterogeneous codon evolution, concatenation, and cytonuclear discordance significantly alter phylogenetic reconstructions in rapidly diversifying Drosophila

Renee Catullo[1, 2], Lars Jermiin[2], John Oakeshott[2]

1. Western Sydney University, Penrith, NSW, Australia

2. Land and Water Flagship, CSIRO, Canberra, ACT, Australia

Recently and rapidly diversifying taxa remain a significant challenge to phylogenetic reconstruction. In such groups, genetic data from the nuclear genome generally contains little phylogenetic signal, often creating reliance on more rapidly evolving mitochondrial data and/or concatenated approaches for phylogenetic reconstructions. For a variety of reasons, the reliance on mitochondrial data and simplified phylogenetic analyses may produce deeply misleading results, although the extent to which phylogenetic reconstructions are affected is rarely tested. We used data on the *Drosophila melanogaster*clade to test the influence of concatenation methods, cytonuclear discordance, and assumptions about nucleotide evolution on phylogenetic reconstructions. Our surprising results demonstrate how easy it can be to generate a well-supported phylogenetic tree that may significantly differ from a species tree in a rapidly diversifying taxa. We demonstrate that even when using large volumes of genetic data, it must still be carefully curated prior to use in phylogenetic analysis.

384

population differentiation, adaptability and heritability in the common sunskink *Lampropholis coggeri*

Felipe de Mello Martins[1], John Llewelyn[2], Ben Phillips[3], Loeske Kruuk[1], Craig Moritz[1]

1. Australian National University, Lyneham/Canberra, ACT, Australia

2. Centre for Tropical Biodiversity and Climate Change, James Cook University, Townsville, Queensland, Australia

3. School of BioSciences, University of Melbourne, Melbourne, Victoria, Australia

Small, isolated populations at the edge of a species' range (peripheral isolates) are potential hotbeds of adaptive diversity under climate change, and an interesting experimental systems in which to answer fundamental questions in ecology and evolution. We aim at exploring the genetic basis of potentially adaptive traits under a climate change scenario using as model the common sunskink *Lampropholis coggeri* as this animal is a well-developed laboratory model, has a well-studied phylogeographic history and ectotherms are considered to be particularly vulnerable to climate change due to their narrow tolerance ranges to environmental temperatures. Using DArT technology, we generated nearly 18000 SNPs randomly distributed in the genome in a captive breeding population previously genotyped for morphological and physiological traits. We used this data to assign paternity to individuals. We also calculated heritability of measured traits using mixed models. Our analyses suggest that a simple model without maternal is a better fit to the data. Many important physiological traits related to thermal tolerance and performance have significant differences between populations that are mantained after acclimation and are also observed on F1. Many of these traits have low heritability; these results matches Fisher's theorem that traits under strong selection have low narrow-sense heritability because of low variance in phenotypes. We conclude that the two populations studied show important phenotypic differences that are not only heritable but are under strong selection.

385

The role of microRNAs in the repeated parallel diversification of lineages of Midas cichlid fish from Nicaragua

Paolo Franchini[1], Peiwen Xiong[1], Carmelo Fruciano[1, 2], Axel Meyer[1]

1. Department of Biology, University of Konstanz, Konstanz, Germany

2. School of Earth, Environmental & Biological Sciences, Queensland University of Technology, Brisbane, Australia

Cichlid fishes are an ideal model system for studying diversification because they provide textbook examples of extremely rapid evolution and speciation through adaptive radiation. Although gene regulation has been widely recognized to be an important mechanism that links diversification in gene function to speciation, so far its role in cichlid speciation has received little attention. We investigated the potential importance of miRNA regulation in the diversification of six cichlid species of the Midas cichlid adaptive radiation (*Amphilophus* spp.) from Nicaraguan lakes. Using several genomic resources we discovered 236 Midas miRNA genes that were then used to predict miRNA target sites in 8,232 Midas 3' UTRs. Using population genomic calculations of SNP diversity we found miRNA genes to be more conserved than protein coding genes. In contrast to what has been observed in other cichlids, but similar to what is typically found in other organisms, we observed genomic signatures of purifying selection on miRNA targets by comparing these sites with less conserved non-target portions of the 3' UTRs. Interestingly, in one species pair that putatively speciated sympatrically in crater Lake Apoyo, we found a different pattern of relaxed purifying selection and high genetic divergence at miRNA targets. Our results suggest that sequence evolution at miRNA binding sites might be an important mechanism that contributes to the rapid phenotypic evolution of the Midas cichlid adaptive radiation from Nicaragua, but also in cichlid fishes more generally.

386

The genetic basis of an adaptive key trait – radula genes in the radiation of *Tylomelania*

Leon Hilgers[1], Michael Hofreiter[2], Stefanie Hartmann[2], Thomas von Rintelen[1]

1. Museum für Naturkunde Berlin - Leibniz Institute for Evolution and Biodiversity Science, Berlin, Germany

2. Institute of Biochemistry and Biology, Faculty of Mathematics and Natural Sciences, University of Potsdam, Potsdam, Brandenburg, Germany

Specializations of the feeding apparatus play a crucial role in many adaptive radiations. Examples of such adaptive key traits include the beak of Darwin's finches and pharyngeal as well as oral jaws of the East African cichlids. Despite intensive research on the genomics of adaptation and speciation, our understanding of the genetic basis of these processes and involved phenotypic traits is still in its infancy. *Tylomelania* is a genus of viviparous freshwater snails endemic to Sulawesi (Indonesia) that comprises riverine and lacustrine clades, the latter of which have undergone radiations in the ancient lakes of the island. In contrast to the riverine species, lacustrine taxa exhibit unparalleled habitat-correlated radula (rasping tongue) diversity. To investigate the genetic basis of this putatively adaptive trait, we generated morph-wise transcriptomes of the mantle and radula-forming tissue of *Tylomelania sarasinorum*, a species that exhibits a striking substrate-correlated radula polymorphism. Here we present the first radula-forming tissue transcriptome and compare sequence and expression information from both ecomorphs to investigate the genetic basis of radula shape and formation. The assembled transcriptomes give first insights into the genetic basis of radula formation and add to the existing information on molluscan shell biomineralization. Differential expression analysis illuminated tissue specific patterns, and, combined with SNP analyses, generated a list of candidate genes that likely contribute to radula diversity. This study is the first step towards uncovering the genetic basis of radula diversification in *Tylomelania* which will ultimately add to our understanding of the genetic underpinnings of adaptive traits.

387

Whole genome sequencing studies of speciation and selection in the Lake Malawi cichlid radiation

Milan Malinsky[1, 2], Hannes Svardal[2], Alexandra Tyers[3], Martin Genner[4], Eric Miska[1, 2], George Turner[3], Richard Durbin[2]

1. University of Cambridge, Cambridge, United Kingdom

2. Wellcome Trust Sanger Institute, Cambridge, United Kingdom

3. Bangor University, Bangor, Wales, United Kingdom

4. University of Bristol, Bristol, United Kingdom

The thousands of haplochromine cichlid fish species in the great lakes of the African rift valley constitute some of the most dramatic vertebrate evolutionary radiations. This presentation will cover analysis of >300 whole-genome sequences spanning three evolutionary timescales: 1) unpublished results covering over 85 species and all the proposed major evolutionary lineages of Lake Malawi; 2) a phylogeographic study connecting Lake Malawi to other East African lake and riverine haplochromines; 3) speciation in Massoko - a small crater lake in the Malawi catchment (Malinsky *et al*., 2015).

Key results include the finding that the maximum genome wide divergence of Lake Malawi cichlids is of a similar level as for species from Victoria (surprising given that the Victoria radiation is thought to be at least 10x younger). We also see strong signatures of gene-flow between the great lakes, likely mediated by riverine fish. To assist time calibration, we estimated the mutation rate from parents-child trios (preliminary estimate: ~$0.6x10e^{-8}$ per bp per generation). In Lake Malawi, the average genome phylogeny shows differences to standard taxonomy, with instances of repeated phenotypic specialization. A large amount of genetic variation is shared across species, with pairwise Fst between 5% and 65% (diversity within species 0.05-0.1%, divergence between species 0.1-0.3%). Because of the shared variation, local phylogeny varies across the genome. In part this is expected due to incomplete lineage sorting, but we also see evidence for gene-flow between separate branches of the Malawi phylogeny. In coding sequences, we see strong evidence of selection in the visual system, including genes not previously studied in cichlids. A detailed study into a specific case of speciation and selection has been carried out for Lake Massoko. To obtain a similar level of detail for pairs of Malawi species, we recently collected 1400 more Lake Malawi samples from ~250 species.

1. Malinsky, M. et al. Genomic islands of speciation separate cichlid ecomorphs in an East African crater lake. Science 350, 1493–1498 (2015).

388

Species boundaries among *Heliconius* butterflies reflect the genetic architecture of speciation

Simon H Martin[1], Dorothea Lindtke[2], John Davey[1], Chris D Jiggins[1]

1. University of Cambridge, Cambridge, CAMBRIDGESHIRE, United Kingdom

2. University of Sheffield, Sheffield, United Kingdom

We still lack a detailed understanding of the genetic mechanisms that allow species to coexist and hybridize without collapsing. Theory predicts that genetic incompatibilities that reduce fitness of hybrids and recombinants may form barriers to gene flow, particularly in genomic regions of reduced recombination. This has proved difficult to test empirically. Conventional methods for studying the landscape of gene flow across the genome are prone to biases, making them difficult to interpret. We addressed this issue by using novel, and less biased, approaches to study relatedness across the genome between hybridizing species of Heliconius butterflies. We compared two pairs of species using multiple resequenced genomes from multiple replicated regions of sympatry. We found that relatedness between species fluctuates on a large scale across the genome, and that patterns of introgression differ between the two species pairs. In one pair, gene flow is correlated with recombination rate, reduced at chromosome

centres and in gene-rich regions. In the other pair, rates of gene flow are more even across the genome. Using simulations, we show that these patterns are consistent with biological differences between the two species pairs. One pair has distinct wing patterns that are under strong ecological selection and may provide a genome-wide barrier to gene flow. The other pair lacks this dramatic ecological difference, so the species boundary depends more on the distribution of genetic incompatibilities and recombination. Our findings therefore supplement theoretical work, showing how the shape of the species boundary reflects the genetic architecture of species differences.

389

Microgeographic adaptation drives the evolution of reproductive isolation in parapatry

Henry L North[1], Maddie E James[1], Melanie J Wilkinson[1], Daniel Ortiz-Barrientos[1]

1. School of Biological Sciences, The University of Queensland, Brisbane, Queensland, Australia

Divergent evolution between taxa adapted to contrasting environments is challenging at a microgeographic scale, where populations are within the dispersal neighbourhood of one another. This is because gene flow homogenises the allelic differences created by natural selection. However, strong selection against migrants and hybrids can mitigate the effects of gene flow. We report a case study of microgeographic adaptation in two ecotypes of the Australian wildflower *Senecio lautus*, which display heritable phenotypes adapted to sand dune and rocky headland environments. These sister taxa interbreed readily in the laboratory but show reproductive isolation in the field despite their parapatric distribution. By using phenotypic and genotyping-by-sequencing data in geographic cline analysis, we are able to quantify divergence in adaptive traits and allele frequencies, natural selection, and the strength of barriers to gene flow. We observe stepped clines in the allele frequency of nine diagnostic loci and in ten phenotypic traits at a spatial scale of less than 50m. The estimated strength of selection was absolute for several traits and barriers to gene flow were complete for some loci. Consistent with the accumulation of complete reproductive isolation between the two ecotypes, the centre of phenotypic and allelic clines were tightly concordant. Despite this, allele frequencies across the genome were homogenous and genetic differentiation was minimal with F_{ST} = 0.029, suggesting that divergence is recent. These findings are consistent with the prediction that extreme fitness differences conferred by as few as one locus can facilitate divergent evolution and potentially speciation irrespective of spatial scale.

390

Genomics, Morphology and Ecology of the Lake Tanganyika Cichlid radiation

Fabrizia Ronco[1], Walter Salzburger[1]

1. University of Basel, Basel, BASEL, Switzerland

The assemblage of cichlid fishes from East African Lake Tanganyika consists of about 200 described and approximately 50 un-described species and shows an extraordinary degree of morphological and ecological diversity. Thus it provides a prime model system to study adaptive radiation in general and in a comparative context in particular as it harbors extremely species-rich as well as species-poor lineages. This integrative study will provide the most comprehensive examination of a cichlid radiation combining whole genome sequencing with geometric morphometric (in 2D and 3D) and stable isotope measurements (as a proxy for ecology) of all ~250 Tanganyika cichlids. This extensive dataset provides the unique opportunity to investigate how morphological and ecological differentiation reflects the phylogenetic history in this unique adaptive radiation.

391

Speciation gene evolution and the origin of chemical mimicry in a sexually deceptive orchid

Philipp M. Schlüter[1], Khalid E.M. Sedeek[1], John Shanklin[2]

1. Department of Systematic and Evolutionary Botany, University of Zurich, Zurich, ZH, Switzerland

2. Department of Biology, Brookhaven National Laboratory, Upton, NY, USA

Pollinator attraction via chemical mimicry of insect sex pheromones acts as automatic magic trait in Mediterranean *Ophrys* orchids, underlying both divergent adaptation and reproductive isolation between different species. Chemical mimicry is also the defining feature of ecological speciation and adaptive radiation of *Ophrys*. *SAD2* and *SAD5*, two members of a small gene family of acyl-ACP desaturases are mainly responsible for controlling the hydrocarbon double-bond composition of the pseudo-pheromone produced by the orchids, and thereby pollinator attraction and reproductive isolation. This makes them both major magic genes and ecological speciation genes. Evolutionary and molecular functional analysis of the desaturase family, including functional testing of resurrected (engineered) ancestral proteins, revealed how sex pheromone mimicry could have evolved from housekeeping functions of the ancestral desaturases and how selection by pollinators may have shaped the evolution of the speciation genes *SAD2* and *SAD5*.

392

Comparative RNA-seq analysis of antifungal immune responses among *Drosophila* species

Yosuke Seto[1], Koichiro Tamura[1, 2]

1. Department of Biological Sciences, Tokyo Metropolitan University, 1-1 Minami-Osawa, Hachioji-shi, Tokyo, Japan

2. Research center for genomics and bioinformatics, Tokyo Metropolitan University, 1-1 Minami-Osawa, Hachioji-shi, Tokyo, Japan

During feeding behavior, many animals are exposed to infection of various microorganisms from their digestive organs. Therefore, the immune system is important for survival against the infection. To elucidate the relationship between the immune system and feeding habitat, we conducted comparative RNA-seq analysis using *Drosophila* species.

Drosophila flies feed and breed on fermented fruits and sap, where a variety of bacteria and fungi propagate. *Drosophila* flies defend themselves from invading microorganisms with innate immune system. In our previous studies, we found that *Drosophila virilis*, which feeds on slime fluxes and decaying parts of trees, is more resistant to *Penicillium* fungus infection than *D. melanogaster*, which feeds on fermented fruits. To investigate the immune mechanism responsible for the difference in the antifungal resistance, we compared the expression patterns of immune-related genes in gut, in salivary gland and in fat body between the two species in response to the infection of *Penicillium* fungus. We found that *D. virilis* used mainly antibacterial peptide genes, Diptericin and Defensin, whereas *D. melanogaster* used mainly antifungal peptide genes, Drosomycin. Additionally, expressions of lysozyme genes were also up-regulated in the infected *D. virilis*. These results indicate that the immune system has been

substantially differentiated during the evolution of *D. melanogaster* and *D. virilis*. The differences in the immune system may have been evolved as an adaptive response to microbial environments.

393

Speciation with extensive gene flow in Lake Victoria cichlid species

Yohey Terai[1], Ryutaro Miyagi[2], Akie Sato[3], Shohei Takuno[1]

1. SOKENDAI (The Graduate University for Advanced Studies), Hayama, KANAGAWA, Japan

2. Department of Biological sciences, Tokyo Metropolitan University, Hachioji, Tokyo, Japan

3. Graduate School of Dental Medicine, Tsurumi University, Yokohama, Kabagawa, Japan

The molecular basis of the incipient stage of speciation is still poorly understood. Here, we report the pattern of genomic differentiation between two Lake Victoria cichlid species. Lake Victoria has dried up and has been refilled roughly 15,000 years ago. This lake harbors several hundreds endemic species of cichlids. They have undergone very recent and rapid speciation events during this short period. These fish species are genetically closely related and share nucleotide polymorphisms among species. So far, only one gene, a long-wavelength sensitive opsin (*LWS*), was identified as a gene bearing fixed genetic differences between species. Further analysis showed *LWS* was responsible for adaptation and speciation in cichlids. In this study, we analyzed genomic DNA sequences from two Lake Victoria cichlids (20 individuals each). The genetic differentiation between two species was extremely low, while 21 differentiated regions (DRs: 14~28 kb) with fixed differences were extracted by Fst sliding-window analysis. These regions contained one to three coding genes, one of which included *LWS*. Several of those genes were associated with vision, development, resistance to hypoxia, and differentiation of sexual traits. The expression level of four genes with fixed differences at upstream of the genes were completely different. Thus, at least the genes in five regions including *LWS* out of 21 regions were indeed responsible for the species differences, suggesting the genes in DRs may be responsible for the adaptation and speciation of Lake Victoria cichlids.

394

Genomics of adaptive divergence in East African cichlid fishes: a comparative approach

Alexandra A.-T. Weber[1], Jelena Rajkov[1], Walter Salzburger[1], Bernd Egger[1]

1. Zoological Institute, University of Basel, Basel, Switzerland

Species that show replicate adaptation of divergent populations along the same environmental gradient are important model systems to examine how new species form in natural populations. *Astatotilapia burtoni* is an East African cichlid fish occurring in Lake Tanganyika (LT) and affluent rivers. In southern LT, *A. burtoni* lake-stream population pairs display different degrees of genetic and morphological differentiation, suggesting that they rest at different stages of the 'speciation continuum'.To investigate the molecular basis of adaptation along the lake-stream environmental gradient, we performed whole genome resequencing of 12 individuals per population in three *A. burtoni* lake-stream population pairs, as well as one lake-stream population pair of another cichlid species, *Haplochromis stappersii (*sympatric with *A. burtoni* in northern LT). We first assessed population structure and morphological differences in both species from northern LT and compared them to the *A. burtoni* populations from the southern part of the lake.

Body shape analyses of *A. burtoni* and *H. stappersi* from the North revealed morphological differences along the lake-stream gradient, but they did not follow the same trajectories as in the southern populations. Using more than 7 millions SNPs, we detected contrasting patterns of genetic differentiation between lake and stream populations from the North and South of LT. Finally, F_{ST} outlier analyses among lake-stream population pairs revealed candidate genes of local adaptation involved in sensory systems, in communication (e.g. coloration) as well as in immunity. Overall, these results provide insights into cichlid speciation genomics and bring further understanding of how selection acts on natural populations.

395

Dominance and selection coefficients inferred from large-scale population data identify candidate recessive genes

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

1. Department of Medicine, Division of Genetics, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, United States

2. Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, USA

The quantification of diploid selection coefficients in specific human variants and genes remains largely elusive. Unlike model organisms, dominance (**h**) and selection (**s**) coefficients in humans must be inferred from natural population data. We present a method to estimate coarse average selection and dominance coefficients per gene by comparing Exome Aggregation Consortium[1] population genetic data in ~35,000 Europeans to simulated diploid alleles in a realistic demography[2]. We match putatively deleterious variants (nonsense and damaging missense) via informative summary statistics of the per-gene frequency spectrum. We classify genes as candidate strong selection recessives (h<0.1), strongly selected "non-recessives" (h>=0.1), under weak selection, nearly neutral, or sub-drift.

To validate our candidate recessive and non-recessive gene sets, we demonstrate significant enrichment in genes under recessive selection (and/or depletion of non-recessives) for autosomal recessive diseases, hearing loss, and in genes identified in consanguineous individuals with depleted homozygous LOF variants[3]. We replicate classical predictions of recessivity in large metabolic pathways (e.g. TCA), consistent with Wright's theory of the physiological origin of dominance[4,5], and GO annotated extracellular localization, and dominance in GO transcription factors[6]. We find significant enrichment for GO infertility, meiosis, and spermatogenesis genes in the recessive strong selection class, but no enrichment for oogenesis, suggesting a large autosomal recessive component to male-specific infertility consistent with mammalian studies in cattle[7].

To our knowledge this is the first large set of human candidate recessive genes (~1500) identified from panmictic population data. This is qualitatively consistent with recessivity observed in most deadly fly and yeast variants[8,9]. Notably, a large recessive component in many human genes is inconsistent with the simplifying assumption of additivity in previous estimates of selection against non-synonymous variants[10,11], since recessive genes under strong selection map to weak selection due to prevalent neutral heterozygotes. Thus, a dominance-aware marginal DFE substantially increases the average selection against deleterious human variants.

1. Exome Aggregation Consortium, et al. (2015) Analysis of protein-coding genetic variation in 60,706 humans. bioRxiv doi: http://dx.doi.org/10.1101/030338.
2. Tennessen JA, et al. (2012) Evolution and functional impact of rare coding variation from deep sequencing of human exomes. Science 337(6090):64–69.

3. Narashimhan VM, et al. (2016) Health and population effects of rare gene knockouts in adult humans with related parents. Science 352(6284):474--477.
4. Wright, S. (1934) Physiological and Evolutionary Theories of Dominance. American Naturalist. 68(714): 24--53.
5. Kacser, H, Burns, JA (1981) The molecular basis of dominance. Genetics 97:639–666.
6. Seidman JG, Seidman C (2002) Transcription factor haploinsufficiency: when half a loaf is not enough. J Clin Invest 109: 451–455.
7. VanRaden PM, et al. (2011) Harmful recessive effects on fertility detected by absence of homozygous haplotypes. J Dairy Sci. 2011 Dec;94(12):6153-61.
8. Mukai T (1972) Mutation rate and dominance of genes affecting viability in Drosophila Melanogaster. Genetics 72:335–355.
9. Agrawal AF and Whitlock MC (2011) Inferences about the distribution of dominance drawn from yeast gene knockout data. Genetics 187:553–566. doi: 10.1534/genetics.110.124560.
10. Boyko, AR, et al. (2008) Assessing the evolutionary impact of amino acid mutations in the human genome. PLoS Genet. 4, e1000083.
11. Kryukov, GV et al. (2009) Power of deep, all-exon resequencing for discovery of human trait genes. Proc. Natl. Acad. Sci. U.S.A. 106, 3871-3876.

396

Profiles of low complexity regions in Apicomplexa

fabia ursula battistuzzi[1], Kristan A Schneider[2], Matthew K Spencer[3], David Fisher[4], Sophia Chaudhry[5], Ananias A Escalante[6]

1. Department of Biological Sciences, Oakland University, Rochester, Michigan, United States
2. Department of MNI, University of Applied Sciences, Mittweida, Germany
3. Department of Geology and Physics, Lake Superior State University, Sault Ste. Marie, Michigan, United States
4. David Eccles School of Business, University of Utah, Salt Lake City, UT, United States
5. Center for Molecular Medicine and Genetics, Wayne State University, Detroit, MI, United States
6. Institute for Genomics and Evolutionary Medicine, Temple University, Philadelphia, PA, United States

Low complexity regions (LCRs) are a ubiquitous feature in genomes and yet their evolutionary history and functional roles are unclear. Previous studies have shown contrasting evidence in favor of both neutral and selective mechanisms of evolution for different sets of LCRs suggesting that modes of identification of these regions may play a role in our ability to discern their evolutionary history. To further investigate this issue, we used a dynamic approach to identify species-specific profiles of genome complexity and, by comparing properties of these sets, determine the influence that starting parameters have on evolutionary inferences. We find that, although qualitatively similar, quantitatively each species has a unique LCR profile which represents the frequency of these regions within each genome. Inferences based on these profiles are more accurate in comparative analyses of genome complexity as they allow to determine the relative complexity of multiple genomes as well as the type of repetitiveness that is most common in each. Based on the dynamic LCR sets obtained, we identified predominant evolutionary mechanisms at different complexity levels, which show neutral mechanisms acting on highly repetitive LCRs (e.g., homopolymers) and selective forces becoming more important as heterogeneity of the LCRs increases. Our results show how inferences based on LCRs are influenced by the parameters used to identify these regions. Sets of LCRs are heterogeneous aggregates of regions that include homo- and hetero-polymers and, as such, evolve according to different mechanisms. LCR profiles provide a new way to investigate genome complexity across species and to determine the driving mechanism of their evolution.

397

Assessing the fitness cost of Bt insecticidal toxin resistance over time.

Chris Ward[1], David G Heckel[2], Simon Baxter[1]

1. University of Adelaide, Adelaide, SA, Australia
2. Max Planck Institute for Chemical Ecology, Jena, Germany

Field evolved resistance to Bt insecticidal toxins has occurred in the diamondback moth, *Plutella xylostella*, and despite being widely studied, the fitness costs associated with resistance remain unclear. In the absence of selection, neutral polymorphisms with no fitness costs can be maintained in large populations through genetic drift. However, if the selection coefficient is high, deleterious alleles should be removed from populations under purifying selection. Insecticide resistance alleles without fitness cost may therefore pose a serious threat in agriculture, as they could persist in populations over time and undergo rapid re-selection.

A Hawaiian population of the diamondback moth, NO-QA, has evolved resistance to Bt insecticides through a 30bp deletion in the ATP-dependent Binding Cassette transporter gene *abcc2*. To assess fitness costs associated with resistance, NO-QA was crossed with a Bt susceptible strain, Waite, and their progeny used to found four replicate population cages with an initial resistance frequency of 0.5. Caged populations were maintained for 19 generations without insecticide exposure then genotyped at generation 10 (n=192) and 19 (n=384) to assess changes *abcc2* resistance frequency over time. The frequency of resistance alleles reduced over time from 0.5 to and average of 0.31, yet replicates were highly variable, indicating genetic drift may have had a substantial influence. Selection coefficients for each replicate population ranged from 0 to less than -0.5, which supported a fitness cost. Reduced genome representation libraries (RAD-seq) for NO-QA, Waite, and four population cage replicates were then sequenced to assess potential effects of strain bias across the genome. After 19 generations, the average genome composition was biased towards the Bt susceptible Waite strain. Nevertheless, this data demonstrates this type of Bt resistance comes with a fitness cost in diamondback moth.

398

Using evolutionary rate covariation to detect novel evolutionary patterns in the eudicot (Brassicaceae) and monocot (Poaceae) plant lineages

Jae Young Choi[1], Nathan Clark[2], Michael Purugganan[1]

1. New York University, New York, NY, United States
2. University of Pittsburgh, Pittsburgh, PA, United States

Genes with a common function are expected to be under the same evolutionary pressure and hence their underlying species phylogenetic rates are often correlated across long evolutionary times. This phenomenon, previously termed Evolutionary Rate Covariation (ERC), has been studied in yeasts, fruit flies, and mammals, which revealed previously uncharacterized genes interacting with known protein complexes and genetic networks.

It has also shown novel evolutionary processes influencing genetic networks. In this study we selected genome sequences from the eudicot (Brassicaceae) and monocot (Poaceae) plant lineages to test the power of the ERC method in plants. Using available functional genomic data in *Arabidopsis* and *Oryza* we have investigated whether changes in evolutionary pressure and expression levels, which were largely responsible for the ERC correlations in the animal and fungal kingdoms, are able to explain the ERC correlations in the plant kingdom. As an example, genes involved in the abiotic stress network were examined to find specific genes in the network that have experienced the highest ERC associations. We then used this information to find novel candidate genes that are involved in the stress network. Our analysis suggested the ERC method as a powerful tool to detect novel genes interacting in known plant genetic networks.

399

Finding needles in a genomic haystack: targeted capture to identify signatures of selection in a non-model plant species

Matthew J Christmas[1], Ed Biffin[2], Martin F Breed[1], Andrew J Lowe[1]

1. University of Adelaide, Adelaide, SA, Australia

2. State Herbarium of South Australia, Adelaide, SA, Australia

Teasing apart neutral and adaptive genomic processes and identifying loci that are targets of selection can be difficult, particularly for non-model species lacking a reference genome. However, identifying such loci and the factors driving selection have the potential to greatly assist conservation practices, especially for the management of species in the face of contemporary and future climate change. Here, I present on assessing adaptive genomic variation within a non-model plant species, the narrow-leaf hopbush (*Dodonaea viscosa* ssp. *angustissima*), commonly used for restoration in Australia. We used a hybrid-capture target enrichment approach to selectively sequence 970 genes across 17 populations along a latitudinal gradient from 30°S to 36°S. 8,462 single-nucleotide polymorphisms (SNPs) were analysed for F_{ST} outliers as well as associations with environmental variables. Using these methods we found 50 SNPs with significant correlations to temperature and water availability, and 24 SNPs to elevation. Genes containing SNPs identified as under environmental selection were diverse, including aquaporin and abscisic acid (ABA) genes, as well as genes with ontologies relating to environmental responses. Redundancy analysis demonstrated that only a small proportion of the total genetic variance was explained by environmental variables. We demonstrate that selection has led to clines in allele frequencies in a number of functional genes, including those linked to leaf shape and stomatal variation, which have been previously observed to vary along the sampled environmental cline. Using our approach, gene regions subject to environmental selection can be readily identified for non-model organisms.

400

Evolution of Tumor Necrosis Factor Superfamily (TNFSF) genes TNFSF 12 and 13: Phylogenetic clues for the emergence of *in genomic* fusion of *TNFSF12-TNFSF13*

Sabyasachi Das[1], Yoichi Sutoh[1], Masayuki Hirano[1], Qifeng Han[1], Jianxu Li[1], Max D. Cooper[1], Brantley R. Herrin[1]

1. Emory University, Atlanta, GA, United States

BAFF and APRIL (TNFSF13 members) are important regulatory factors for lymphocyte activation and survival. Another TNFSF13 gene that encodes a BAFF and APRIL-like molecule (BALM) was found in fish. Here we report the molecular characterization of the first TNFSF13 homolog in lampreys, a jawless vertebrate representative. In an investigation of the evolution of *BAFF*, *APRIL*, *BALM* and a closely related gene called *TWEAK* (*TNFSF12*), we identified an ancestral TNFSF13 gene in jawless vertebrates, but not the *TWEAK* gene. Hence, while *TWEAK* evolved in jawed vertebrates the*TNFSF13* gene appeared before the divergence of jawed and jawless vertebrates. Considering the encoded protein of ancestral *TNFSF13* gene in lamprey possess more BAFF like features than that of APRIL, we could notice that *BAFF* is present in all vertebrates, but *APRIL* and/or *BALM*independently lost in different lineages. For example, *BALM* is absent in all tetrapod genomes, and *APRIL* is lost in birds and several fish species. *TWEAK* is also lost in bird lineage suggesting that the genetic network of immune related genes have greatly reconstructed in birds genome. The comparative genome and transcriptome analyses suggest that an *in-genomic* fusion between *APRIL* and closely related *TWEAK* genes that produce a hybrid molecule called TWE-PRIL originated in mammalian lineage. Like mammalian *BAFF* and *APRIL* the ancestral TNFSF13 in lamprey exhibits a wide range of tissue and cellular expression including innate lymphoid cells and T cell-like (VLRA and VLRC) and B cell-like (VLRB) lymphocytes in early stage of vertebrate evolution.

401

Large scale phylogenetic patterns in waterbeetles

Aurélie Désamoré[1], Johannes Bergsten[1]

1. Natural History Museum of Stockholm, Stockholm, Sweden

Coleoptera is the largest order in the Animal Kingdom, comprising a quarter of all-known animal species. With more than 4200 described species, Dytiscidae (predaceous diving beetles) represents one of the largest and most diverse beetle families and is the most commonly encountered group of aquatic insects. In contrast to most of other insects, they have an adult stage that is truly aquatic resulting from a secondary return to the water environment during their evolution. They are occurring in various habitat types, where they form multi-species assemblages due to their high diversity and large variation in body size and habitat specificity.

We re-analyzed the molecular data from the recently published phylogeny of 164 Dytiscidae based on 9 DNA sequence fragments with the exclusion of highly variable regions and of the morphological characters. We dated the phylogeny with 12 fossils as a calibration and tested for shift in diversification rates and in body size. We also reconstructed the biogeographical history of the group using model-based likelihood.

This reanalyzed phylogeny revealed stronger support for important basal nodes. However, some relationships remain unsolved and call for a genomic approach to better resolve the evolution of Dytiscidae. The crown group of Dytiscidae was dated as ca. 211 Ma (Triassic) that is similar to the age recovered in the dating of the whole Coleoptera phylogeny (ca 219 Ma). The origin of the family was inferred as American with several independent dispersions to Palearctic and African regions. Despite the fact that the family is the most species rich among aquatic insects, we found no evidence for an acceleration of speciation during their evolution. However, the rate of morphological evolution, as represented by body size, revealed large heterogeneity among clades, showing that phenotypic differentiation has been decoupled from species diversification.

402

Untying the Evolution of Vision Using the Gordian Worms

James Fleming[1], Davide Pisani[1], Nicholas Roberts[1], Mark Blaxter[2], Georgios Koutsovoulos[2], Philipp Schiffer[3]

1. University of Bristol, Bristol, United Kingdom

2. University of Edinburgh, Edinburgh, United Kingdom

3. University College London, London, United Kingdom

Opsin proteins regulate a number of photoreceptive perceptions across the metazoa, including vision. The Ecdysozoa possess a wide variation in the morphology and acuity of visual systems. However, the opsins of the basal clades have largely been unexplored, and little is known about their visual modalities.

The Nematomorpha (Gordian, or Horsehair worms) is a parasitoid phylum comprising around 320 named species. Due to their close relationship to the highly derived and much more species-rich Nematoda (roundworms), understanding the Nematomorpha is key to our understanding of both the evolutionary history of the Ecdysozoa, and vision, as the Nematoda do not utilise opsin-mediated light reception. We present new opsins from both transcriptome and genome data of the Nematomorpha and show that the loss of opsin-mediated light reception in the Nematode worms is to be considered a lineage specific phenomenon.

403

Combining Shigella Tn-seq data with Gold-standard E. coli Gene Deletion Data Suggests Rare Transitions between Essential and Non-essential Gene Functionality

Nikki E Freed[1], Dirk Bumann[2], Olin Silander[1]

1. Massey University, Albany, AUCKLAND - NORTH SHORE, New Zealand

2. Infection biology, University of Basel, Basel, Switzerland

Gene essentiality - whether or not a gene is necessary for cell growth - is a fundamental component of gene function. It is not well established how quickly gene essentiality can change, as few studies have compared empirical measures of essentiality between closely related organisms. Here we present the results of a Tn-seq experiment designed to detect essential protein coding genes in the bacterial pathogen Shigella flexneri 2a 2457T on a genome-wide scale. Superficial analysis of this data suggested that 451 protein-coding genes in this Shigella strain are critical for robust cellular growth on rich media. Comparison of this set of genes with a gold-standard data set of essential genes in the closely related Escherichia coli K12 BW25113 revealed that an excessive number of genes appeared essential in Shigella but non-essential in E. coli. Importantly, and in converse to this comparison, we found no genes that were essential in E. coli and non-essential in Shigella, implying that many genes were artefactually inferred as essential in Shigella. Controlling for such artefacts resulted in a much smaller set of discrepant genes. Among these, we identified three sets of functionally related genes, two of which have previously been implicated as critical for Shigella growth, but which are dispensable for E. coli growth. The data presented here highlight the small number of protein coding genes for which we have strong evidence that their essentiality status differs between the closely related bacterial taxa E. coli and Shigella. A set of genes involved in acetate utilization provides a canonical example. These results leave open the possibility of developing strain-specific antibiotic treatments targeting such differentially essential genes, but suggest that such opportunities may be rare in closely related bacteria.

404

Origins and conservation of mammalian BRAF pseudogenes.

Cian Glenfield[1], Aoife McLysaght[1]

1. Smurfit Institute of Genetics, Trinity College Dublin, Dublin

Competitive endogenous RNAs (ceRNAs) represent a novel class of post-transcriptional gene regulators. CeRNAs function by competing for microRNAs which have binding sites shared with other RNA transcripts. Expressed pseudogenes and other long non-coding RNAs can thus alter the levels of mRNA from protein-coding genes, resulting in reduced protein expression. Proper regulation of ceRNA expression appears to be critical in certain cases, as several studies have shown that altering ceRNA levels can have an impact on cancer development and progression. However, few studies to date have considered the evolutionary impact and origins of ceRNAs, presumably in part due to difficulties with pseudogene and other long non-coding RNA discovery and annotation. Here we report our findings on the evolutionary analysis of the pseudogene BRAFP1 as a conserved ceRNA. We found that the pseudogene is present in syntenic locations in each species of the Catarrhini lineage, which is comprised of the Old-world monkeys and Apes. Multiple sequence alignment reveals that the 3'UTRs of the pseudogenes have a lower substitution rate than their pseudo-protein-coding regions, as well as a similar substitution rate to their respective parent gene 3'UTRs. In addition, we found several miRNA binding sites that are conserved between each parent gene and pseudogene. Our preliminary results indicate that pseudogenes and likely other long non-coding RNAs can represent a novel method of conserved post-transcriptional gene regulation. We anticipate this analysis will help facilitate further discovery of conserved ceRNAs and their consequences.

405

Transient Receptor Potential Gene Family Evolution in two echinoderms, sea urchin and starfish

Jun Gojobori[1]

1. SOKENDAI (The Graduate University for Advanced Studies), Hayama, KANAGAWA, Japan

Sensing the temperature of the surrounding environment should be an important trait for organisms in terms of adaptation. Several members of transient receptor potential (TRP) superfamily are known to function as thermo sensors. These genes are called as thermoTRPs. TRPs are ion channels which have six transmembrane domains. ThermoTRPs open their channel by the heat or cold stimuli, but the number of thermoTRPs members or the temperature they sense is frequently changed during evolution. I conducted analysis to find out the candidates of TRPs in the genomes and transcriptomes of two echinoderms, sea urchin and starfish. The larvae of them have been shown to have thermotaxis and the thermoTRPs of them are supposed to play a great role. Based on HMM search using known TRPs, I found that sea urchin and starfish both potentially have almost the same number of TRP genes as vertebrates. I also found that sea urchin has more TRPA genes than fruit fly or human. The unusual expansion of TRPA genes in sea urchin may have functionally important for its thermotaxis and adaptation.

406

Evolution of Animal Allorecognition: A Case Study in Sponges

Laura F Grice[1], Marie EA Gauthier[1], Kerry E Roper[1], Xavier Fernàndez-Busquets[2, 3, 4], Sandie M Degnan[1], Bernard M Degnan[1]

1. School of Biological Sciences, The University of Queensland, Brisbane, Queensland, Australia

2. Nanomalaria Group, Institute for Bioengineering of Catalonia (IBEC), Barcelona, Spain

3. Barcelona Institute for Global Health (ISGlobal), Hospital Clínic-Universitat de Barcelona, Barcelona, Spain

4. Nanoscience and Nanotechnology Institute (IN2UB), University of Barcelona, Barcelona, Spain

Although the ability to discriminate self from nonself is a highly conserved animal trait, the systems enabling animal allorecognition vary markedly between taxa, suggesting independent evolutionary origins. The disparity between extant recognition systems means that understanding of the evolutionary processes shaping their underlying genes is limited. Here we explore an allorecognition gene family across an entire animal phylum, the sponges. A peculiarity of sponge physiology allows dissociated, histocompatible cells to sort into aggregates, and later, functional sponges. This process is mediated by aggregation factors (AFs), proteoglycans which have also been implicated in the sponge immune response to tissue contact, but which have not been analysed at a genomic level. We show that the demosponge *Amphimedon queenslandica* genome has six tightly-clustered AF genes, which encode proteins containing multiple Calx-beta and von Willebrand domains and a newly-defined Wreath domain. Despite these genes having remarkably similar exon, intron phase, and domain structures, high nucleotide and amino acid polymorphism exists between individuals, and appears to be introduced both by genomically-encoded variants and, in at least one AF gene, extensive single-nucleotide RNA editing of up to 5% of surveyed nucleotide positions. Analysis of 24 sponge genomes and transcriptomes spanning all four poriferan classes reveals that all surveyed demosponges possess AFs with Wreath domains and other similar domains to *Amphimedon* AFs. However, no two demosponges share matching AF gene repertoires. Representatives of other sponge classes completely lack recognisable AFs and Wreath domains. These results are consistent with the AF gene family originating at the base of the demosponges and undergoing rapid evolution via domain shuffling, recruitment and loss, resulting in the range of AFs present in modern demosponges. The continual evolution of AFs in sponges provides an explanation as to how other allorecognition genes present in the animal kingdom obtain their unique structure and organisation.

407

Worldwide Genetic Variation in Serotonin Pathway Genes Associated with Bipolar Disorder

Jayla Harvey[1], Michael Campbell[1]

1. Howard University, Washington, DISTRICT OF COLUMBIA, United States

Bipolar disorder (BPD) is a brain illness characterized by dramatic shifts in mood and activity levels, affecting ~60 million people worldwide. Despite the severity of these symptoms, management of BPD can be successful if diagnosis and treatment are initiated early. Consequently, research has focused on identifying genetic factors that facilitate early detection and/or treatment of this illness. In particular, genes in the serotonin pathway have been under intense investigation given their strong association with other behavior-related disorders with similar symptoms. However, little is still known about patterns of genetic variation in the human serotonin system in natural populations, which can impact treatment. To address this current gap in knowledge, we examined 20 genes in the serotonin pathway, totaling >1.6 million bases of sequence data, in ~1400 individuals from worldwide populations. Here, we report striking patterns of diversity in the *SLC6A4* gene, including an excess of high-frequency polymorphisms and long-range haplotypes, consistent with a model of positive selection in African populations. We also inferred that non-coding polymorphisms at *SLC6A4* are likely the targets of selection, raising the possibility that these variants may play a role in gene expression. Although the precise function of these polymorphisms is currently unknown, we argue that they are or have been selectively advantageous during evolutionary history, representing new candidate loci for further study. Because serotonin transport inhibitors are first-line treatments for BPD, our study, the largest of its kind to date, will be informative for the development of targeted interventions based on more "personalized" genomic information.

408

Transcriptional evidence for a reproduction_immunity trade-off in Japanese quail (*Coturnix japonica*)

Kate E Ihle[1]

1. University of Zurich, Zurich, ZURICH, Switzerland

The cost of reproduction is thought to be a major regulator of life-history evolution. Increased investment in reproduction is thought to come at the expense of self-maintenance, and thus limit parental lifespan. The regulation of the balance between reproduction and self-maintenance is still poorly understood, but investment in reproduction is predicted to come at the cost of immune function. In birds maternal investment in egg components is energetically expensive, and large variations in the level of maternally provided resources have been documented in natural populations. In wild populations, the limited availability of nutrients and other recourses is often suggested o mediate life-history trade-offs. Here we examine the evidence for a reproduction-immunity trade-off at the transcriptional level in the absence of resource limitation in a captive population of Japanese quail (*Coturnix japonica*) divergently selected for high versus low maternal investment in egg size. As all animals shared a resource-rich environment, any trade-offs between reproductive and immune investment are more likely to be related to intrinsic factors than to external resource availability. Using whole-transcriptome RNAseq, we identified 30 consistently differentially expressed genes in ovarian follicle cells in two replicates of the high and low maternal investment lines. In the high maternal investment line genes associated with reproductive investment are upregulated, while genes linked to immune function were downregulated. Our data provide experimental evidence for an intrinsic reproduction-immunity trade-off in the absence of resource limitations in a precocial bird.

409

A genomewide survey of genes for muscle structural proteins that enabled the rise of chordates

Jun Inoue[1], Noriyuki Satoh[1]

1. Okinawa Institute of Science and Technology Graduate University, Kunigami-gun, OKINAWA, Japan

Evolution of both free-living and sessile chordates is closely associated with the acquisition of tadpole-like body plan. Our ancestor adopted a new style of swimming by beating of tail formed by notochord and muscle. Comparison of genomes is expected to provide insight into the molecular genetic basis of such novel features of chordates. To infer the evolutionary process in muscle structural proteins, we estimated gene trees based on sequences derived from genome data representing major deuterostome lineages. In order to reconstruct the gene trees employing exhaustive sequence sampling, we constructed an analysis pipeline consisting of two-step approach: 1) selection of ortholog candidates using BLAST search

and neighbor joining analysis, and 2) estimation of gene trees using maximum likelihood analysis. The resultant gene trees suggest that muscle protein isoform families of chordates (e.g, actin and myosin heavy chain) tend to be the products of independent gene duplications in each lineage. In most cases, the genes in ascidians, apart from those in lancelets and other deuterostomes, form a sister group with a clade including highly diversified genes in vertebrates. Our results suggest that the tadpole-like body plan, originated before the separation of ascidian and vertebrate lineages, diverged independently by increasing the number of genes within each lineage.

410

Population genetic processes affecting the genomic pattern of sequence diversity in influenza virus H3N2

Kangchon Kim[1], Yuseob Kim[1, 2]

1. Interdisciplinary Program of EcoCreative, Ewha Womans University, Seoul, Korea

2. Department of Life Sciences, Ewha Womans University, Seoul, Korea

Relative contributions of population genetic processes to molecular evolution have been an important and controversial question in evolutionary genetics. We quantify the impacts of population genetic processes underlying evolution of human influenza virus H3N2 on reducing the effective population size of HA (hemagglutinin) segment. For this analysis, we use computer simulation of viral population reproducing in discrete generations where each virus sequence represents the state of viruses infecting one host. We used mutation rate and selection coefficients for beneficial and deleterious alleles estimated by observed changes of variant allele frequencies and nonsynonymous-to-synonymous substitution ratios. The effective size of population in which selective sweeps occur is inferred by the rate of soft selective sweeps. This size relative to the census size of viral population is used to estimate how much reduction in the effective size occurs by other processes: background selection and meta-population dynamics. The variation-reducing power of background selection and meta-population dynamics is greater than that of recurrent positive selection and must be crucial in explaining the observed level of sequence diversity in H3N2. We also found that per-site synonymous diversity of genomic segments varies with segment length and the rate of adaptive evolution, which can be explained by interplay between negative and positive selection. The joint analysis of these processes on the genomic pattern of variation provides insight on the rate of reassortments in H3N2 viruses, which might otherwise be difficult to obtain.

411

Rapid replacement of centromeres by a variant-type repetitive DNA in a primate taxon

Akihiko Koga[1]

1. Primate Research Institute, Kyoto University, Inuyama City, Japan

Centromeres usually contain large amounts of tandem repeat DNA, which often exceeds the megabase-pair level in size. Alpha satellite DNA is a major centromeric repetitive DNA of simian primates. Humans carry two types of alpha satellite DNA: simple repeats of 171-bp repeat units, and those exhibiting higher-order repeat structures (multiple copies of 171-bp units form a larger repeat unit and the larger units appear periodically). The latter type is known to be evolutionarily new and to play more significant roles in centromere formation. We have recently identified two types of alpha satellite DNA in owl monkeys (genus *Aotus*; New World monkey). The two types were named OwlAlp1 and OwlAlp2. With respect to the structure, the two types of human alpha satellite DNA differ in repeat organization, but the difference between OwlAlp1 and OwlAlp2 is in the size of their repeat units. The repeat units of OwlAlp1 (185 bp) correspond to part of those of OwlAlp2 (344 bp). All other New World monkeys examined thus far (marmosets, squirrel monkeys, capuchins, tamarins etc.) appear to have only alpha satellite DNA close in size and sequence to OwlAlp2. Thus, OwlAlp1 is considered to be evolutionarily new, derived from OwlAlp2 by partial deletion. OwlAlp1 occupies the centromeric constriction region of all chromosomes, whereas OwlAlp2 is present in pericentric regions of most, but not all, chromosomes, suggesting that OwlAlp1 plays more significant roles. These features support the view that OwlAlp1 replaced OwlAlp2 as the principal centromeric repetitive DNA in the owl monkey lineage after its divergence from those leading to other New World monkeys.

412

Demographic impact on genome-wide signatures of selection

Verena E. Kutschera[1], Nicolas Dussex[2], Robert Fleischer[3], Neil J. Gemmell[2], Russell D. Gray[4, 5], Gavin Hunt[4], Michael G. Ritchie[6], Kim Rutherford[2], Christian Rutz[6], R. Axel W. Wiberg[6], Jochen B. W. Wolf[1]

1. Department of Evolutionary Biology, Evolutionary Biology Centre, Uppsala University, Uppsala, Sweden

2. Department of Anatomy, University of Otago, Dunedin, New Zealand

3. Smithsonian Conservation Biology Institute, National Zoological Park, Washington, DC, USA

4. Department of Psychology, University of Auckland, Auckland, New Zealand

5. Max Planck Institute for Science of Human History, Jena, Germany

6. Center for Biological Diversity, School of Biology, University of St. Andrews, St. Andrews, United Kingdom

Evolutionary change is driven by a combination of neutral and selective processes. Selection is the prevailing evolutionary force sifting through genetic variation allowing organisms to adapt and seize novel opportunities (positive selection), but likewise to maintain systemic functionality (purifying selection). As selection efficacy strongly depends on the effective population size (N_e), its interplay with genetic drift needs to be considered. However, owing to a historical lack of genome-wide data, our knowledge about the relative importance and strength of selection acting on the genome is limited.

Comparative genomic approaches are powerful to address this question. We generated genomic resources for several species within the avian genus *Corvus* (crows, rooks, and jackdaws). Here we focus on a species pair with a strong contrast in expected population sizes using population genomic re-sequencing data from an island species, the New Caledonian crow (*C. moneduloides*), and from a widespread species, the European crow (*C. corone*). We compare estimates of the distribution of fitness effects of new mutations (DFE), and of levels of positive selection using McDonald-Kreitman type statistics. The marked difference in N_e between the species under investigation allows us to isolate the effect of genetic drift on the selective genomic landscape, thus addressing a central question arising from evolutionary theory.

413

Fitness pleiotropy and the phenotypic basis of adaptation in experimentally evolving yeast

Yuping Li[1], Sandeep Venkataram[1], Atish Agarwala[1], Dmitri Petrov[1], Gavin Sherlock, Daniel Fisher[1]

1. Stanford University, Stanford, CALIFORNIA, United States

Understanding how the fitness of a beneficial mutation selected in one environment varies in other environments is essential for understanding the evolutionary process. However, difficulty in obtaining a large number of independent adaptive mutations, and being able to measure their fitness across multiple environments has constrained our ability to study the pleiotropy of beneficial mutations. We have overcome these limitations using a DNA barcode approach, which we have used to isolate 4800 yeast lineages independently evolved in glucose-limited batch culture conditions. High throughput fitness measurements have shown that thousands of these lineages are adaptive, and whole-genome sequencing has identified the genetic basis of adaptation in 300 lineages. To investigate fitness pleiotropy, we have systematically manipulated the exponential or stationary growth phases within each growth cycle, and remeasured fitness. Our experiments showed that the fitnesses of the adapted clones are highly pleiotropic across these conditions, and in some cases exhibit antagonistic pleiotropy. In addition, the measured fitness is dependent on both the identity of the specific genes carrying a mutation and on the mutation types (e.g. missense *vs* frameshift). We have also conducted detailed physiological experiments to study the phenotypic basis of the observed fitness pleiotropy. Measurement of the growth cycle and metabolism of the adapted strains indicates that the adaptive mutations affect multiple phenotypic traits and helps explain the strong fitness pleiotropy observed. Taken together, our results suggest that fitness is exquisitely sensitive to specific environmental conditions and that adaptive mutations exhibit high phenotypic pleiotropy to drive these fitness effects.

414

Gene Duplicability of Core Genes Is Highly Consistent across All Angiosperms

Zhen Li[1, 2, 3], Jonas Defoort[1, 2, 3], Setareh Tasdighian[1, 2, 3], Steven Maere[1, 2, 3], Yves Van de Peer[1, 2, 3, 4], Riet De Smet[1, 2, 3]

1. Department of Plant Systems Biology, VIB, Ghent, Belgium

2. Department of Plant Biotechnology and Bioinformatics, Ghent University, Ghent, Belgium

3. Bioinformatics Institute Ghent, Ghent University, Ghent, Belgium

4. Genomics Research Institute, University of Pretoria, Pretoria, South Africa

Gene duplication is an important mechanism for adding to genomic novelty. Hence, which genes undergo duplication and are preserved following duplication is an important question. It has been observed that gene duplicability, or the ability of genes to be retained following duplication, is a nonrandom process, with certain genes being more amenable to survive duplication events than others. Primarily, gene essentiality and the type of duplication (small-scale versus large-scale) have been shown in different species to influence the (long-term) survival of novel genes. However, an overarching view of "gene duplicability" is lacking, mainly due to the fact that previous studies usually focused on individual species and did not account for the influence of genomic context and the time of duplication. Here, we present a large-scale study in which we investigated duplicate retention for 9178 gene families shared between 37 flowering plant species, referred to as angiosperm core gene families. For most gene families, we observe a strikingly consistent pattern of gene duplicability across species, with gene families being either primarily single-copy or multicopy in all species. An intermediate class contains gene families that are often retained in duplicate for periods extending to tens of millions of years after whole-genome duplication, but ultimately appear to be largely restored to singleton status, suggesting that these genes may be dosage balance sensitive. The distinction between single-copy and multicopy gene families is reflected in their functional annotation, with single-copy genes being mainly involved in the maintenance of genome stability and organelle function and multicopy genes in signaling, transport, and metabolism. The intermediate class was overrepresented in regulatory genes, further suggesting that these represent putative dosage-balance-sensitive genes.

415

Tangled up in two: a burst of genome duplications at the end of the Cretaceous and the consequences for plant evolution

Yves Van de Peer[1, 2, 3, 4], Kevin Vanneste[1, 2, 3], Steven Maere[1, 2, 3]

1. Department of Plant Systems Biology, VIB, Ghent, Belgium

2. Department of Plant Biotechnology and Bioinformatics, Ghent University, Ghent, Belgium

3. Bioinformatics Institute Ghent, Ghent University, Ghent, Belgium

4. Genomics Research Institute, University of Pretoria, Pretoria, South Africa

Genome sequencing has demonstrated that besides frequent small-scale duplications, large-scale duplication events such as whole genome duplications (WGDs) are found on many branches of the evolutionary tree of life. Especially in the plant lineage, there is evidence for recurrent WGDs, and the ancestor of all angiosperms was in fact most likely a polyploid species. The number of WGDs found in sequenced plant genomes allows us to investigate questions about the roles of WGDs that were hitherto impossible to address. An intriguing observation is that many plant WGDs seem associated with periods of increased environmental stress and/or fluctuations, a trend that is evident for both present-day polyploids and palaeopolyploids formed around the Cretaceous–Palaeogene (K–Pg) extinction at 66 Ma. I will revisit the WGDs in plants that mark the K–Pg boundary, and discuss some specific examples of biological innovations and/or diversifications that may be linked to these WGDs. I will review evidence for the processes that could have contributed to increased polyploid establishment at the K–Pg boundary, and discuss the implications on subsequent plant evolution in the Cenozoic.

416

Rises and falls of opsin genes in 46 fish genomes and their implications for environmental adaptation

Jinn-Jy Lin[3, 1, 4, 2], Feng-Yu Wang[5], Tzi-Yuan Wang[3], Wen-Hsiung Li[3, 1]

1. Institute of Molecular and Cellular Biology, National Tsing Hua University, Hsinchu, Taiwan

2. Bioinformatics Program, Taiwan International Graduate Program, Institute of Information Science, Academia Sinica, Taipei, Taiwan

3. Biodiversity Research Center, Academia Sinica, Taipei, Taiwan

4. Academia Sinica, Taipei, TAIWAN, Taiwan

5. Taiwan Ocean Research Institute, National Applied Research Laboratories, Kaohsiung, Taiwan

The composition and functionality of opsin genes in a fish affect its visual capabilities and adaption to its habitat. Previously not many fish genomes were available and the changes in opsin gene sequences and copy number during fish evolution could not be fully revealed. We analyzed 46 fish

genomes to study the genomic organization and evolution of opsin genes. We found that each round of whole genome duplication (WGD) had a strong effect on the syntenic organization of opsin gene-bearing loci and different opsin gene families experienced different kinds of changes. The tandem duplication was the most frequent duplication event that changed the opsin gene copy number in lineage-specific manner. Finally, we examined the evolution of key tuning sites in opsins in relation to the visual adaptation of fishes to different habitats such as epipelagic or demersal. Our study provides a detailed view of evolutionary changes in tuning sites, copy number and synteny of opsin genes in fishes and sheds light on the role of opsin genes in fine-tuning fish's adaptation to diverse aqueous environments.

417

Single cell and metagenomics help understand early stage archaeal endosymbionts

Anders E Lind[1], William Lewis[2], Lionel Guy[1], Martin Embley[2], Thijs JG Ettema[1]

1. Uppsala University, Uppsala, UPPSALA, Sweden

2. Institute for Cell and Molecular Biosciences, University of Newcastle, Newcastle, United Kingdom

There are few known cases of archaeal endosymbiosis[1,2]. Here we present the draft genomes of two such endosymbionts, both living inside eukaryotic ciliates. The hosts in question are *Nyctotherus ovalis* and *Metopus contortus*, both known to harbor methane-producing endosymbionts. We have applied culture independent methods such as single cell and metagenomics in order to get insights into the evolutionary history of these organisms. Both methods show remarkably similar results in terms of assembly efficiency and completeness, and both methods seem to be viable methods for whole genome sequencing of endosymbiotic organisms.

Their genomes shows that these are indeed evolutionary separate events, and most likely also recent events. Both endosymbionts show telltale signs of adaptation to endosymbiosis[3], albeit at an early stage. Neither of the genomes shows signs of heavy genome size reduction, but there is a gene loss due to pseudogenization in both genomes, where the endosymbionts of *N. ovalis* seem to be further along the process. Further study of these genomes might lead to insights into how archaea can escape an eukaryotic hosts defences and adapt to an intracellular lifestyle.

1. van Hoek, A. H., et al. (1998). Molecular biology and evolution 15(9): 1195-1206.
2. Embley, T. M., et al. (1992). FEMS Microbiology Letters 97(1-2): 57-61.
3. Toft, C. and S. G. E. Andersson (2010). Nat Rev Genet 11(7): 465-475.

418

A molecular palaeobiological approach of arthropod terrestrialisation

Jesus Lozano-Fernandez[1], Robert Carton[2], Alastair R Tanner[1], Mark N Puttick[3], Mark Blaxter[4], Jakob Vinther[1, 3], Jørgen Olesen[5], Gonzalo Giribet[6], Gregory D Edgecombe[7], Davide Pisani[1, 3]

1. School of Biological Sciences, University of Bristol, Bristol, United Kingdom

2. Department of Biology, The National University of Ireland, Maynooth, Ireland

3. School of Earth Sciences, University of Bristol, Bristol, United Kingdom

4. Institute of Evolutionary Biology, University of Edinburgh, Edinburgh, United Kingdom

5. Natural History Museum of Denmark, University of Copenhagen, Copenhagen, Denmark

6. Museum of Comparative Zoology, Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA, United States of America

7. Department of Earth Sciences, The Natural History Museum, London, United Kingdom

Animals have marine origins, and only few animal phyla include lineages that can complete every phase of their life cycle outside the water. The process through which animals adapted to life on land is referred to as terrestrialisation and it is one of the most extreme cases of adaptation. Arthropoda (insects, spiders, centipedes and their allies) represent the largest majority of terrestrial biodiversity and have an extensive and rich fossil record that suggests they were the first fully terrestrial animals. Arthropods colonised the land multiple times independently, which allow rigorous comparison of the alternative solutions adopted by the different (but genomically and morphological comparable) groups to the same adaptive challenge.

In this study we implemented a molecular palaeobiological approach, merging molecular and fossil evidence, to elucidate the deepest history of the terrestrial arthropods. We focused on the three, independent, Palaeozoic arthropod terrestrialisation events (those of Myriapoda, Hexapoda and Arachnida) and showed that a marine route to the colonisation of land is the most likely scenario. Molecular clock analyses confirmed an origin for the three terrestrial lineages bracketed between the Cambrian and the Silurian. While molecular divergence times for Arachnida are consistent with the fossil record, Myriapoda and Hexapoda are inferred to have colonised land earlier. An estimated origin of myriapods by the early Cambrian substantially predates trace or body fossil evidence, precedes the appearance of embryophytes and raise the possibility of independent terrestrialisation events in crown-group myriapod lineages, consistent with morphological arguments for convergence in tracheal systems.

419

Diversity in gene expression in cell populations in relation to the selectionism view of tumor evolution

Chunyan Li[1], Yali Hou[1], Jin Xu[1], Aiqun Zhang[2], Jiahong Dong[2, 3], Chung-I Wu[1, 4], Xuemei Lu[1]

1. Beijing Institute of Genomics, Chinese Academy of Sciences, Chao Yang, Beijing 100101, China

2. Hospital and Institute of Hepatobiliary Surgery, Chinese PLA General Hospital, Beijng 100853, China

3. Department of Hepatopancreatobiliary Surgery, Beijing Tsinghua Changgung Hospital, Medical Center, Tsinghua University, Beijing, China

4. Department of Ecology and Evolution, University of Chicago, Chicago IL, USA

While intratumor diversity driven by selection has been the prevailing view in cancer biology, pervasive selection generally does not lead to the high genetic diversity observed within a single tumor. Indeed, recent population genetic analyses of intratumor diversity have been unable to reject the neutral interpretation and it seems necessary to test the selectionism view directly. Here, we utilized gene expression data as surrogate for functional significance in intra- and inter-tumor comparisons. Selectively-driven expression divergence is expected to be higher than neutral expression divergence. Instead, we observed little expression differentiation among samples of the same tumors, which is even lower than the expression differences among normal samples, the latter being the baseline level of neutral divergence. To further test the hypothesis of neutral evolution, we selected a tumor that is unusually diverse in both nucleotide variation and chromosomal alteration for detailed studies. This case

enables us to calibrate the level of expression divergence against that of genetic divergence. We observe that intratumor divergence in gene expression profile lags far behind genetic divergence, indicating insufficient phenotypic differences for selection to operate. All these analyses suggest that natural selection does not operate effectively within tumors. The evolutionary as well as medical implications are discussed.

420

Phylogenies derived from somatic mutations agree with physical topologies in *Eucalyptus*

Adam J Orr[1], Robert Lanfear[2], Reed A Cartwright[1]

1. School of Life Sciences, Arizona State University, Tempe, Arizona, United States

2. Department of Biological Sciences, Macquarie University, Sydney, Australia

Eucalyptus melliodora, a tree native to eastern Australia, has strong timber and a high nectar load, making it economically important and a vital source of food for nectar-consuming species. In 1993, an individual of the species was discovered that harbors a somatic mutation conferring herbivore resistance to a number of branches on the tree via differential terpenoid production. Transcriptomic analysis has provided limited determination of the genetic source of this variation. We generated whole-genome Illumina sequences for 24 samples from this tree (3 samples from 8 branches), for a total of 700x coverage. We called variants using both a reference-free De-Bruijn variant caller DiscoSNP++ and GATK using the high-quality reference for Eucalyptus grandis, a closely-related species. We find that the phylogeny of the variants identified by both methods reflects the branching pattern of the tree, though the phylogeny is affected by short interior nodes. We also discuss the challenges of estimating phylogenies for somatic tissues. This data presents an opportunity to study the rate and processes of somatic mutations in plants, with connections to cancer and the evolution and maintenance of multicelluarity.

421

Evolution of olfactory receptor gene family in vertebrates

Hardip R Patel[1], Denis O'Meally[2], Paul Waters[3], Simon Easteal[4]

1. Australian National University, Canberra, ACT, Australia

2. University of Sydney, Sydney, NSW

3. University of New South Wales, Sydney, NSW, Australia

4. The Australian National University, Acton, ACT, Australia

Olfactory receptor (OR) genes belong to one of the largest vertebrate gene families, but the number of OR genes can vary greatly among species. Elephant sharks, for instance, have only one functional OR gene, whereas African elephants have ~2,000. The evolutionary processes that gave rise to variation in the number OR genes have been studied extensively with inconclusive results. Analysis of copy number variation in humans indicates that evolution of OR genes can largely be explained by neutral process such as “genomic” drift and random birth-and-death of OR genes. In contrast, broader comparative studies have found that vertebrate OR gene repertoires reflect the properties of ecological niches and anatomical features rather than phylogenetic relationships and that they are therefore adaptive. Systematic evolutionary study of OR genes in vertebrates is limited by: ***(a)*** lack of a unified framework to accurately identify OR genes in vertebrate genomes, ***(b)*** inconsistent estimates of the number of OR genes for the same species, ***(c)*** inclusion of non-OR G-protein coupled receptor sequences (false positives) in analyses, and ***(d)*** lack of transparency in reporting genomic co-ordinates for sequences analysed. We have created a framework for systematic evolutionary analysis of vertebrate OR genes that addresses some of these deficiencies. We have also identified ~100,000 OR genes present in all 250 vertebrate species represented in the NCBI database. The analytical framework we have developed and this extensive sample of genes from diverse genomes enable systematic investigation of the mode and tempo of OR gene family evolution and of the evolutionary processes responsible for vertebrate OR repertoires. Using these resources, I will present results of our analysis of OR gene sub-family expansion and contraction across the vertebrates.

422

TreeFam and Ensembl: Phylogenetic resources

Mateus Patricio[1], Matthieu Muffato[1], Bronwen Aken[1], Paul Flicek[1]

1. European Molecular Biology Laboratory, European Bioinformatics Institute, Cambridge, Cambridgeshire, United Kingdom

The Ensembl and Ensembl Genomes projects create and distribute genome annotations for a wide range of genomes, including model organisms. The number of publicly available genomes is increasingly rapidly, providing an opportunity for new insights via comparative genomics. TreeFam produces phylogenetic trees and orthology predictions, though previously only for metazoans. Here we describe advances in TreeFam, targeted to achieve scalability to all Ensembl eukaryotes.

The key component is a new library of HMM (Hidden Markov Model) profiles that was created from Panther and TreeFam, with custom profiles to fill gaps in gene coverage. The library represents gene families across all eukaryotes.

We have designed a new workflow that uses this library to classify protein sequences from thousands of genomes into families in a quick and robust manner. The workflow’s full-build mode generates phylogenetic trees and orthologies anew across all species. The faster ‘update’ mode inserts data from new species or new gene annotations into the existing phylogenetic trees and orthologies.

The first step in this new workflow is to match incoming protein sequences to our library of gene families. For each family, we then create a multiple sequence alignment which is used to infer the best amino-acid replacement model and to reconstruct a phylogenetic tree. Each phylogenetic tree is reconciled with a species tree in order to infer consistent homology relationships following the speciation and duplication events reported. In update mode, only those alignments with new protein data are recomputed. The whole workflow is fully automated using eHive, our standard pipeline management system.

Gene families produced by TreeFam’s new workflow were released for vertebrates in Ensembl 84 (March 2016). The results can be viewed on our website at www.ensembl.org.

423

Sponges, not the comb jellies, are the sister group of all the other animals: further evidence from the analyses of compositional biases

Davide Pisani[1]

1. University of Bristol, Bristol, N/A, United Kingdom

There is significant disagreement on the phylogenetic relationships at the root of the animal tree of life. While initial analyses of genomic-scale datasets associated with the publication of the first two, completely sequenced, comb jellies (phylum Ctenophora) genomes (Ryan et al. 2013; Moroz et al. 2014; and Whelan et al. 2015) suggested that this lineage might represent the sister group of all the other animals, further re-analyses of the datasets associated with these studies (Pisani et al. 2015) showed that a position of the ctenophores at the root of the animal tree is most likely a tree reconstruction artifact. The results of Pisani and collaborators sparked a heated debate (e.g. Halanych et al. 2016 Vs Pisani et al. 2016) with Halanych et al. (2016) suggesting that, contrary to Pisani et al. (2015), the placement of the sponges at the root of the animal tree might represent a compositional attraction between Silicean sponges and the outgroups. Here, we will present results of reanalyses of all recently published datasets (Ryan et al. 2013; Whelan et al. 2015; Chang et al. 2015; Cannon et al. 2016) that bears on this problem and show that contrary to Halanych et al. (2016) correcting for compositional heterogeneity invariably strenghten support for sponges as the sister group of all the other animals. This strongly confirms the results of Pisani et al. (2015). We shall use these results to illustrate how, phylogenetic patterns are best extracted and distinguished from non-phylogenetic noise in the analyses of genomic scale datasets.

1. Cannon JT, Vellutini BC, Smith J 3rd, Ronquist F, Jondelius U, Hejnol A. Xenacoelomorpha is the sister group to Nephrozoa. Nature. 2016 Feb 4;530(7588):89-93. doi: 10.1038/nature16520.
2. Chang ES, Neuhof M, Rubinstein ND, Diamant A, Philippe H, Huchon D, Cartwright P. Genomic insights into the evolutionary origin of Myxozoa within Cnidaria. Proc Natl Acad Sci U S A. 2015 Dec 1;112(48):14912-7. doi: 10.1073/pnas.1511468112.
3. Halanych KM, Whelan NV, Kocot KM, Kohn AB, Moroz LL. Miscues misplace sponges. Proc Natl Acad Sci U S A. 2016 Feb 23;113(8):E946-7. doi:10.1073/pnas.1525332113.
4. Moroz LL, ey al. The ctenophore genome and the evolutionary origins of neural systems. Nature. 2014 Jun 5;510(7503):109-14. doi:10.1038/nature13400.
5. Pisani D, Pett W, Dohrmann M, Feuda R, Rota-Stabelli O, Philippe H, Lartillot N, Wörheide G. Genomic data do not support comb jellies as the sister group to all other animals. Proc Natl Acad Sci U S A. 2015 Dec 15;112(50):15402-7. doi:10.1073/pnas.1518127112.
6. Pisani D, Pett W, Dohrmann M, Feuda R, Rota-Stabelli O, Philippe H, Lartillot N, Wörheide G. Reply to Halanych et al.: Ctenophore misplacement is corroborated by independent datasets. Proc Natl Acad Sci U S A. 2016 Feb 23;113(8):E948-9.doi: 10.1073/pnas.1525718113.
7. Ryan JF, et al. The genome of the ctenophore Mnemiopsis leidyi and its implications for cell type evolution. Science. 2013 Dec 13;342(6164):1242592. doi: 10.1126/science.1242592.
8. Whelan NV, Kocot KM, Moroz LL, Halanych KM. Error, signal, and the placement of Ctenophora sister to all other animals. Proc Natl Acad Sci U S A. 2015 May 5;112(18):5773-8. doi: 10.1073/pnas.1503453112.

424

Chromosomal speciation in rock-wallabies: linking population processes to evolutionary processes across the phylogeny

Sally Potter[1, 2], **Jason G Bragg**[1, 3], **Craig Moritz**[1], **Janine E Deakin**[4], **Mark DB Eldridge**[2]

1. The Australian National University, Acton, ACT, Australia

2. Australian Museum Research Institute, Sydney, NSW, Australia

3. Royal Botanical Garden, Sydney, NSW, Australia

4. Institute for Applied Ecology, University of Canberra, Bruce, ACT, Australia

Australian rock-wallabies have extensive chromosomal rearrangements across the genus (*Petrogale*). Chromosomal rearrangements are known to cause recombination suppression in the genome and thus create reproductive barriers and drive speciation. Whether this is a cause or consequence is often difficult to decipher. Using a targeted exon capture approach, we resolve the phylogenetic relationships across *Petrogale* to understand how chromosomal variation has evolved within this genus. In addition, we focus more detailed sampling across three parapatric species from Queensland to understand the links between chromosomal rearrangements, gene flow and speciation. Recent results indicate high levels of gene flow despite complex chromosomal rearrangements. We explore how gene flow patterns may change across the genome in relation to chromosome rearrangements. This is an exciting Australian system to explore the relationship between genome divergence and chromosomal rearrangements in speciation.

425

Lineage specific expansions in detoxification-related and transporter gene families and sub-classes are strongly associated with polyphagy in herbivorous insects

Rahul Rane[1, 2], **Tom Walsh**[2], **Lars Jermiin**[2], **Stephen Pearce**[2], **Gordon Karl**[2], **Richards Stephen**[3], **Lee Ronald**[1], **Ary Anthony Hoffmann**[1], **John Oakeshott**[2]

1. University of Melbourne, Carlton, VIC, Australia

2. CSIRO, Canberra, ACT, Australia

3. Baylor College of Medicine, Houston, TX, USA

Abstract:

Host polyphagy in insects is hypothesized to be associated with a greater gene repertoire in gene families directly or indirectly involved in xenobiotic detoxification or sensing. Here we test this hypothesis by focussing on 58 species and nine such gene families which includes the highest numbers of genes ever reported in insects for the eight families. To do so we created a novel orthology pipeline that was able to sensitively identify increased conservation as tested in *Drosophila* species while identifying lineage specific gene expansions across the genome. The workflow to study the 58 species was therefore based on duplication and orthology sensitive characterisation of gene families. We found that polyphagous species had more genes and especially duplications when compared with restricted specialists from the same insect order. Species feeding on leaf tissues and seeds which were expected to represent the most toxic diets had particularly broad gene repertoires. These patterns were mainly driven by P450, CCE, ABC transporter and OBP gene families and their detoxification sub-classes, with OBPs being present in lower numbers in the five most aggressive polyphagous pest species considered in the study. The data suggest a strong biological signal for evolution of polyphagy and/or tolerance of toxic diets.

426

Phylogenetic relationships of the three major teleost lineages

Vydianathan Ravi[1], Chao Bian[2], Yinchang Hu[3], Xidong Mu[3], Xueyan Shen[4], Jia Li[2], László Orbán[4], Byrappa Venkatesh[1], Qiong Shi[2]

*1. Institute of Molecular and Cell Biology (IMCB), A*STAR, Singapore, Singapore, SINGAPORE*

2. BGI-Shenzhen, Shenzhen, China

3. Key Laboratory of Tropical & Subtropical Fishery Resource Application & Cultivation, Ministry of Agriculture, Pearl River Fisheries Research Institute, Chinese Academy of Fishery Sciences, Guangzhou, China

4. Reproductive Genomics Group, Temasek Life Sciences Laboratory, Singapore

Teleost fishes are the largest and most diverse group of extant vertebrates. They are divided into three major lineages - the Osteoglossomorpha (bonytongues and mooneyes), Elopomorpha (eels and relatives) and Clupeocephala (all the remaining teleosts). The phylogenetic relationship between these three groups has not been resolved satisfactorily. Mitogenome-based phylogenetic studies have placed Osteoglossomorpha as the most ancestral teleost group, with Elopomorpha and Clupeocephala forming a monophyletic group. On the other hand, recent studies based on nuclear genes as well as ultraconserved elements have suggested Elopomorpha as the most ancestral teleost group. We have sequenced the whole-genome of an Osteoglossomorph, the Asian arowana (*Scleropages formosus*) which has a 780 Mb genome and predicted 22,016 protein-coding genes. We carried out phylogenomic analysis using a genome-wide set of one-to-one orthologues obtained from arowana, several other teleosts, spotted gar and coelacanth. Our analysis of this large set of protein sequences supported a sister group relationship between Osteoglossomorpha and Elopomorpha, with Clupeocephala constituting an outgroup, in contrast to the relationships proposed by previous analyses based on smaller datasets.

427

Patterns of gene duplication and loss in the mammalian lineage mirror association with pathogenic CNVs in human

Alan M Rice[1], Aoife McLysaght[1]

1. Trinity College Dublin, Dublin, DUBLIN, Ireland

Copy number variations (CNVs) account for genome variation an order of magnitude larger than single nucleotide polymorphisms. While much of this variation has no phenotypic consequences, some variants have been associated with neurodevelopmental disorders including autism, epilepsy, intellectual disability, and schizophrenia. Copy number changes of specific dosage-sensitive genes may be causative of this pathogenicity. To understand these disorders there is a need to identity causative genes, usually performed through case-control association studies. However, identifying specific causative copy number changes remains difficult given study size requirements to achieve the necessary power. Here we show that patterns of gene duplication and loss in the mammalian lineage are associated with human CNV pathogenicity. We found that pathogenic copy number alterations are significantly enriched for genes involved in development and that genes found on pathogenic CNVs have greater copy number conservation across mammals. Conversely, genes found in benign CNV regions have more variable copy number across the tested species, showing greater duplication frequency and more missing orthologs. These results demonstrate that population CNV trends translate to a reciprocal evolutionary pattern where pathogenic variations in copy number are sufficiently deleterious to be selected against. Whether specific genes are dosage-sensitive can be predicted by characteristic evolutionary patterns and hallmarks of selection. We anticipate that these evolutionary metrics will provide insights to regions of the genome of currently-unknown clinical significance. Furthermore, well-established pathogenic regions can be further refined at gene resolution.

428

Building an evolutionary framework for the functional characterisation of an arthropod-specific gene family

Jack Scanlan[1], Lars S Jermiin[2], Charles Robin[1]

1. University of Melbourne, Parkville, Vic, Australia

2. Land and Water, CSIRO , Canberra, ACT, Australia

Discovering the functions of uncharacterised genes is a major challenge in modern biology, but it is rarely undertaken within a robust evolutionary framework. EcKinases, a largely arthropod-specific family of small-molecule kinases, are implicated in the regulation of ecdysteroids in reproduction and development, but some lines of evidence suggest they detoxify phytoecdysteroids in the diets of crop pest insects, and they may act as a novel, non-canonical family of Phase II detoxification enzymes. However, this family is broadly uncharacterised, and its evolutionary history has not been studied. To address this, we have manually annotated EcKinases in the genomes of over 60 species of arthropods and used Bayesian phylogenetic methodologies to reconstruct their evolution phylum-wide. EcKinases were found in the genomes of hexapods and crustaceans, but not chelicerates and myriapods. Gene content per genome ranges from 12 (in many bee species) to 104 (in the German cockroach). Some taxa, like bees, exhibit remarkable stability in gene content, while others, like the *Drosophila* genus and related Dipteran taxa, have experienced rapid gene gain and loss. Polyphagous Lepidopteran species have undergone independent gene blooms in clades predicted to be ecdysteroid kinases, suggesting they may play a role in phytochemical detoxification. Various single-copy clades are highly conserved in insects, pointing to roles in biological processes fundamental to this taxon. These results and ongoing work are providing an informative framework for characterising EcKinase gene function across the arthropod phylum.

429

Characterizing Sex-Biased Gene Expression in the Green Anole

Shawn Rupp[1, 2], Melissa Wilson Sayres[2]

1. Arizona State University, Chandler, ARIZONA, United States Minor Outlying Islands

2. School of Life Sciences, Arizona State University, Tempe, Arizona, United States

In species with highly heteromorphic sex chromosomes, the degradation of one of the sex chromosomes will result in unequal gene expression between the sexes and between the sex chromosomes and the autosomes. Dosage compensation is a process whereby genes on male and female sex chromosomes achieve equal gene expression. We compared levels of transcription between males and females, and between the X

chromosome and the autosomes in the green anole, Anolis carolinensis. Our results from the green anole build on data from mammals, fruit flies, and nematodes, supporting the hypothesis that chromosome-wide dosage compensation may be a universal feature of XX/XY sex determination systems.

430

Experimental Analysis of GC Content Evolution in Bacteria

Mrudula Sane[1], Deepa Agashe[1]

1. National Centre for Biological Sciences, Tata Institute of Fundamental Reasearch, Bangalore, KARNATAKA, India

Bacterial genome GC content varies from ~13% to 75% GC. Understanding the source and maintenance of this variation is important for understanding the evolution of codon usage bias and protein-coding sequences. Both neutral processes (like mutation) and selective processes (environmental selection) are thought to shape GC content. One environmental factor, nitrogen limitation, is thought to select for AT-rich genomes. We hypothesized that if this is true, over evolutionary time scales genomes with an underlying GC->AT mutational bias would have a selective advantage over the wild-type under nitrogen-limitation. Our data suggest that under nitrogen-limitation, mutants of *Escherichia coli* with altered mutational biases vary in initial fitness, suggesting that under selection, these mutants would have different probablilities of becoming fixed in the population. We find that a mutant with a GC->AT bias does not have a competitive selective advantage over WT, but a mutant with an AT->GC bias has 30% higher fitness compared to the WT in nitrogen limitation, in a 24-hour competition. We find that the GC->AT biased mutant has a selective advantage over WT in rich media. This suggests that under nitrogen limitation, a GC->AT biased mutant would only rise to high frequencies under genetic drift, and we are currently testing whether this trend continues over longer timescales. We are addressing this question by evolving mutants with biased mutational spectra under various environmental conditions implicated in GC content evolution and asking whether the spectrum of mutations fixed under selection is different from that fixed under drift.

431

Rapid evolutionary response to Tasmanian devil facial tumor disease

Andrew Storfer[1], Brendan Epstein[1], Rodrigo Hamede[2], Menna Jones[2], Barbara Schonfeld[2], Sarah Hendricks[3], Cody Wiench[3], Elizabeth Murchison[4], Paul Hohenlohe[3]

1. Washington State University, Pullman, WA, United States

2. Biological Sciences, University of Tasmania, Hobart, TAS, Australia

3. Biological Sciences, University of Idaho, Moscow, ID, USA

4. Kings College, Cambridge University, Cambridge , England, UK

Cancer poses one of the greatest human health challenges we face today, and although rare, cancers have been transmitted from mother to fetus or via organ transplants in immune suppressed patients. In Tasmanian devils (*Sarcophilus harrisii*), however, a recently emerged transmissible cancer is nearly 100% fatal and nearly all populations are infected. Devil facial tumor disease has resulted in localized declines exceeding 90% and an overall species decline of approximately 80% in less than 20 years. Because transmission of devil facial tumor disease is density-independent, disease-induced extinction has been predicted based on epidemiological models. However, long-diseased populations have persisted, raising the possibility of resistance evolution. Here, we report the first genomic evidence of a rapid evolutionary response to strong selection imposed by devil facial tumor disease, and such a response has rarely, if ever been documented in wild populations. We identify two genomic regions that exhibit concordant signatures of selection across three populations, including large, parallel allele frequency changes from before infection to only a few generations after infection. Both of these regions contain genes related to immune function or cancer risk in other mammals. Devil facial tumor disease spreads between hosts by suppressing or evading the immune system, and our results suggest that hosts are also evolving immune-modulated resistance that aid in species persistence in the face of this devastating disease.

432

Whole-genome sequence of the Tibetan frog *Nanorana parkeri* and the comparative evolution of tetrapod genomes

Yanbo Sun[1], Zijun Xiong[2], Xueyan Xiang[2], Jing Che[1], Yaping Zhang[1]

1. Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, YUNNAN, China

2. BGI-Shenzhen, Shenzhen, Guangdong, China

The development of efficient sequencing techniques has resulted in large numbers of genomes being available for evolutionary studies. However, only one genome is available for all amphibians, that of *Xenopus tropicalis*, which is distantly related from the majority of frogs. More than 96% of frogs belong to the Neobatrachia and no genome exists for this group. This dearth of amphibian genomes greatly restricts genomic studies of amphibians and, more generally, our understanding of tetrapod genome evolution. To fill this gap, we provide the *de novo* genome of a Tibetan Plateau frog,*Nanorana parkeri*, and compare it to that of *X. tropicalis* and other vertebrates. This genome encodes more than 20,000 protein-coding genes, a number similar to that of *Xenopus*. Although the genome size of *Nanorana* is considerably larger than that of *Xenopus* (2.3 vs 1.5 Gb), most of the difference is due to the respective number of transposable elements in the two genomes. The two frogs exhibit considerable conserved whole-genome synteny despite having diverged about 266 Ma, indicating a slow rate of DNA structural evolution in anurans. Multi-genome synteny blocks further show that amphibians have fewer inter-chromosomal rearrangements than mammals but have a comparable rate of intra-chromosomal rearrangements.With the new genome used as reference, multi-tissue transcriptomes of five *Nanorana* species with much different habitat altitudes were further generated, and genes underlying the *N. parkeri*'s adaptations to extreme environments were identified with comparative analysis methods. The new genome offers an improved understanding of evolution of tetrapod genomes, and also provides a genomic reference for other evolutionary studies.

433

The impact of dosage sensitive gene families on plant evolution

Setareh Tasdighian[1, 2], Steven Maere[1, 2], Yves Van de Peer[1, 2]

1. Plant systems biology, VIB, Ghent, Belgium

2. Biochemistry and Biotechnology, University of Ghent, Ghent, Belgium

Whole genome duplications (WGDs) are believed to play a major role in angiosperm evolution. Previous studies have found that some functional categories of genes, including regulatory and developmental categories, expand almost exclusively through genome duplication, likely because their expansion through small-scale duplications is counteracted by dosage balance effects. However, the duplication dynamics of individual gene families have not been studied in detail. We developed a stochastic birth-death model to study the size evolution of gene families across a species phylogeny, taking into account both small-scale and large-scale duplication events. We use this model on a set of angiosperm species with known WGD history to assess the dosage balance sensitivity of individual gene families, and we interpret the results in the context of the potential impact of WGDs on evolutionary innovation in angiosperms.

434

The number of prolactin cleavage sites generating vasoinhibins varies in primates

Jakob Triebel[1], Daniel Rios-Barrera[2], Juan Pablo Robles[3], Manuel Aguilar[3], Gonzalo Martínez de la Escalera[3], Carmen Clapp[3], Thomas Bertsch[1]

1. Institute for Clinical Chemistry, Laboratory Medicine and Transfusion Medicine, Paracelsus Medical University, Nürnberg, Bavaria, Germany

2. Directors' Research, Developmental Biology Unit, European Molecular Biology Laboratory, Heidelberg, Baden-Würtemberg, Germany

3. Instituto de Neurobiología, Universidad Nacional Autónoma de México - Campus UNAM-Juriquilla, Querétaro, México

BACKGROUND

The prolactin/vasoinhibin axis in humans features the generation, secretion, and actions of the pleiotropic pituitary hormones prolactin and vasoinhibins under control of the hypothalamus, the pituitary gland, and the target tissue microenvironment. Vasoinhibins are generated by proteolytic cleavage of full-length prolactin at various cleavage sites. The evolutionary history of vasoinhibins is largely unknown.

METHODS

The prolactin protein sequences of the primate species Human, Chimpanzee, Gorilla, Orangutan, Gibbon, Vervet AGM, Olive baboon, Macaque, Marmoset, Tarsier, Bushbaby, and Mouse Lemur, and the prolactin gene tree were retrieved from the ENSEMBL data base. A multiple sequence alignment was performed, using Clustal Omega. Five known cleavage sites within the human prolactin protein sequence, defined by fully conserved sequence motifs required for the generation of vasoinhibins with molecular masses of 15, 16.8, 17.2, 17.7, and 18 kilodalton (kDa), were the focus of the comparison.

RESULTS

The prolactin protein sequence of all hominoidea (Human, Chimpanzee, Gorilla, Orangutan, and Gibbon) demonstrated the presence of all 5 cleavage sites present in the human sequence. Species from the taxon Old world monkeys (Vervet AGM, Olive baboon, Macaque) lack the cleavage site for the 17.2 kDa vasoinhibin. The marmoset (Simians) prolactin sequence does not feature the 16.8 kDa vasoinhibin cleavage site, the Tarsier (Dry nose primates) lacks the 16.8, and the 15 kDa cleavage sites, the Bushbaby the 15, 16.8, and the 17.7 kDa cleavage sites, and the Mouse Lemur (Wet nose lemurs) the 15 and 16.8 kDa cleavage sites.

CONCLUSION

The variation in the number of cleavage sites likely translates into a corresponding difference in the number of vasoinhibins present in the respective species. The ascending number of cleavage sites throughout primate evolution may represent prolactin gain-of-function events and constitutes an endocrinological distinctive feature between primates, which could affect the pleiotropic profile of biological effects of the prolactin/vasoinhibin axis.

435

Evolution of the CRKs and other DUF26-containing gene families in plants

Aleksia Vaattovaara[1], Chunxiang Li[2], Ari Löytynoja[2], Jarkko Salojärvi[1], Jaakko Kangasjärvi[1], Michael Wrzaczek[1]

1. Division of Plant Biology, Department of Biosciences, Viikki Plant Science Centre, University of Helsinki, Helsinki, Finland

2. Institute of Biotechnology, University of Helsinki, Helsinki, Finland

Linking gene family expansions to the functional evolution of proteins is an interesting challenge in evolutionary biology. Plants contain a large number of receptor-like protein kinases (RLKs) to be able to sense and response to changes in their environment. The cysteine-rich receptor-like protein kinases (CRKs) are distinguished from the other RLKs based on the structure of their extracellular region which contains DUF26 (domain of unknown function 26; also known as stress-antifung domain, PF01657) domains. DUF26 domains are also found in two closely related gene families, the plasmodesmata-localized proteins (PDLPs) and the cysteine-rich receptor-like secreted proteins (CRSPs). The DUF26 domain is plant-specific and contains the conserved cysteine motif C-8X-C-2X-C.

In order to understand their evolution, we have identified and manually curated gene models for CRKs, PDLPs and CRSPs from more than 20 plant species covering most plant lineages. We can identify genes with DUF26 domain from land plants but not from the sequenced algae species. Our main interest is to understand why so many genes in these gene families are maintained after duplication events and why different phylogenetic subgroups have expanded in different plant lineages. Our data suggests that genes containing two DUF26 domains appeared for the first time in the lycophytes. Intriguingly, in genes with two DUF26 domains, the first and the second DUF26 domain have differentiated into specific forms with unique sequence context surrounding the conserved cysteines. There is also considerable variation within DUF26 domains between the different phylogenetic subgroups of the CRKs in *Arabidopsis*. This variation might have functional and structural importance for the extracellular domain of the CRKs.

436

Genome-wide analysis of bitter taste receptor genes in birds and the genomic basis of adaptation to a microclimate contrast in the spiny mouse

Kai Wang[1], Kexin Li[2], Eviatar Nevo[2], Huabin Zhao[1]

1. Wuhan University, Wuhan, HUBEI, China

2. University of Haifa , Haifa, Israel

I am Kai Wang, a graduate student from College of Life Sciences at Wuhan University in China. I am applying for the "Young Investigator Travel Award" of SMBE 2016. So far I have been involved in two projects: one project was finished and the other is still ongoing.

The first project was published in GBE last year, which was entitled "Birds generally carry a small repertoire of bitter taste receptor genes". Using recently released 48 avian genomes, we characterized the evolution of avian bitter taste receptor genes (*Tas2r*s) and tested whether dietary toxins have shaped the repertoire size of avian *Tas2r*s. Our analyses appear to support that herbivorous and insectivorous birds demand more functional *Tas2r*s than carnivorous birds feeding on noninsect animals, and highlight the critical role of taste perception in birds.

The second project is to uncover the genomic basis of adaptation in the spiny mouse (*Acomys cahirinus*) to a microclimate contrast in the two opposite slopes of "Evolution Canyon" in Israel, where show strong abiotic contrasts including differences in temperature and humidity. We sequenced and annotated the draft genome of the spiny mouse, and resequenced the whole genomes of several individuals from both slopes, aiming to examine patterns of genetic differentiation at the genome-wide level between these two recently diverged populations inhabiting the two slopes.

437

Evolutionally conserved mechanisms of regeneration in chordates: Uncovering signaling pathways required for WBR in *Botrylloides leachi.*

Lisa Zondag[1], Simon Blanchoud[1], Kim Rutherford[1], Neil Gemmell[1], Megan J Wilson[1]

1. Department of Anatomy, University of Otago, Dunedin, Otago, New Zealand

Regenerative capacity differs greatly across organisms and the ability to regenerate declines as morphological complexity increases. Within the chordate phylum, vertebrate animals exhibits a very limited regenerative potential, however, the sea squirt *Botrylloides leachi* is a chordate with a remarkable ability to undergo whole body regeneration (WBR). A fully functional adult organism (zooid) can regenerate from a minuscule piece of vascular tissue within only 8 days. In order to compare the molecular mechanisms underlying WBR in a chordate to the regeneration process in other animals, we have analysed the transcriptome of *B. leachi* at each stage of regeneration, in addition to sequencing the genome.

Genomic analysis identified signaling factor families that had expanded in the colonial sea squirt lineage but not solitary sea squirts, which have a limited ability to regenerate. Following *de novo* transcriptome assembly (6 transcriptomes in total), differential expression analysis was performed to identify genes up or/and down-regulated during WBR. Differential gene expression analysis indicates that both wound healing and an immune response are activated during early steps of regeneration in *B. leachi* WBR. These processes are also key to regeneration in response to injury in vertebrate models of regeneration such as the limb regeneation in salamanders. This suggests that chordate animals may employ a homologous series of molecular events during regeneration events wheather it be WBR or tissue regeneration.

438

Hologenomic adaptations underlying the evolution of sanguivory in the vampire bat

Marie Lisandra Zepeda Mendoza[1], Zijun Xiong[2], Anne Kathrine Runge[1], Marina Escalera-Zamudio[3], Julien Thézé[4], Jose Alfredo Samaniego Castruita[1], Daniel Streicker[5], Hannah K Frank[6], Elizabeth Loza-Rubio[7], Shengmao Liu[2], Kristine Bohman[1], Nicole White[8], Oliver A Ryder[9], Aris Katzourakis[4], Oliver G Pybus[4], Yang Li[2], Maria L Méndez-Ojeda[10], Aldo Carmona Baez[1], Shiping Liu[2], Gareth Jones[11], Alex D Greenwood[3], Mads Frost Bertelsen[12], Mike Bunce[13], Guojie Zhang[2], Thomas Sicheritz-Pontén[14], M Thomas P Gilbert[1]

1. Centre for GeoGenetics, Natural History Museum of Denmark, Denmark, Copenhagen

2. BGI–Shenzhen, Shenzhen, China

3. Leibniz Institute for Zoo and Wildlife Research (IZW), Berlin, Germany

4. Dept of Zoology, University of Oxford, Oxford

5. Institute of Biodiversity, Animal Health and Comparative Medicine, MRC-University of Glasgow Centre for Virus Research, University of Glasgow, Scotland

6. Department of Biological Sciences, Stanford University, California, USA

7. Centro Nacional de Investigación Disciplinaria en Microbiología Animal, INIFAP, Mexico

8. Australian Wildlife Forensic Services, Department of Environment and Agriculture, Curtin University, Perth, Australia

9. San Diego Zoo Institute for Conservation Research, California, USA

10. Facultad de Medicina Veterinaria y Zootecnia, Universidad Veracruzana, Veracruz, Mexico

11. School of Biological Sciences, University of Bristol, Bristol, England

12. Center for Zoo and Wild Animal Health, Copenhagen Zoo, Copenhagen, Denmark

13. Trace and Environmental DNA laboratory, Department of Environment and Agriculture, Curtin University, Perth, Australia

14. Center for Biological Sequence Analysis, Department of Systems Biology, Technical University of Denmark, Lyngby, Denmark

The order Chiroptera, bats, exhibits extremely high genetic and dietary diversity, and includes the only three obligate sanguivorous mammals, the vampire bats. Relying on blood as the sole dietary source poses significant challenges, ranging from morphological to nutritional. To study the evolution of sanguivory we used a hologenomic approach, in which we identified adaptive changes in the vampire genome as well as in its gut microbiome. To this end, we generated a high-quality reference genome (N50=26.9 Mb) for the common vampire bat, *Desmodus rotundus*, using a combination of *de novo*assembly and a Hi-C-based contiguity refinement technology. We then performed comparative genomic analyses against bats with other diets (frugivorous, insectivorous, and carnivorous), including gene selection, gene loss, and gene family expansion-contraction. We also generated metagenomics datasets by shotgun sequencing faecal samples of *D. rotundus* and bats with other diets. We identified taxa and functions that were differentially abundant or present only in the vampire bat. Our combined genomics and metagenomics results highlight how both genome and gut microbiome played key roles in the evolution of sanguivory, through affecting traits such as energy metabolism, immunity, digestive system morphogenesis, and osmotic homeostasis. The common vampire bat represents a perfect example of the study of the evolution of complex phenotypes in non-model organisms by analysing its genome and gut microbiome in a complementary fashion. Overall, our results highlight that studies not accounting for the action of both the genome and the microbiome provide incomplete insights into the evolution of complex adaptations.

439

Epigenetic variation in interspecific hybrids of the genus Arabidopsis.

Agustin L. Arce[1], Fei He[1], Gregor Schmitz[1], Juliette de Meaux[1]

1. Botanical Institute - University of Cologne, Cologne, NORD RHEIN-WESTPHALIA, Germany

Despite the major role it played in plant evolution, interspecific hybridization is generally thought to cause dramatic changes in gene regulation. This so called “genomic shock” might be mediated by epigenetic deregulation and the concomitant reactivation of transposable elements (TEs). Indeed, the epigenetic profiles reflect variation in TE content across plant genomes. Previous work in *Arabidopsis thaliana* x *Arabidopsis lyrata* interspecific hybrids showed that orthologous TEs present a preferential expression of the *A. lyrata* allele, which was associated to the reduced presence of silencing epigenetic marks.

To study the effect of interspecific hybridization on the epigenomic landscape, we used ChIP-Seq to characterize the profiles of H3K9me2 and H3K27me3 epigenetic marks in *A. thaliana, A. lyrata,* and in three hybrids between different accessions of these species. The distribution of H3K27me3silencing marks, which normally labels genes that are developmentally or environmentally regulated, coincided in all three hybrids with regions typified in *A. thaliana* as H3K27me3 rich. By contrast, two hybrids showed a distribution of TE silencing H3K9me2 marks similar to the pattern observed in the parents, while the third showed a markedly modified pattern. These results suggest the existence of within-species variation in the potential to reprogram epigenetic marks associated to TE silencing in hybrids.

440

The evolutionary optimization of polyadenylation by 3’ UTR

Wenqing Cao[1]

1. Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing, China

Polyadenylation adds a poly(A) tail to an mRNA, which is an essential step for the choice of mRNA isoforms and the regulation of mRNA level in eukaryotes. Therefore, defects in polyadenylation profoundly alter cell viability, growth, and development. Although previous studies based on evolutionary conservation identified a few elements in the 3’ UTR, the function of other sequences is still mysterious, partly because the high level of degeneracy in coding the signal of polyadenylation in the 3’ UTR. To systematically investigate the function of 3’ UTR sequences in regulating polyadenylation, we generated a library containing 3,628 yeast strains, with a variant of 3’ UTR inserted right after a GFP coding sequence in each strain. We quantified the polyadenylation efficiency for each variant by calculating the proportion of the readthrough transcripts and correlated it with sequence properties. We identified several strategies by which polyadenylation efficiencies are coded in the 3’ UTR. Particularly, guanine-rich motifs play a central role in regulating polyadenylation efficiency. We further confirmed that sequences of 3’ UTR in yeast genome are largely in agree with the coding strategies we identified from our 3’ UTR variant library. This study expands our understanding on the coding rules and evolutionary dynamics of 3’ UTR, which together pave the road for the ultimate goal of understanding every nucleotide in a genome.

441

Alignment of biological networks

Michael Charleston[1], **Alexandru Radu**[2], **Martin McGrane**[2]

1. School of Physical Sciences, University of Tasmania, Hobart, TAS, Australia

2. School of Information Technologies, University of Sydney, Sydney, NSW, Australia

Systems biology is a place of networks: gene regulatory networks, protein-protein interaction networks, food webs, social and contact networks.

Nodes of biological networks can be anything from microRNAs to genes to proteins, individuals to populations to species, and combinations thereof. The edges in the networks represent the different kinds of interactions these nodes can have with each other. These networks change over time, between species, and across ecosystems, but they often retain common features, the revelation of which can help us understand the most important aspects of the networks, and what changes may have taken place to transform one into another.

Biological networks have topological features that we can leverage to ease such comparisons; for example an abundance of certain motifs, the connectedness and degree distribution of nodes, and the overall diameter. We have developed methods (Node Fingerprinting [1] and Node Handprinting [2]) that leverage these features to create state of the art tools to align even large biological networks, with high accuracy and speed, and minimal memory. We have extended these alignment methods to estimate paths between networks, that describe plausible evolutionary routes through "network space" which can in turn provide estimates of evolutionary distance between networks – the Biological Network Edit Distance.

This poster describes our new methods, demonstrates their accuracy, and shows their potential for informing biologists interested in studying the relationships among biological networks.

1. [1] Node fingerprinting: an efficient heuristic for aligning biological networks. A Radu, M Charleston. Journal of Computational Biology 21 (10), 760-770, 2014.
2. [2] Node handprinting: A scalable and accurate algorithm for aligning multiple biological networks. A Radu, M Charleston. Journal of Computational Biology 22 (7), 687-697, 2015.
3. [3] Biological Network Edit Distance. M McGrane, M Charleston. Journal of Computational Biology, 2016 (in press).

442

The role of synonymous codon usage in regulating mRNA level explains the well-known correlation between them in genomes

Siyu Chen[1]

1. Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing, China

The phenomenon that synonymous codons are used with different frequencies (also known as codon usage bias) exists in a wide range of organisms. It has been extensively reported that codon usage bias is positively correlated with gene expression level, which has been explained by the natural selection on translational accuracy and/or efficiency to promote codon usage bias in highly expressed genes, although they are both challenged by recent studies. In this study, we investigated an opposite possibility that codon usage bias can directly regulate the mRNA level. We generated two synthetic *GFP* libraries of in total 3,556 variants that varied at 12 synonymous sites, and measured the mRNA levels of these variants. Unexpectedly, we observed that genes using a higher proportion of preferred codons tend to exhibit higher expression levels, suggesting that codon usage bias plays an important role in regulating gene expression. In other words, the correlation between codon usage bias and expression level could exist even in the absence of natural selection. Our study demonstrates the pleiotropic function of codon usage bias, provides a new explanation for the relationship between codon usage bias and gene expression level in the genome, and paves the road for synthesizing of artificial organisms.

443

Identification and Characterization of *TYR* gene in Cynomolgus Monkey (*Macaca fascicularis*)

Se-Hee Choe[1, 2], Ja-Rang Lee[1], Hyeon-Mu Cho[1, 2], Sang-Je Park[1], Young-Hyun Kim[1, 2], Sang-Rae Lee[1, 2], Kyu-Tae Chang[1, 2], Jae-Won Huh[1, 2]

1. National Primate Research Center, Korea Research Institute of Bioscience & Biotechnology (KRIBB), Cheongju, Chungbuk, Republic of Korea

2. University of Science & Technology (UST), National Primate Research Center, Korea Research Institute of Bioscience and Biotechnology (KRIBB), Cheongju, Chungbuk, Republic of Korea

Tyrosinase is copper-containing enzyme that regulates and controls melanin biosynthesis. Human *TYR* gene consists of five exons spanning about 65kb on Chromosome 11q14-q21. Mutations within the *TYR* gene could lead to the oculocutaneous albinism (OCA) due to the failure of melanin formation. Through many studies about albinism have been investigated, *TYR* gene in the non-human primate has not yet been characterized.

In this study, we identified full-length *TYR* gene in cynomolgus monkey using RACE technique. Total, two different transcript variants are identified. First type of transcript is consist of well conserved five exons and ORF compared with human. However, second type of transcript is consist of only four exons created by exon 3 skipping and following frameshift event. The result of RT-PCR showed that two transcript variants were expressed in primates including human. Second type transcript variant generated by alternative splicing mechanism with exon skipping and its transcript variant encode malfunctioning tyrosinase by frameshift. As a result, we firstly identified full-length sequence of *TYR* gene and characterized in cynomolgus monkey and other primates. Therefore, their sequences could be useful information for biological research in phenotype analysis.

444

The sequence determinants of translational readthrough constrains the evolution of 3' UTR

Xiao Chu[1]

1. Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing, China

The stop codon readthrough during translation has been widely reported and is in general deleterious unless in virus which has a compacted genome and uses it as a potential mechanism to diversify proteome. Here, taking advantage of high-throughput sequencing technology, we systematically studied the impact of sequence context on the level of readthrough. To this end, we constructed a plasmid with the expression of a fusion protein of dTomato-GFP from TDH3 promoter in *Saccharomyces cerevisiae*. A stop codon was inserted between dTomato and GFP, together with 9 bp random sequences at both sides. We dissected the readthrough into two aspects, readthrough probability and readthrough strength, which represent the proportion of cells with readthrough in the population and the average degree of readthrough in the cells with readthrough, respectively. We found that the variance of readthrough probability is much larger than that of readthrough strength, suggesting that natural selection may mainly operate on readthrough probability. Interestingly, the adenosine following stop codon leads to low readthrough probability and high readthrough strength, which causes abundant readthrough in a small fraction of cells, enlarging the cell-to-cell variation on stop codon readthrough in a cell population. We further analyzed the first nucleotide in the 3'UTR of yeast genome and found that adenosine is the most frequently used nucleotide, suggesting natural selection on translational readthrough constrains the evolution of 3' UTR. Taken together, our study distinguished two independent aspects of translational readthrough, identified the coding rules of translational readthrough in the genome, and shed light on the genomic evolution of 3'UTR in yeast.

445

Early evolution and dynamics of DNA methylation in animals

Alex de Mendoza[1], William Hatleberg[2], Technau Ulrich[3], Kevin Pang[4], Bernard M Degnan[2], Ryan Lister[1]

1. University of Western Australia, Crawley, WA, Australia

2. University of Queensland, Brisbane, Queensland, Australia

3. University of Vienna, Vienna, Austria

4. SARS Institute , Bergen, Norway

DNA methylation is a key epigenetic modification required for vertebrate development, deployed and erased in promoters and enhancer regulatory elements in a dynamic manner. Nevertheless, many of these patterns and dynamics are not observed in the sparsely methylated genomes of the few invertebrates analysed to date, where mostly gene bodies of constitutively expressed genes are methylated. To address if this stark contrast between vertebrate and invertebrate epigenomes is an ancestral feature, we have generated developmental series of whole-genome base resolution maps of cytosine methylation and hydroxymethylation for 3 representatives of the earliest branching positions of animal evolution: sponges, ctenophores and cnidarians. Despite the three species have a conserved toolkit of DNA methyltransferases and TET enzymes, the cnidarian and ctenophore methylomes are rather static, as it has previously reported for insects, and low levels of genome wide methylation are observed, concentrated in gene bodies. However, the sponge methylome is heavily methylated, showing a methylation profile similar to that of vertebrates. Highly expressed genes have unmethylated promoters, which are CpG dense, resembling vertebrate CpG Islands. Moreover, we find differentially methylated regions that change in a directional manner along development, 50% directly in gene promoters, correlating with changes in gene expression. In contrast, hydroxymethylation does not seem to mark active demethylation in the sponge genome. Overall, our analyses show a clear example of epigenome convergence between sponges and vertebrates, revealing unprecedented plasticity in animal methylomes.

446

Patterns of chromatin accessibility and transcription factor binding in human and chimpanzee pluripotent stem cells.

Irene Gallego Romero[1, 2], Shyam Gopalakrishnan[3], Yoav Gilad[4]

1. School of Biological Sciences, Nanyang Technological University, Singapore

2. Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore

3. Natural History Museum of Denmark, University of Copenhagen, Copenhagen, Denmark

4. Department of Human Genetics, University of Chicago, Chicago, IL, USA

Evolutionary change in closely related species has long been hypothesised to be driven by changes in gene regulation. To test this hypothesis, we have generated maps of genome-wide chromatin accessibility using ATAC-seq in induced pluripotent stem cell (iPSC) lines derived from 6 humans and 7 chimpanzees (*Pan troglodytes*), as well as identified patterns of transcription factor (TF) binding activity across 1261 TFs in the same cell lines. We observe that sharing of chromatin accessibility patterns between the two species is strongest near orthologous transcription start sites (orthoTSS), and decreases with increasing distance from these. However, when we combine these data with a previously published RNA-sequencing dataset from the same cell lines, we find that significant inter-species differences in chromatin accessibility near orthoTSS cannot explain the majority of differentially expressed genes between the two species. Similarly, when we focus on transcription factor binding patterns across the two species, we find that sites most likely to bound in both species tend to have high PWM scores and are located close to orthoTSS. Intriguingly, some of the transcription factors with the most divergent inter-species binding patterns have been implicated in early developmental processes, suggesting that the differences we observe at the pluripotent stage might underlie other interspecies cellular-level, and potentially even organismal-level, differences between humans and chimpanzees.

447

Estimating the distribution of effect sizes of variants affecting human gene expression

Emily C Glassberg[1], Ziyue Gao[2, 3], Jonathan K Pritchard[1, 2, 3]

1. Biology, Stanford University, Stanford, CA, USA

2. Howard Hughes Medical Institute, Chevy Chase, MD, USA

3. Genetics, Stanford University, Stanford, CA, USA

Variants affecting gene expression (eQTLs) help explain the genetic basis of human complex traits both in terms of interpreting genome-wide association studies and understanding genome regulatory architecture more generally.[1] While many human complex traits are highly polygenic, expression of each individual gene is thought to be regulated by a smaller set of variants.[(e.g. 2)] Therefore, studying the genetic architecture of gene regulation is a promising point of entry for developing mechanistic understandings of complex traits. However, due to (i) differences in statistical power between SNPs of varying minor allele frequencies and (ii) linkage between causal SNPs, it is difficult to draw conclusions about the space of potential expression-altering mutations from statistically significant eQTL signals alone. Here, we develop a likelihood-based method that utilizes findings from existing eQTL studies to estimate the proportion and effect size distribution of eQTLs. We evaluate the performance of our method using simulations and show that our inference is unbiased when causal variants are sparse, as is believed to be the case. The performance of our method decreases as the density of and linkage between causal variants increases. However, this effect stems not from our inference method, but from biases in effect size estimates introduced by the statistical methods used in eQTL mapping. We therefore employ alternative mapping strategies to reduce biases in effect size estimation and consequently improve inference. Applying this pipeline to existing human eQTL data[3,4], we estimate the distribution of effect sizes of causal variants. From this, we infer the expected neutral rate of evolution of gene expression traits. Finally, we compare these estimates across cell types and gene sets to learn about variation in human genome regulatory architecture and gene expression constraints.

1. Albert FW and Kruglyak L 2015. 'The role of regulatory variation in complex traits and disease.' Nat Rev Gen (16): 197-212.
2. Wheeler, HE et al. 2016. 'Survey of the Heritability and Sparsity of Gene Expression Traits Across Human Tissues.' bioRXiv.
3. Lappalainen, T et al 2013 'Transcriptome and genome sequencing uncovers functional variation in humans' Nature (501): 506-11.
4. GTEx Consortium 2015. 'The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans.' Science (348): 648-60.

448

Adaptive Rewiring of a Gene Regulatory Network

Jungeui Hong[1], Nathan Brandt[1], Timothy Hughes[2], Ally Yang[2], David Gresham[1]

1. New York University, New York, NEW YORK, United States

2. University of Toronto, Toronto, Canada

Understanding the molecular basis and dynamics of adaptive evolution of gene expression is a central problem in evolutionary biology. Typically, retrospective analyses are employed to infer adaptive changes in gene expression. To study the dynamics and outcome of gene expression evolution under conditions of strong selection, we performed experimental evolution of yeast cells growing in nitrogen-limited chemostats. Following several hundred generations we found significant divergence of nitrogen-responsive gene expression in lineages with increased fitness. Using high throughput sequencing we identified repeated selection of non-synonymous mutations in the zinc finger DNA binding domain of the GATA transcription factor, GAT1, an activator of the nitrogen catabolite repression (NCR) regulon. We investigated the functional effect of adaptive mutations using a combination of biochemical assays, protein binding microarrays, transcriptional reporters and mathematical modeling. We find that the functional effects of GAT1 mutations are exerted both directly, and indirectly by rewiring incoherent feed-forward loops comprising different GATA transcription factors. Using targeted ultra-deep sequencing we find that evolving populations contain multiple GAT1 mutations at low frequencies (10^{-2} -10^{-3}) during the initial stages of selection that fail to subsequently increase to appreciable frequencies due to clonal interference. Our study demonstrates that under strong selection the evolution of gene expression is highly repeatable and that adaptive changes in gene expression can result from both direct and indirect effects within the context of a gene regulatory network.

449

Translational regulation constrains the evolution of coding sequences after the start codon in *Saccharomyces cerevisiae*

Qing Huan[1]

1. Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing, China

It has been generally accepted that initiation is the rate-limiting step of translation. However, the transition between translational initiation and translational elongation may also play a role in regulating the translational efficiency of a gene, which is defined as the number of proteins produced per mRNA per unit time. In *Saccharomyces cerevisiae*, genes encoding ribosome subunits, which in general have higher translational efficiency, preferentially use 'GNN' as their second codon, suggesting that the mRNA sequence after the start codon may be important to promoter the translational efficiency of a gene. To systematically examine the function of nucleotides after the start codon, we inserted 12 random nucleotides between the start codon and other coding sequences of GFP and measured their impacts on the translational efficiency (GFP level/mRNA

level). We observed that GNN containing variants exhibited higher translational efficiency. Further, we performed Ribo-seq experiment to gauge the translational dynamics and found that knocking-out certain initiation factors may eliminate the positive impact of GNN on the translational efficiency, suggesting that GNN promotes translational efficiency by interacting with those initiation factors. Together, our results demonstrate the importance of the transition between translational initiation and translational elongation in regulating translational efficiency and shed light on the genome evolution of the coding sequence context after the start codon.

450

Genome-wide expression profiling analysis identifies aging-associated gene expression change

Jae-Won Huh[1, 2], Ja-Rang Lee[1], Se-Hee Choe[1, 2], Hyeon-Mu Cho[1, 2], Sang-Je Park[1], Young-Hyun Kim[1, 2], Sang-Rae Lee[1, 2], Kyu-Tae Chang[1, 2]

1. National Primate Research Center, Korea Research Institute of Bioscience and Biotechnology (KRIBB), Cheongju, Chungbuk, Republic of Korea

2. University of Science & Technology (UST), National Primate Research Center, Korea Research Institute of Bioscience and Biotechnology (KRIBB), Cheongju , Chungbuk, Republic of Korea

Chlorocebus aethiops is an old world monkey, has been traditionally used as an experimental model for biomedical research. In addition, recent technological advances of next generation sequencing is useful for unravel the genetic mechanisms of senescence, aging and age-related diseases. Here we analyzed the blood transcriptome of aged *C. aethiops* from nine individuals ranging in age from 17 to 25 years old over the next year using the Illumina sequencing platform. We identified 418 to 676 differentially expressed genes (DEGs) in each nine individuals which tend to increase with age. Among the DEGs, 35 genes were differentially expressed between two time points in nine individuals commonly. A major proportion of the common 35 DEGs were belonged to gene ontology categories involved in translation, translational elongation, and regulation of cellular protein metabolic process. Additionally, we added third-year blood samples, and 13 common DEG s were experimentally confirmed. This study reported the gene expression changes during aging *C. aethiops* transcriptome, and common genes identified have the potential as the particular target of aging.

451

Functional diversity of bZIP transcription factors in the sponge *Amphimedon queenslandica*: insights into the ancestral animal regulatory genome

Katia Jindrich[1], Kathrein E Roper[1], Sussan Lemon[1], Benedict Yuen[1], Sandie Degnan[1], Bernie Degnan[1]

1. School of Biological Sciences, The University of Queensland, St Lucia, QLD, Australia

What genomic innovations supported the emergence of multicellular animals, over 600 million years ago, remains one of the most fundamental questions in evolutionary biology. Increasing evidence suggests that the elaboration of the regulatory mechanisms controlling gene expression, rather than gene innovation, underlies this transition. Since transcriptional regulation is largely achieved through the binding of specific transcription factors to specific cis-regulatory DNA, understanding the early evolution of these master orchestrators is key to retracing the origin of animals (metazoans). Basic leucine zipper (bZIP) transcription factors constitute one of the most ancient and conserved families of transcriptional regulators. They play a pivotal role in multiple pathways that regulate cell decisions and behaviours in all kingdoms of life. Here, we explore the early evolution and putative roles of bZIPs in a representative of one of the oldest surviving animal groups, the marine sponge *Amphimedon queenslandica*. Phylogenetic analyses identify 17 bZIPs in *Amphimedon*, originating from a repertoire of 7 and 12 bZIPs in the metazoan and holozoan ancestor, respectively. As expected for regulatory molecules, most bZIPs display high temporal specificity, cell-type specific localization and are dynamically expressed throughout *Amphimedon* development. Specific sponge bZIPs appear to be involved in a variety of contexts, including cell fate decisions, circadian regulation and response to pathogens. Integrating these observations with ongoing ChIP-Seq experiments, we infer that many of the roles bZIPs play in bilaterians have a more ancient origin and were present in the last common ancestor of all contemporary animals.

452

Developing maps of fitness consequences for plant genomes

Zoé Joly-Lopez[1], Adrian E Platts[1, 2], Adam Siepel[2], Michael D Purugganan[1]

1. Center For Genomics And Systems Biology, New York University, New York, NY, United States

2. Simons Center for Quantitative Biology, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, USA

Predicting the fitness consequences of mutations, and their concomitant impacts on molecular and cellular function as well as organismal phenotypes, is an important challenge in biology that has new relevance in an era when both functional and population-genomic data are readily available. The ability to construct genomewide maps of fitness consequences in plant genomes is a recent development that has potential implications for our ability to predict the fitness effects of mutations and discover functional elements, especially in the non-coding regions of the genome. We present a method developed recently for the human genome that we are applying to plant genomes. This approach combines population genomics and divergence data to estimate the probability of fitness effects of mutations in classes of sites defined by a common function (e.g. TFBSs) by integrating intra-specific polymorphisms and between-species divergence data with functional genomic information. Its foundation relies on a statistical method called **I**nference **N**atural **S**election from **I**nterspersed **G**enomically co**H**erent elemen**T**s (INSIGHT), which is conceptually similar to population genetics methods that use patterns of polymorphism and divergence to identify departures from neutral expectations. The contrast between polymorphism and divergence is a powerful approach to inferring recent selection and the INSIGHT approach to pooling dispersed sites enables the characterization of noncoding elements that may have been subject to recent selection. The adaptation of this approach to plant genomes has confirmed major differences by which we can functionally partition the genomes of *Arabidopsis thaliana* and *Oryza sativa*. Maps of fitness consequences in plants, combined with traditional genetic approaches, could accelerate discovery of functional elements such as regulatory sequences in non-coding DNA and genetic polymorphisms associated with key traits, including agronomically-important traits.

453

Instability of 5’ overlaps between protein-coding genes

Izabela Makałowska[1], Wojciech Rosikiewicz[1], Yutaka Suzuki[2]

1. Institute of Molecular Biology and Biotechnology, Adam Mickiewicz University, Poznań, Poland

2. Laboratory of Functional Genomics, The University of Tokyo, Tokyo, Japan

Numerous studies showed that occurrence of sense-antisense gene pairs is quite common in human and other genomes. Such pairs are often formed by two protein-coding genes however, with the discovery of antisense lncRNAs, the interest in sense-antisense overlaps between protein-coding genes significantly decreased. Nevertheless, considering recent advances in sequencing technologies we reinvestigated the phenomenon of this very specific case of antisense transcription. In our studies we utilized TSS-seq data from 73 human and 13 mouse libraries and performed multilevel analysis to examine overlap conservation across species, various tissues, experimental conditions and individuals. We identified 592 human and 113 mouse gene pairs that overlap at their 5' end in at least one library. None of them overlapped in all libraries and majority utilized overlapping TSSs in selected tissues/cell lines only. Our results show that 5' overlap between two genes is highly instable. This instability is not just a simple consequence of tissue specific factors. The switch between two states, non-overlapping and overlapping, very often reflects the response of the cell to changing environmental conditions. We identified, for example, 15 gene pairs that shifted from non-overlapping to overlapping state in reaction to a transfection. Moreover, analysis of 26 adenocarcinoma cell lines showed that overlap between two genes might not be conserved even in cell lines of the same type and cultured under the same conditions. In addition, we showed that overlap between genes does not have a negative effect on their expression. On the contrary, genes have higher expression level in tissues/cell lines in which they overlap while compared with the expression in cell lines where both genes are expressed but do not overlap. However, as our results show, this elevated expression is not directly related to the fact that genes overlap but is resulting from the acquisition of additional promoters.

454

Sequence uniqueness determines the accuracy of isoform resolvability from short read RNA-seq data

Jeremy RB Newman[1], Ana Conesa[2, 3], Lauren M McIntyre[1, 2]

1. Department of Molecular Genetics and Microbiology, University of Florida, Gainesville, Florida, United States

2. Genetics Institute, University of Florida, Gainesville, Florida, United States

3. Department of Microbiology and Cell Science, Institute for Food and Agricultural Sciences, University of Florida, Gainesville, Florida, United States

There is a growing recognition of isoform-specific effects in biological processes and disease. Algorithms have been developed to estimate isoform abundance from short-read RNA sequencing data. However their performance varies wildly, implying that current algorithms cannot resolve the uncertainty inherent in isoform reconstruction. The use of full-length transcripts obtained through PacBio sequencing can reduce uncertainty about the reference transcriptome, although this is prohibitively expensive as a high-throughput option. We sought to leverage evidence provided by long-read references to determine the exact resolution of isoform-level information retrievable from short-read sequencing. We developed an algorithm to identify the fragments within a gene that are unique to an isoform or common to multiple isoforms, and use this to quantify the uncertainty in reads assigned to specific isoforms. We tested this algorithm by using short-read sequencing from mouse to determine what isoforms are resolvable, and which are indistinguishable, and compared this is to PacBio sequencing obtained from the same samples. We tested whether isoform resolvability was influenced by read length and transcriptome complexity by simulating reads from mouse, human and Drosophila transcripts. Transcripts with highly similar sequences are not able to readily distinguishable, and sequence uniqueness is dependent on read length and transcriptome complexity, with longer reads and fewer isoforms per gene both increasing sequence uniqueness and isoform resolvability. We conclude that the resolvability of isoforms from short-read RNA-seq data is highly dependent on the identification of sequence uniqueness, and that the transcriptome-wide resolution of isoforms is not possible from short read data alone.

455

Epigenetic age estimation of humpback whale populations

Andrea Polanowski[1], Simon N Jarman[1], Jooke Robbins[2], David Chandler[3]

1. Australian Antarctic Division, Kingston, TAS, Australia

2. Center for Coastal Studies, Provincetown, MA 02657, USA

3. Australian Genome Research Facility , Perth, Western Australia, Australia

Epigenetic assays for estimating the age of humans have recently been developed based on age-induced changes in DNA methylation of specific genes. We used this information on age associated DNA methylation in human genes to identify homologous gene regions in humpback whales. In 2014, we successfully developed an epigenetic method for estimating humpback whale age with DNA from skin biopsy samples. Forty-five known-age humpback skin samples from a unique, intensively-studied population with individuals tracked since birth were employed to calibrate relationships between cytosine methylation and age. Methylation levels of 37 cytosines in regulatory regions of eight humpback whale genes were assayed and 7 were found to have a significant age-related profile. Three CpG sites with the strongest age-related methylation profiles were selected for the assay. This epigenetic age assay has an R2 of 0.82 and predicts age from skin samples with a standard deviation of 2.991 years. In 2015 we included another 27 known aged humpback whales to calibrate the assay. We found we had more variability and the R2 decreased to 0.73, however the age relationship remains significant. To demonstrate the potential of this technique, we constructed the first modern age profile of humpback whales off eastern Australia and compared the results to population structure 5 decades earlier. The modern age structure shows that there is still a deficit of older whales in this population that was hunted to near extinction in the early 1960s. Estimation of animal age by these sorts of non-lethal epigenetic biomarkers has great potential for enabling research on the population biology of many long-lived animals.

456

Regulation of mRNA level by codon usage bias

Wenfeng Qian[1]

1. Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing, China

Synonymous codons are used in different frequencies in genomes, the phenomenon that is known as codon usage bias. It has been extensively reported that the codon usage bias is strongly correlated to mRNA level, a phenomenon that is frequently explained by natural selection on translational accuracy and/or efficiency. That is, reduced translational accuracy and efficiency leads to cellular toxicity and ribosome sequestering, respectively, and such impact scales with gene expression level, which leads to stronger natural selection on synonymous codon usage among highly expressed genes. In sharp contrast, the potential role of codon usage bias in regulating mRNA level has been ignored for long, although it

could explain the correlation between mRNA level and codon usage bias as well. Here, we directly measured the impact of synonymous mutations on mRNA level, by generating in total 3,556 synonymous variants of the gene encoding green fluorescent protein (*GFP*), inducing their expression from the genome of the budding yeast, and measuring their mRNA levels with high-throughput sequencing. Surprisingly, we found that preferred codons up-regulate mRNA levels in general. In other words, even if natural selection on translation is not operating, a strong correlation between codon usage bias and mRNA level could be observed in genomes. Our study reveals the vital role of codon usage bias in regulating mRNA level, provides an alternative explanation for the well-known correlation between codon usage bias and mRNA level, and calls for re-evaluation of theories on the evolution of codon usage bias.

457

High-Throughput Identification of *Cis*-Regulatory Rewiring Events in Yeast

Shrutii Sarda[1], **Sridhar Hannenhalli**[1]

1. Center for Bioinformatics and Computational Biology, University of Maryland, College Park, Maryland, United States

A co-regulated module of genes ("regulon") can have evolutionarily conserved expression patterns and yet have diverged upstream regulators across species. For instance, the ribosomal genes regulon is regulated by the transcription factor (TF) *TBF1* in *C. albicans*, while in *S. cerevisiae* it is regulated by *RAP1*. Only a handful of such rewiring events have been established, and the prevalence or conditions conducive to such events are not well known. Here, we develop a novel probabilistic scoring method to comprehensively screen for regulatory rewiring within regulons across 23 yeast species. Investigation of 1713 regulons and 176 TFs yielded 5353 significant rewiring events at 5% FDR. Besides successfully recapitulating known rewiring events, our analyses also suggests TF candidates for certain processes reported to be under distinct regulatory controls in *S. cerevisiae* and *C. albicans*, for which the implied regulators are not known: 1) oxidative stress response (Sc-*MSN2* to Ca-*FKH2*),and 2) nutrient modulation (Sc-*RTG1* to Ca-*GCN4*/Ca-*UME6*). Further, a stringent screen to detect TF rewiring at individual genes identified 1446 events at 10% FDR. Overall, these events are supported by strong co-expression between the predicted regulator and its target gene(s) in a species-specific fashion (>50-fold). Independent functional analyses of rewiring TF pairs revealed greater functional interactions and shared biological processes between them (p=1e-03).

Our study represents the first comprehensive assessment of regulatory rewiring; with a novel approach that has generated a unique high-confidence resource of several specific events, suggesting that evolutionary rewiring is relatively frequent and may be a significant mechanism of regulatory innovation.

1. Sarda S, Hannenhalli S. 2015. High throughput identification of cis-regulatory rewiring events in yeast. Mol. Biol. Evol. 32:3047–3063. http://mbe.oxfordjournals.org/content/32/12/3047.abstract

458

Adaptive RNA editing in squid and octopus

mengyi sun[1], **Jianzhi Zhang**[1]

1. University of Michigan, Ann Arbor, MI, United States

Abstract:

RNA editing alters mRNA sequence and contributes to transcriptome diversity. Because only a handful of functional RNA-editing cases are known, it is debated whether most RNA editing observed is beneficial. Recent studies in squid and octopus reported an unusually large number of A to I editing events, many of which are nonsynonymous and have high editing levels. Such a large dataset provides an opportunity to address the aforementioned debate. We calculated the fraction of sites edited and the median editing level for synonymous editing and non-synonymous editing, respectively. In squid, although the fraction of sites edited is smaller for nonsynonymous editing than for synonymous editing, the median editing level is greater for the former than for the latter. When our analysis is limited to high editing levels (>50%), the fraction of sites edited becomes significantly greater for nonsynonymous editing than synonymous editing. In octopus, both the fraction of sites edited and the median editing level are greater for nonsynonymous editing than for synonymous editing. These patterns suggest adaptive RNA editing at the genomic scale in squid and octopus, but contrast previous findings in humans. We are searching for the potential benefit of widespread RNA editing in cephalopods and studying the evolutionary origin of this phenomenon.

459

Creating a New Code: Constructing a Synthetic Epigenetic Memory System in Mammalian Cells

Andrew H.Y. Wang[1], **Timothy A. Hore**[1]

1. Department of Anatomy, University of Otago, Dunedin, New Zealand

Cytosine methylation is essential for normal embryonic development, X chromosome inactivation, genomic imprinting and transposon silencing. In mammals, DNA methylation at CG dinucleotides is unique because it has the ability to be maintained in the absence of signals which created it. This molecular memory system is dependent upon the symmetry of CG dinucleotides, allowing methylation marks to be copied to a cognate CG on the reverse strand following replication through the action of DNA methyltransferase DNMT1. Significantly, molecular memory could be transmitted by further symmetric DNA sequences, including the CHG context (where H is any non-C nucleotides). In plants, CHG nucleotides carry epigenetic information in a manner that is CMT3 methyltransferase dependent. We propose to overexpress the plant CMT3 protein in cultured mammalian cells, and create for the first time epigenetic memory outside of the canonical CG context. In this research we will be testing the catalytic activity of the CMT3 enzyme using bisulphite sequencing to measure the global methylation level after CMT3 transfection, either in the present of existing CG methylation (when CMT3 is introduced into wild-type cells) or in the absence of other methylation system (when CMT3 is introduced into DNMT-TKO cells). Additionally we will also endeavour to investigate the potential implications such a modification to the epigenome could have to the developmental potency and morphology of the mouse embryonic stem cells. This work has significance for the emerging field of 'synthetic epigenetics' and could have extensive biotechnological and medical applications.

460

Waking the sleeping dragon - molecular insights into the hibernation of the central bearded dragon

Alex Capraro[1], Denis O'Meally[2], Paul Waters[1]

1. University Of New South Wales, Sydney, NSW, Australia

2. University of Canberra, Canberra

Hibernation is a common strategy present throughout the animal kingdom that allows organisms to survive the harsh environmental conditions of winter. In order to achieve this, hibernators modulate many biological processes, including basal metabolic rate, oxygen consumption and heart rate. Additionally, hibernators employ a number of defence mechanisms in order to mitigate physiological stress caused by starvation, oxidative stress, hypoxia, and immune challenge. Molecular-based research into hibernation is almost entirely focused on mammalian systems; however, a wealth of information remains untapped in reptilies. I will describe RNA sequencing results from liver and large intestinal tissue of central bearded dragons (*Pogona vitticeps*) in late hibernation and post-feed after arousal. Hibernation was associated with up-regulation of genes associated with stress response and cell cycle regulation, in addition to protein modification processes and epigenetic regulatory mechanisms. Arousal from hibernation was associated with enrichment of a multitude of metabolic processes, as well as gene expression patterns that indicate the activation of the innate and adaptive immune system. My results provide exciting novel insights into the biological processes and regulatory mechanisms that govern hibernation.

461

Evidence for post-transcriptional regulation of dosage compensation in platypus

William B Horspool[1], Carly Conlan[2], Paul D Waters[1]

1. University Of New South Wales, Sydney, NSW, Australia

2. Australian National University, Canberra, ACT, Australia

In mammals males have a single X chromosome, whereas females have two. A dosage compensation system was thought to have evolved to increase gene expression from the single X in males to the equivalent of two Xs. This system was assumed to be essential to survival, but within vertebrates many examples of partial or incomplete dosage compensation challenge this view. For example, in platypus, expression from the Xs is heavily female biased. All previous large scale studies of dosage compensation have focused on total RNA transcript levels. In this study RNA-seq, polysome fractionation and iTRAQ was used to analyse platypus and opossum dosage compensation at different stages from the genome to the proteome. For the first time this analysis demonstrates that post transcriptional dosage compensation might be occurring in platypus, a species previously thought to have an ineffective dosage compensation system. This has large implications for inconsistencies observed in the dosage compensation systems of other species.

462

Unravelling marsupial X inactivation with shRNA knockdowns

Kim L McIntyre[1], Shafagh A Waters[2], Kevin V Morris[1], Paul D Waters[1]

1. University Of New South Wales, Sydney, NSW, Australia

2. The University Of New South Wales, Sydney, NSW, Australia

X chromosome inactivation is an epigenetic phenomenon occurring in eutherian and marsupial mammals to rectify imbalances in X-linked gene dosage. One X chromosome is transcriptionally silenced early in female embryonic development, which is then stably maintained in somatic tissue. In eutherians, the lncRNA *Xist* is transcribed from the future inactive X, coating it *in cis* to form the distinctive *Xist* cloud. *Xist* is known to interact with chromatin-modifying complexes. In marsupials, the independently evolved lncRNA *Rsx* silences the inactive X and forms the same distinctive cloud as *Xist*. Here I use RNA fluorescence *in situ* hybridization (RNA-FISH) to investigate how knocking down epigenetic modifiers changes *Rsx* localisation and reactivation of X-linked genes.

463

Genomic evidence for elevated translational efficiency through mRNA looping

Shaohuan Wu[1]

1. Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Chaoyang District, BEIJING, China

mRNA translation is subject to sophisticated control, among which is mRNA looping.mRNA looping is mediated by the loop complex eIF4E-eIF4G-PABP and implied to facilitate translational initiation. Here we performed ribsome profiling in a yeast strain with PABP knocked down, to investigate the translational regulation on a genome-wide scale. We found that translational efficiency decrease is more dramatic in potentially looped mRNAs inferred from a previous RIP-seq study. We further observed that shorter transcripts exhibit more dramatic decrease in translational efficiency in the PABP deficient strain, suggesting that transcript length is potentially a major determinant of the degree of mRNA looping (percentage of looped mRNAs for each mRNA species). To our knowledge, this study presents the first genome-wide evidence that mRNA looping facilitates translation initiation and implicates the possible mechanism of mRNA looping, which potentially explains why ribosome protein coding genes are characterized by incredibly shorter transcript length compared to other genes.

464

Partial Translational Gene Regulation in Chicken revealved by Quantitative Mass Spectrometry

Luohao Xu[1], Severin Uebbing[1], Hans Ellegren[1], Jonas Bergquist, Niclas Backström

1. Evolutionary Biology Center, Uppsala University, Uppsala

There is increasing evidence that dosage compensation is not a ubiquitous feature following sex chromosome evolution, especially not in organisms where females are the heterogametic sex, like in birds. Even when it occurs, compensation can be incomplete and limited to dosage-sensitive genes. However, previous work has mainly studied transcriptional regulation of sex-linked genes, which may not reflect expression at the protein level. Here, we used liquid chromatography–tandem mass spectrometry to detect and quantify expressed levels of more than 2,400 proteins in ten

different tissues of male and female chicken embryos. For comparison, transcriptome sequencing was performed in the same individuals, five of each sex. The proteomic analysis revealed that dosage compensation was incomplete, with a mean male-to-female (M:F) expression ratio of Z-linked genes of 1.32 across tissues, similar to that at the RNA level (1.29). The mean Z chromosome-to-autosome expression ratio was close to 1 in males and lower than 1 in females, consistent with partly reduced Z chromosome expression in females. Although our results exclude a general mechanism for chromosome-wide dosage compensation at translation, 30% of all proteins encoded from Z-linked genes showed a significant change in the M:F ratio compared with the corresponding ratio at the RNA level. This resulted in a pattern where some genes showed balanced expression between sexes and some close to 2-fold higher expression in males. This suggests that proteomic analyses will be necessary to reveal a more complete picture of gene regulation and sex chromosome evolution.

465

Underrepresentation of active histone modification marks in evolutionarily young genes

Daqi Yu, Wenwen Shi, Yong E. Zhang

It is known that evolutionarily new genes can rapidly evolve essential roles in fundamental biological processes. Nevertheless, the underlying molecular mechanism of how they acquire their novel transcriptional pattern is less characterized except for the role of cis-regulatory evolution. Epigenetic modification offers an alternative possibility. Here, we examined how histone modifications have changed among different gene age groups in Drosophila melanogaster by integrative analyses of an updated new gene dataset and published epigenomic data. We found a robust pattern across various datasets where both the coverage and intensity of active histone modifications, histone 3 lysine 4 trimethylation and lysine 36 trimethylation, increased with evolutionary age. Such a temporal correlation is negative and much weaker for the repressive histone mark, lysine 9 trimethylation, which is expected given its major association with heterochromatin. By further comparison with neighboring old genes, the depletion of active marks of new genes could be only partially explained by the local epigenetic context. All these data are consistent with the observation that older genes bear relatively higher expression levels and suggest that the evolution of histone modifications could be implicated in transcriptional evolution after gene birth.

466

Comparative Study of Skin Gene Expression Patterns between Humans and Apes

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

1. SOKENDAI (The Graduate University for Advanced Studies), Hayama, KANAGAWA, Japan

2. Primate Research Institute, Inuyama, Aichi, Japan

Human skin has extensive characteristics in morphology and physiology compared to other primates. Skin is directly exposed to the external environment and plays essential roles, including maintenance of body temperature and moisture. The aim of our study is to understand how human-specific characteristics of skin had been acquired during the evolutionary process at genetic level. We comprehensively compared the expression levels of genes in skin between humans and apes to identify genes with human-specific expression patterns, which may contribute to human-specific characteristics.

We conducted mRNA expression analysis by using the next-generation sequencing (RNA-Seq) of skin samples of humans (n=3), chimpanzees (n=3), gorillas (n=3), and orangutans (n=3). We extracted genes showing significantly different expression patterns for each species compared to the other primates. The expression levels of 23 and 21 genes were significantly higher and lower in humans compared to apes, respectively. The number of genes with significantly different expression levels was higher in the human lineage (44 genes in total) relative to chimpanzee (22 genes in total) and gorilla (14 genes in total). We conducted gene ontology analysis for 44 genes with human-specific expression patterns to know what functions these genes would be related to. This analysis revealed that several genes have roles in developmental and metabolic process.

The larger number of genes with human-specific expression patterns may contribute to acquirement of human-specific characteristics in skin.

467

Emergence of functional novelties in a molecular lock-and-key system: allelic diversification at the self-incompatibility locus in Arabidopsis.

Maxime Chantreau[1], Céline Poux[1], Xavier Vekemans[1], Vincent Castric[1]

1. UMR CNRS-8198, Laboratoire Evo-Eco-Paleo, Université de Lille Science et Technologies, Villeneuve D'Ascq, France

The process by which new phenotypic traits can emerge in a population is a challenging question in the field of evolutionary biology, especially for phenotypes controlled by complex genetic networks requiring coordinated molecular changes between interacting partners. Addressing this issue requires an integrated understanding of both the genotype-to-phenotype map and the fitness landscape, which are hard to determine empirically for most phenotypes. To shed light on this biological process, we focused on the self-incompatibility system in outcrossing *Arabidopsis* species. Self-incompatibility is a reproductive system by which hermaphrodite flowering plants recognize and specifically reject self-pollen. In the *Brassicaceae*, it is based on a molecular lock-and-key mechanism involving two genes (the male *SCR* and the female *SRK* gene). Both display large allelic series and are tightly linked in a small non-recombining region which ensures strict haplotypic association between co-adapted alleles. The existence of this large diversity of S-haplotype has been recognized early on, and several theoretical models have been proposed for how it might arise. However, each of them has received some criticism so that, we currently still don't have a good understanding of how new S-alleles come into existence.

We decided to decipher this mechanism by travelling in time. An ancestral resurrection approach is ongoing, whereby we are currently regenerating *in planta* the ancestral *SRK* alleles of two closely related but functionally distinct *A.halleri* haplotypes by genetic transformation into *A.thaliana*. Native *SCR*and *SRK* alleles from both haplotypes have already been transformed and successful restoration of the self-incompatible phenotype has been validated through cross-pollination experiments. We are now generating plants expressing the ancestral *SCR* and *SRK* alleles, which we will use to functionally challenge the ancestral components with their two extant descendant alleles. This will enable us to determine the mutational path followed to create a new functionally divergent self-incompatible allele.

468

A genome-wide test of the modular structure of pleiotropy

Julie Collet[1], Katrina McGuigan[1], Scott Allen[1], Stephen Chenoweth[1], Mark Blows[1]

1. University of Queensland, St Lucia, QLD, Australia

The evolution of morphological traits is constrained by the nature and extent of their pleiotropic relationship with other traits. Modularity, where traits with a similar function are pleiotropic whilst pleiotropy is limited between traits belonging to different functions, is thought to enhance evolvability because: (i) modularity reduces the range of effects of deleterious mutations, as mutations would only affect the traits belonging to the targeted module rather than the entire organism, (ii) all traits of a module can respond to natural selection as a unit (iii) it preserves the module's function during evolutionary change. Despite its key role to explain the evolution of complex organisms, modularity of the genotype-phenotype map has scarcely been directly tested. Here, we adopt an original approach using gene expression traits from 41 mutation accumulation lines as phenotypes, to directly test whether functional modules show enhanced pleiotropy. Using gene expression traits as phenotypes has many advantages that include measuring thousands of traits at the same time and measuring traits that span a large number of functions, easily attributable thanks to several repositories (eg. GO terms and KEGG pathways). Contrary to the prediction of modularity, we found little evidence for higher levels of mutational pleiotropy within functional modules than in random sets of traits. We further compared the strength of selection against mutations that affected traits within functional modules to mutations that affected traits spanning several functions. We found strong selection against traits spanning a very large array of functions. However, against theoretical expectations, we found no evidence for weaker selection on function-specific mutational pleiotropy compared to selection on mutational pleiotropy across functional modules. Together, our results indicate that trait functions do not predict modularity of the genotype-phenotype map, and that mutations targeting a large number of functions are strongly selected against.

469

Seeing colour in the dark: Lessons learned from the evolution of vision in 100 fish species

Fabio Cortesi[1], Zuzana Musilova[2], Michael Matschiner[3], Sara M Stieb[1], Fanny de Busserolles[1], Martin Malmstrøm[3], Ole K Tørresen[3], Reinhold Hanel[4], Sissel Jentoft[3], Karen L Carleton[5], Justin Marshall[1], Walter Salzburger[6]

1. The Queensland Brain Institute, The University of Queensland, Brisbane, QLD, Australia

2. Department of Zoology, Charles University, Prague, Czech Republic

3. Centre for Ecological and Evolutionary Synthesis, University of Oslo, Oslo, Norway

4. Institute of Fisheries Ecology, Thünen Institute, Hamburg, Germany

5. Department of Biology, University of Maryland, College Park, MD, USA

6. Zoological Institute, University of Basel, Basel, Switzerland

Imagine living a thousand meters below sea level surrounded by darkness but for a few bioluminescent rays of light. Now imagine living in a clear mountainous lake three thousand meters above sea level where the plethora of light can cause blindness anytime. There are many examples of animals that have adapted their vision to cope with different light conditions, yet little is known about the molecular basis of these adaptations. Thanks to the advent of New Generation Sequencing Techniques, it has now become feasible to study the evolution and function of vision, along with its underlying molecular machinery (opsins and their regulatory elements), one-on-one, in nature. Here, we report transcriptome based functional analyses of opsin expression in more than 30 fish species. These species were chosen based on a comparative dataset looking at visual evolution in 100 fish genomes spanning the teleost phylogeny. We show that the expression of opsins is very dynamic including co-expression patterns and retinal regionalization (using in-situ hybridization), which correlate with the light environment and/or visual tasks of different fishes. In particular, several species have evolved ingenious strategies to adapt to extremes of their photic environments. These include three independent cases of multiple rhodopsin use in species that colonize light deprived habitats, which may well be the first evidence for 'true' rhodopsin-based colour vision in vertebrates. Understanding how fish perceive their world and how this allows them to adapt to different environments is crucial for future management purposes, especially since many fish species are facing imminent danger due to anthropogenic activities.

470

Genomic innovations at the onset of animal multicellularity: insights from the sponge histone PTM landscape

Federico Gaiti[1], Selene L. Fernandez-Valverde[2], Bernard M. Degnan[1], Milos Tanurdzic[1]

1. School of Biological Sciences, University of Queensland, St Lucia, QLD, Australia

2. Cátedras CONACyT, Laboratorio Nacional de Genómica para la Biodiversidad (LANGEBIO), CINVESTAV, Irapuato, Guanajuato, México

Recent comparative analyses of histone post-translational modifications (PTMs) demonstrated that the complex regulatory landscape present in bilaterians evolved before they diverged from cnidarians. However, it remains unclear if these regulatory innovations, which appear to underlie eumetazoan complexity and diversity, evolved earlier and were part of the genomic landscape of the very first crown animals. Here, using ChIP-Seq, we extend these analyses to generate genome-wide maps of regulatory elements in the sponge *Amphimedon queenslandica*. As a sponge, *Amphimedon* has one of the least complex animal body plans and is an extant representative of one of the oldest animal phyletic lineages. Despites it morphological simplicity - it lacks a gut, nerves and muscles - we show that this sponge shares a conserved gene regulatory landscape with eumetazoans (cnidarians + bilaterians). We found that *Amphimedon* distal *cis*-regulatory sites are characterized by the same combination of histone PTMs – H3K4me1 and H3K27ac – typically associated with eumetazoan enhancers and are preferentially enriched in the vicinity of developmental regulatory genes. Moreover, many of these distal *cis*-regulatory sites are located in microsyntenic gene blocks that are deeply conserved between sponges and eumetazoans, consistent with *cis*-regulation constraining genome architectures since the origin of animals. Overall, these results argue that a major shift in genome *cis*-regulatory complexity occurred along the metazoan stem, concomitant with the evolution of animal multicellularity. With a complex gene regulatory landscape already in place at the dawn of animals, we hypothesize that quantitative rather than qualitative differences in regulatory mechanisms led to the evolution of the diversity of eumetazoan body plans, mainly the expansion of developmental gene families (encoding transcription factors and components of signaling pathways), *cis*-regulatory DNA, non-coding RNAs and the subsequent enlargement and rewiring of regulatory networks.

471

Light environment change induces differential expression of guppy *opsins*

Alexandrea M Kranz[1], Leonard G Forgan[1], Gemma L Cole[1], John A Endler[1]

1. Deakin University, Waurn Ponds, VIC, Australia

Light environments have a critical impact on species that use vision to survive and reproduce. Visual systems must accommodate changes in light that occur from minutes to years, yet we do not know how animal visual systems respond to spectral (color) changes over the longer time scales.

Here we use an evolutionary experiment with different light environments to not only experimentally demonstrate that the visual system of guppies can evolve, but also show the effect of our evolutionary experiment on the plasticity of the visual system. We demonstrated changes in the visual system by examining the expression levels of nine guppy *opsin* genes; the genes that encode the light sensitive proteins in photoreceptors. In response to light, we found that some *opsin* expression levels underwent divergent evolution across generations and that the plastic response of a sub-set of *opsin* was determined by the light environment that the population evolved under.

472

Expansions of ependymin-related proteins in diverse invertebrate taxa and their co-option into novel functions

Carmel McDougall[1], Scott F Cummins[2], Daniel J Jackson[3], Bernard M Degnan[1]

1. School of Biological Sciences, The University of Queensland, Brisbane, QLD, Australia

2. Faculty of Science, Health and Education, University of the Sunshine Coast, Maroochydore, QLD, Australia

3. Department of Geobiology, Georg-August University of Göttingen, Göttingen, Germany

Ependymins are fish-specific glycoproteins that play key roles in CNS plasticity and memory formation. Almost a decade ago it was revealed that these ependymins evolved from a related group of proteins comprising both vertebrate and invertebrate members, however the function of these ependymin-related proteins (Epdrs) remains unknown. We have systematically identified Epdrs from both whole genomes and transcriptomes to build a broad picture of the distribution of these genes across Metazoa. We find that the genes encoding these proteins are commonly lost, and have not identified any members in the ecdysozoan species examined to date. We also find that this gene family has undergone significant expansions in other taxa, including the cephalochordate amphioxus, the brachiopod *Lingula*, the molluscs *Crassostrea, Lottia, Pinctada* and *Haliotis*, the echinoderm, *Acanthaster planci*, the placozoan *Trichoplax* and the sponge *Amphimedon*. We investigate the functions of these expanded Epdr complements in different species and find that they play diverse, lineage-specific roles, such as shell patterning in *Haliotis*, and alarm response in *Acanthaster*. Along with the originally described role of ependymin proteins in the CNS, the broad range of functions reported here reveals that this family of proteins is particularly amenable to co-option into novel roles.

473

Brain of the Blind: Transcriptome evolution in the golden-line cavefish

Fanwei Meng[1], Yahui Zhao[1], Tom Titus[2], Chunguang Zhang[1], John H Postlethwait[2]

1. Institute of Zoology, Chinese Academy of Sciences, Chaoyang Distrtict, BEIJING, China

2. Institute of Neuroscience, University of Oregon, , , Eugene, OR 97403, USA

The genus *Sinocyclocheilus* (golden-line fish) includes a dozen species of cave-dwelling blind fish and over thirty surface-dwelling sighted species. Cave environments are not only dark, but also generally nutrient poor. Cave-dwelling fish (cavefish) have convergently evolved a series of morphological adaptations to their eyes, pigmentation, brain, olfactory, and digestive systems. Although the brain expends a substantial fraction of an animal's whole-body energy use, it is as yet unknown whether energy metabolism in the cavefish brain has adapted in food-limited cave environments.

We used RNA-seq to investigate the evolution of the cavefish (*S. anophthalmus*, eyeless golden-line) transcriptome by comparing gene expression to a surface-dwelling species (*S. angustiporus*, surface golden-line) from the same genus. Gene pathways related to cholesterol biosynthesis were significantly enriched in genes downregulated in cavefish. Two sterol regulatory element binding transcription factor genes, *srebf1* and *srebf2*, and their down-stream genes, including genes encoding enzymes in cholesterol synthesis (*hmgcs1*, *hmgcra*, *sqlea*, *fdft1*, *cyp51*, *lss*, *dhcr7*, and *idi1*) were downregulated in the cavefish brain compared to surface fish brain.

Our comparative analyses provide molecular evidence for the regulation of cholesterol metabolism in the cavefish brain. Gene expression studies suggest that adaptations to dark, nutrient-poor cave environments in eyeless golden-line cavefish involved reduction in the rate of biosynthesis of cholesterol, thereby saving energy in adaptation to the cave environment.

474

Color vision re-evolution in deep-sea fishes: multiple rhodopsin genes as adaptation to an extreme environment

Zuzana Musilova[1, 2], Fabio Cortesi[3], Michael Matschiner[4], Martin Malmstrøm[4], Ole Tørresen[4], Reinhold Hanel[5], Wayne I.L. Davies[6], Karen Carleton [7], Sissel Jentoft[4], N. Justin Marshall[3], Walter Salzburger[1]

1. Zoological Institute, University of Basel, Basel, Switzerland

2. Charles University in Prague, Prague, CZ, Czech Republic

3. Queensland Brain Institute, University of Queensland, Brisbane, Australia

4. Centre for Ecological and Evolutionary Synthesis (CEES), University of Oslo, Oslo, Norway

5. Institute of Fisheries Ecology, Thuenen Institut, Hamburg, Germany

6. Functional Genomics and Photobiology Group, Zoology, School of Animal Biology, University of Western Australia, Perth, Australia

7. Department of Biology, University of Maryland, College Park, Maryland, USA

Deep-sea fishes have evolved numerous morphological and physiological adaptations, including larger eyes or rod-only retinas, to counteract the low light conditions of their environment. Color vision in vertebrates is based on the expression of different visual genes (cone opsins) in the photopic cone receptors of the retina, while the scotopic rod receptors mostly expresses a single visual gene (rhodopsin) thought to mediate color blind vision under 'dim-light' conditions. Numerous duplications of the opsin genes in teleost fishes extended the molecular substrate for subsequent adaptation to variable light conditions. However, the molecular mechanisms of dim-light-only vision and the associated loss of color vision remains poorly understood. Here we report dynamics of opsin gene evolution based on 100 genomes spanning the teleost phylogeny, with emphasis on the deep-sea fish lineages. We found strong evidence for various cone opsin losses and pseudogenization in many of the deep-sea fish species, confirming the absence of conventional cone-opsin based color vision in these lineages. Interestingly, two deep-sea fish orders, which have lost most of their cone opsins, have consequently evolved up to 30 different rhodopsins with the potential to perceive various dim-light wavelengths. Thirty rhodopsin genes from one species show traces of strong adaptive evolution and we have further confirmed differences in the wavelength

sensitivity by the opsin protein in-vitro expression. This strongly supports our findings of what might be the first evidence for rhodopsin-based color vision in vertebrates.

475

Genome-wide detection of morphotype-specific alternative splicing in chicken feather transcriptomes

Chen Siang Ng[1], Chin-Kuan Chen[2], Siao-Man Wu[2], I-Ming Liu[1], Ping Wu[3], Cheng-Ming Chuong[3], Wen-Hsiung Li[2]

1. National Tsing Hua University, Hsinchu, TAIWAN, Taiwan

2. Biodiversity Research Center, Academia Sinica, Taipei, Taiwan

3. Department of Pathology, University of Southern California, Los Angeles, CA, USA

Feathers are an excellent model for studying the development and evolution of morphological traits because of having diverse forms with hierarchical branching patterns. Various types of morphological changes are allowed to occur by the complex structure of feathers. The genetic basis of the structural differences between different parts of a feather and between different types of feather is a fundamental question in the study of feather diversity, yet the actual genetic material responsible for feather diversity remains largely unresolved. Here, we performed high-throughput mRNA sequencing to investigate gene expression profiling and alternative splicing events. We present bioinformatics analysis of several morphotype-specific splice isoforms which can be related to gene functions in different morphotypes of feathers. This study laid the ground work for studying the evolutionary origin and diversification of feathers as abundant data were produced for the study of feather morphogenesis.

476

Vertebrate Hox genes evolution: insights from the embryonic transcriptome of the hagfish

Juan Pascual-Anaya[1], Adrian Ruiz-Villalba[2], Wataru Takagi[1], Fumiaki Sugahara[3], Shigeru Kuratani[1]

1. RIKEN, Kobe, HYOGO, Japan

2. Department of Anatomy,Embryology and Physiology, Academic Medical Center (AMC), Amsterdam, The Netherlands

3. Divition of Biology, Hyogo College of Medicine, Nishinomiya, Hyogo, Japan

Hox genes are master developmental genes that control the identity of different structures and organs along the principal anterior-posterior axis of animals, as well as secondary axes such as the limbs in tetrapods. Two striking properties of Hox genes expression patterns are temporal and spatial colinearity, which are thought to be crucial for the correct specification of eutmetazoan body plans. However, little is known about Hox genes, clusters and their functions in agnathans or cyclostomes (lampreys and hagfishes), the sister group of gnathostomes or jawed vertebrates (i.e. sharks, fishes and tetrapods) and thus in key phylogenetic position to understand the last common ancestor of vertebrates. Here, we present a reference transcriptome of the Japanese inshore hagfish, *Eptatretus burgeri*, a rare specimen whose embryos are extremely difficult to obtain either in captivity or in natural conditions. We have systematically screened the hagfish transcriptome for *Hox* genes, and a hagfish BAC library for *Hox* clusters, finding a total of 40 *Hox* genes, with up to 6 *Hox4* paralogues (indicative of the presence of up to 6 putative *Hox* clusters) and a striking phenomenon of somatic differential lost of Hox genes: at least 2 Hox genes are lost in some blood cell types. We have also studied the behavior of Hox gene expression during development in the hagfish and lamprey and compare it with that of gnathostomes (using shark), and found out that colinearity is not that well defined in the hagfish as it is in other vertebrates. Our results might help to understand the big morphological disparities between cyclostomes and gnathostomes, since the regulation of Hox genes seem to be different between the two groups.

477

From Origination to Function: Life-course of Human-specific *De Novo* Protein

Qing Sunny Shen[1], Jia-Yu Chen[1], Chuan-Yun Li[1]

1. Beijing Key Laboratory of Cardiometabolic Molecular Medicine, Institute of Molecular Medicine, Peking University, Beijing, China

De novo protein-coding genes emerge from ancestral non-coding DNAs, generating proteins different from those encoded by known protein-coding genes, and may contribute a lot to species-specific traits. According to our previous study, most of the *de novo* protein-coding genes encoded long non-coding RNAs (lncRNA) in outgroups with a similar transcript structure and correlated tissue expression profile, which implies that some *de novo* protein-coding genes may originate from ancestral lncRNAs. However, although this theory has been widely accepted, the transition process as well as the functional importance of these genes is not well addressed to date.

Recently, we identified 64 hominoid-specific *de novo* genes and reported the mechanism for the origination of functional *de novo* proteins from ancestral lncRNAs. We revealed that even though the lncRNA precursors of *de novo* genes are equipped with precise splicing structures and specific tissue expression profiles, they are generally not more selectively constrained than other lncRNA loci. Besides, the existence of these newly-originated *de novo proteins* is not beyond anticipation under neutral expectation, as they generally have longer theoretical lifespan than their current age, due to their GC-rich sequence property enabling stable ORFs with lower chance of non-sense mutations. That is to say, all about the emergence and the retention of these *de novo* genes are likely driven by neutral forces. However, on the basis of the polymorphism profile provided by RhesusBase (http://www.rhesusbase.org), we found that there indeed exist signatures of purifying selection on these genes in human population, which indicates a proportion of these newly-originated proteins are already functional in human. Taken together, we proposed a mechanism for creation of functional *de novo* proteins from ancestral lncRNAs during the primate evolution, which may contribute to human-specific genetic novelties by taking advantage of existed genomic contexts.

1. C Xie et al. Hominoid-specific de novo protein-coding genes originating from long non-coding RNAs. PLoS genetics, 2012.
2. JY Chen et al. Emergence, Retention and Selection: A Trilogy of Origination for Functional De Novo Proteins from Ancestral LncRNAs in Primates. PLoS genetics, 2015
3. XM Zhong et al. RhesusBase PopGateway: Genome-wide Population Genetics Atlas in Rhesus Macaque. Molecular Biology and Evolution, 2016

478

Cell-type specific transcriptomes in the sponge *Amphimedon queenslandica* sheds light on the ancestral metazoan body plan

Shunsuke Sogabe[1], Daniel Stoupin[1], Kerry Roper[1], Sandie Degnan[1], Bernie Degnan[1]

1. School of Biological Sciences, The University of Queensland, St. Lucia, QLD, Australia

The metazoan last common ancestor probably had minimally epithelial and mesenchymal cells. One or both cell types are likely to have had the capacity to self-renew and transdifferentiate (i.e. act as stem cells). They could convert into each other via epithelial-mesenchyme and mesenchyme-epithelial transitions. Using the single-cell RNA-Seq technique CEL-Seq, here we assess the transcriptomes of pinacocytes, choanocytes and archeocytes from the sponge *Amphimedon queenslandica*. Pinacocytes comprise external and internal epithelial-like layers, choanocytes form internal ciliated chambers that function in feeding and are part of the internal epithelium, and archeocytes are amoebocytic mesenchyme cells. Choanocytes and archeocytes also act as stem cells, with the latter giving rise to most sponge cell types. Principle component analysis distinguishes these three cell type transcriptomes, with pinacocytes enriched in genes involved in cell adhesion, cell-cell signaling, cell surface receptors - including sensory rhodopsin-like GPCRs - and internal signal transduction (e.g. GPCR signaling pathway, MAPK cascade) consistent sensory epithelia. Genes significantly upregulated in choanocytes are enriched in metabolic and signaling (e.g. Notch) pathways, suggesting a role in feeding and intercellular communication; choanocytes have a greater enrichment in metazoan-specific and novel genes. Many genes upregulated in archeocytes are associated with cell cycle regulation, mitosis, transcription, translation and RNA processing, suggesting these cells are poised to either divide or change expression profiles. Archeocytes also have the most divergent transcriptomes, consistent with these pluripotent stem cells being in a range of cell states. With gene expression in*Amphimedon* cells being akin to cognate eumetazoan epithelial, mesenchymal and stem cells, we infer that these cell types were part of the crown metazoan body plan that gave rise to extant sponges and eumetazoans.

479

Rapid expansion of pigmentation genes in Penaeid prawns with absolute conservation of gene function.

Alyssa Budd[1], Mark Tonks[2], Sue Cheers[1], Tracey M Hinton[3], Nicholas M Wade[1]

1. CSIRO Agriculture, Brisbane, QLD, Australia

2. CSIRO Oceans and Atmosphere, Brisbane, QLD, Australia

3. Australian Animal Health Laboratory, CSIRO Biosecurity, Geelong, VIC, Australia

Crustaceans have a unique mechanism for producing their colour (Wade *et al*., 2009). The red carotenoid astaxanthin (Axn) is the central chromophore to a protein complex called crustacyanin (CRCN) that modifies the colour of Axn to any other colour in the visible spectrum. In this way, crustaceans produce their cryptic shell colours and patterns that are used in camouflage, survival, reproduction and communication. Traditionally, there were only thought to be 2 genes encoding crustacean colour proteins, CRCN-A and CRCN-C. Using a combination of degenerate PCR and next generation sequence data mining, this study identified 36 potential new CRCN genes from 11 species of Penaeid prawns, with some species containing up to 6 different CRCN isoforms. Despite their nucleotide sequence differences, the ratio of non-synonymous to synonymous mutations showed that the majority of sequences had significant evidence of stabilising selection. As a consequence, predictive 3D structure modelling showed absolute preservation of CRCN-A or CRCN-C structure across all sequences. In *Penaeus monodon*, RNA interference was used to functionally down-regulate three different CRCN isoforms, and this caused distinct gene-specific colour changes depending on which CRCN isoform was down regulated. Expression of these three isoforms was also differentially regulated across the moult cycle. These results demonstrate that these are functional CRCN gene isoforms, and that there has been a rapid expansion of the CRCN gene lineage in Penaeid prawns. Our study showed that the production and regulation of pigmentation in crustaceans is more complex than initially thought, but the mechanism by which colour is produced in crustaceans is strictly conserved.

1. Wade NM, Tollenaere A, Hall MR and Degnan BD (2009) Molecular Biology and Evolution 26(8):1851–1864.

480

Functional diversification of chaperonin paralogs in cyanobacteria with cell differentiation

Julia Weissenbach[1], Judith Ilhan[1], David Bogumil[2], Karina Stucken[1], Tal Dagan[1]

1. Christain- Albrechts University Kiel, Kiel, SCHLESWIG-HOLSTEIN, Germany

2. Ben-Gurion University of the Negev, The Department of Life Sciences & the National Institute for Biotechnology in the Negev, Beersheba, Israel

The chaperonin GroEL and its co-factor GroES promote protein folding in an ATP-dependent manner and are known to have an influence on adaptation to diverse stress conditions. The chaperonin complex includes GroEL, that forms a barrel-like oligomer, and GroES that forms the lid. In most eubacteria the GroESL chaperonin is encoded by a single-copy bicistronic operon that includes the *groES* and *groEL* genes. Comparative analysis of cyanobacterial genomes showed that the GroESL chaperonin genes were duplicated at least twice during the evolution of heterocystous, filamentous cyanobacteria. Here we study the functional diversification of groEL/groES in the multi-seriate filament forming cyanobacterium *Chlorogloeopsis fritschii* PCC 6912. The genome of *C. fritschii* encodes two *groESL* operons (*groESL1*, *groESL2*) and a monocistronic *groEL* gene (*groEL3*). A comparison of gene expression under stress conditions shows that *groEL1* is upregulated during temperature stress whereas the monocistronic *groEL3* is upregulated under light stress. The expression of *groEL2* is induced upon nitrogen deficiency during heterocyst differentiation. In addition, expression of *groEL1* is localized in a distinct pattern under diazotrophy. To establish the GroEL-GroES specificity, we tested for protein interactions between the chaperonin subunits *in vivo*. Subunits encoded in the two operons form hybrid complexes, whereas GroEL3 subunits are not forming oligomers nor interact with any of the two co-chaperonins. Interaction between GroES2 and GroEL2 could not be documented, indicating that the GroESL2 operon does not encode a functional chaperonin complex. Experiments of functional complementation in *E. coli* confirm that *groESL1* can substitute the native operon. Furthermore, *groEL2* could complement the native *groEL* only in combination with *groES1*. GroEL3 was not functionally complement in any combination of co-chaperonins. Our results demonstrate that the evolutionary consequences of *groEL* duplication include specialization as a housekeeping gene of *groEL1*, subfunctionalization of *groEL2* and neofunctionalization of *groEL3.*

481

Preliminary insights into selection in the platypus

Portia Westall[1], Hilary Martin[2], Elizabeth Batty[2], Colin Freeman[2], Tasman Daish[3], Frank Grutzner[3], Tom Grant[4], Rory Bowden[2], Peter Donnelly[2], Jaime Gongora[5]

1. University of Sydney, Lane Cove, NSW, Australia

2. Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, England
3. Univeristy of Adelaide, Adelaide, SA, Australia
4. Biological, Earth and Environmental Sciences, University of New South Wales, Sydney, NSW, Australia
5. Facuty of Veterinary Science, The University of Sydney, Camperdown, NSW, Australia

The platypus is endemic to Australia and exhibits a fascinating suite of characteristics, being a specialised semi-aquatic, fossorial, carnivorous and egg laying mammal. It occurs naturally in freshwater streams, rivers and lakes of eastern Australia, including Tasmania and King Island. Mitochondrial and nuclear DNA has revealed the existence of least three evolutionary significant units within the platypus and existent of discrete genetic populations at the regional and river basin level with very limited gene flow. In addition, platypus genome and transcriptome data analyses have revealed unique signatures of evolution and the role of different selective pressures on gene expression. Given this and the position of the platypus as part of the most basal mammal group, this species is ideal for investigating the evolution and genetic factors underlying of different biological processes in mammals. This includes whether genetic structuring and differences between the documented platypus evolutionary lineages have resulted in gene, and potentially genome-wide differences, that could provide insights to natural selection. Here we assess the genetic diversity of genes associated with metabolism and electroreception using genome sequences from platypuses from New South Wales. This is the first whole-genome sequencing project to look at the landscape of coding genes being part of a collaborative research program with the University of Oxford in which platypus population structure and history, and fine-scale recombination rates are being studied.

482

Live birth and the genetic basis of evolutionary innovation

Camilla M Whittington[1], Mike B Thompson[1], Oliver W Griffith[1], Weihong Qi[2], Anthony B Wilson[3]

1. University of Sydney, Sydney, NSW, Australia
2. Functional Genomics Centre Zurich, Zurich, Switzerland
3. Department of Biology, Brooklyn College, City University of New York, New York, NY, USA

Evolutionary innovations such as eyes and live birth (viviparity) are dramatic, adaptive novelties that have shaped the evolutionary trajectories of animals. Viviparity is an important biological innovation that has required a set of complex phenotypic changes to allow internal incubation of embryos, radically changing the way in which organisms interact with their environment and transmit their genes to the next generation. As viviparity has evolved convergently hundreds of times in mammals, reptiles, fish, amphibians, and invertebrates, it is an ideal model to study evolutionary innovations, offering the opportunity to compare and contrast naturally replicated evolutionary experiments.

There are at least 23 independent origins of viviparity in fishes, with syngnathid fishes (seahorses and pipefish) unique in exhibiting male pregnancy. Male seahorses and pipefish have evolved specialized brooding pouches that provide protection, gas exchange, osmoregulation, and limited nutrient provisioning to developing embryos. Pouch structures differ widely across the Syngnathidae, offering an ideal opportunity to study the evolution of reproductive complexity. However, the physiological and genetic changes facilitating male pregnancy are largely unknown. We used transcriptome profiling to examine pouch gene expression at successive gestational stages in a syngnathid with the most complex brood pouch morphology, the seahorse *Hippocampus abdominalis*. Using a unique time-calibrated RNA-seq data set including brood pouch at key stages of embryonic development, we identified transcriptional changes associated with brood pouch remodelling, nutrient and waste transport, gas exchange, osmoregulation, and immunological protection of developing embryos at conception, development and parturition. Key seahorse transcripts share homology with genes of reproductive function in pregnant mammals, reptiles, and other live-bearing fish, suggesting a common toolkit of genes regulating pregnancy in divergent evolutionary lineages. Our work shows that there are common mechanisms that underpin the development of evolutionary innovations across divergent species.

483

Parallel evolution of the auxin pathway in an Australian wildflower

Melanie J Wilkinson[1], Federico Roda[1, 2], Maddie E James[1], Henry L North[1], Greg M Walter[1], Christine A Beveridge[1], Daniel Ortiz-Barrientos[1]

1. School of Biological Sciences, The University of Queensland, Brisbane, Queensland, Australia
2. Department of Organismic and Evolutionary Biology, The Arnold Arboretum of Harvard University, Boston, MA, USA

Traits may repeatedly evolve in independent populations inhabiting similar environments. Whether this occurs with populations utilising different genetic mechanisms or the same genetic mechanisms (either through the same mutations, genes or biochemical pathway) remains largely unknown, leaving us ignorant to how evolution by natural selection proceeds at the molecular level. Here, we contribute to filling this gap with the *Senecio lautus* system, where populations have repeatedly evolved divergent growth forms in two contrasting adjacent environments. We use a combination of genetic, phenotypic and physiological experiments in natural and mapping populations to isolate the genetic mechanisms involved in the evolution of these two growth forms. Our results indicate that the auxin pathway, a pathway that controls plant growth and development is repeatedly differentiated in many natural populations adapting to contrasting environments. Furthermore, we found an auxin controlled phenotype, response to a change in gravity, to be divergent in many but not all independent *S. lautus populations.* Overall the results indicate that *S. lautus* populations use both the same and different genetic mechanisms in the evolution of growth form.

484

The genetic basis of evolutionary transitions in early development

Christina Zakas[1], Matthew Rockman[1]

1. Department of Biology, New York University, New York, NY, United States

Phenotypic evolution in animals is constrained by the mechanics of early development. How do major transitions in development occur? Historically, efforts to address this question have been limited to comparative methods. The polychaete annelid *Streblospio benedicti* provides a unique opportunity to use forward genetics to experimentally dissect a major transition in animal development. *S. benedicti* is ideal because it produces two distinct offspring types that differ in egg size, early development, and larval morphology. *S. benedicti* is thus a genetic model for the evolutionarily common transition between indirect and direct development. Using genetic crosses between these types, we constructed the first annelid genetic map, which reveals the distribution of genetic factors affecting a suite of genetically separable developmental phenotypes. Because early development is strongly influenced by maternal effects, our cross design disentangles maternal and zygotic genetic effects and shows that a

transition from indirect to direct development requires contributions from both the zygotic and maternal genome; an increase in egg size alone is not sufficient to change development mode. By identifying the loci responsible for regulating early development, we uncover how the dimorphic developmental program is determined on a whole-genome level.

485

Re-categorising the protein structure universe

Ashar J Malik[1], Anthony M Poole[2], Jane R Allison[1, 2, 3]

1. Centre for Theoretical Chemistry and Physics, Institute of Natural and Mathematical Sciences, Massey University, Auckland, New Zealand

2. Biomolecular Interaction Centre, School of Biological Sciences, University of Canterbury, Christchurch, Canterbury, New Zealand

3. Maurice Wilkins Centre, University of Auckland, Auckland, New Zealand

Protein structure exhibits greater evolutionary preservation than sequence, due to its closer relationship to function and therefore to phenotype. The use of protein structure to probe evolutionary relationships has been attempted previously[1,2] and shown to be promising. Protein structures have also been categorised according to both structural and sequence similarity in databases such as SCOP. Here we investigate the capacity of standard measures of protein structural similarity to cope with proteins of different sequence and length, and elucidate the limits of their usefulness. We define the sequence overlap length at which otherwise identical protein structures are no longer deemed similar, and the degree of structural similarity at which these metrics can no longer distinguish two protein structures. Ultimately, evolutionary relationships derived from comparing protein structure using these tools have the potential to reorganise protein structural databases, changing the way we use and interpret these resources as well as our understanding of how molecular-level changes drive evolution.

1. 1. Lundin, D., Poole, A. M., Sjöberg, B.-M., & Högbom, M. Use of Structural Phylogenetic Networks for Classification of the Ferritin-like Superfamily. J. Biol. Chem., 2012, 287(24), 20565–20575.
2. 2. Lundin, D., Berggren, G., Logan, D.T. & Sjöberg, B.-M. The Origin and Evolution of Ribonucleotide Reduction. Life, 2015, 5(1), 604–636.

486

Domain-swap polymerization drives the self-assembly of the bacterial flagellar motor

Matt AB Baker[1], Robert M Hynson[2], Lorraine A Ganuelas[2], Nasim Shah Mohammadi[2], Chu Wai Liew[2], Anthony A Rey[2], Anthony P Duff[3], Andrew E. Whitten[3], Cy M. Jeffries[4], Nicolas J Delalez[5], Yusuke V Morimoto[6], Daniela Stock[2], Judith P Armitage[5], Andrew J. Turberfield[7], Keiichi Namba[6], Richard M Berry[7], Lawrence K Lee[1]

1. Victor Chang Cardiac Research Institute / UNSW Centre for Single Molecule Science, Darlinghurst, NSW, Australia

2. Victor Chang Cardiac Research Institute, Darlinghurst, NSW, Australia

3. Australian Nuclear and Science Technology Organisation, Lucas Heights

4. European Molecular Biology Laboratory, Hamburg

5. Biochemistry, Oxford University, Oxford

6. Graduate School of Frontier Biosciences, Osaka University, Osaka

7. Physics, Oxford University, Oxford

Natural systems at all scales emerge from many components capable of nothing more than simple interactions with their neighbours. This self-organization is one of life's defining features. As an example, the flagellar motor, one of nature's most impressive machines, self-assembles from many protein parts to form a tiny nanoscale motor that can spin five times faster than a Formula1 engine. The paradox of molecular assembly is that many protein subunits avoid premature aggregation, yet at the right time self-assemble to form such complex systems. Here we demonstrate that FliG, one of the first motor proteins to assemble, forms ordered ring structures via domain-swap polymerization, which in other proteins has been associated with uncontrolled and deleterious protein aggregation. Solution and crystal structural data, in combination with *in vivo* biochemical crosslinking experiments and evolutionary covariance analysis, reveal that FliG exists predominantly as a monomer in solution but only as domain-swapped polymers in assembled flagellar motors. We propose a general structural and thermodynamic model for self-assembly, where a structural template controls assembly and shapes polymer formation into rings. This provides a general mechanism by which subunits can assemble into fixed-size rings and reveals the molecular interactions that govern the emergence of complex molecular machinery.

1. Baker et. al. (Nature Structure and Molecular Biology 2016) doi:10.1038/nsmb.3172

487

To convert or not to convert: homology requirements for meiotic gene conversion

Diego A Hartasánchez[1], Marina Brasó-Vives[1], Arcadi Navarro[1]

1. Institute of Evolutionary Biology (Universitat Pompeu Fabra - CSIC), Barcelona, BARCELONA, Spain

Gene conversion is a recombination mechanism in which information is unilaterally transferred from one DNA duplex to another. During meiosis, double-strand-breaks might be repaired by different mechanisms of homologous recombination such as single-strand annealing, double-strand-break repair or synthesis-dependent strand annealing. The latter is thought to be the cause of the large majority of meiotic gene conversion events and extensive experimental work has been performed to understand the precise molecular mechanisms and molecules involved in this pathway.

For gene conversion to occur, there needs to be a high degree of homology between the invading strand of DNA and the invaded donor strand from which complementary DNA will be synthesized. In fact, there seems to be a minimal efficient processing segment, that is, a 100% identity tract between donor and receptor strands, necessary for the gene conversion process to initiate. Further evidence indicates that there might also be an identity requirement for the resolution of the repair pathway. Additionally, experiments performed in yeast suggest that the mismatch-repair machinery is not only involved in the repair of heteroduplexes formed during meiosis, but is also responsible for ensuring homologous recombination.

In light of this new evidence, we present a complete model of gene conversion that includes identity requirements and mismatch-repair. Results from our simulations suggest that the length of gene conversion tracts are a consequence of the action of these mechanisms and are consistent with the synthesis-dependent strand annealing model for gene conversion.

488

Extreme convergent evolution in defensin proteins and quantitative maps of their sequence space

Thomas Shafee[1], Marilyn Anderson[1]

1. La Trobe Institute for Molecular Science, Melbourne, Victoria, Australia

The defensins are ancient, diverse and fast-evolving cysteine-rich proteins found across the eukaryotes. They variously display antimicrobial, signalling and ion channel disruption activities. However, their highly divergent sequences have caused traditional methods of sequence analysis to fail, hampering our understanding of how such proteins evolve and how to engineer their activities. To address this shortfall, we have applied structure-based analyses, and developed new methods of cysteine-rich protein alignment and quantitative maps of protein sequence space.

Through these methods we have shown that the defensins consist of two, independent superfamilies that have undergone some of the most extreme convergent evolution currently known for protein sequence, structure and function. The use of disulphides to display loop sequences makes the defensins highly evolvable, but also imposes evolutionary constraints which have funnelled their convergence.

Structural homology is used to guide multiple sequence alignments by barcoding cysteines known to be genuinely homologous. We use these alignments to generate quantitative maps of protein sequence space. Using multivariate analysis to rotate and project these mega-dimensional spaces into a human-understandable space reveals naturally occurring clusters of sequences with similar biophysical properties. Finally, this allows us to mine the existing diversity generated by evolution to design cluster-central 'archetypal' sequences, somewhat analogous to ancestral sequence reconstruction for engineering increased activity, promiscuity and stability.

The techniques developed for this particularly difficult case are applicable to other protein superfamilies, and complement established sequence analysis methods.

489

How Nemo sees its colourful world: variability of visual pigment genes (opsins) in anemomefish (Amphiprioninae)

Sara Stieb[1], Fabio Cortesi[1], Brian Dalton[2], Fanny de Busserolles[1], Wen-Sung Chung[1], Karen Carleton[2], Justin Marshall[1]

1. University of Queensland, St Lucia, QLD, Australia

2. Department of Biology, University of Maryland, College Park, MD 20742, USA

One of the most famous and interesting subfamilies of damselfish (Pomacentridae) are species of anemonefish. They form dominant hierarchies with one sexually matured pair and subdominant males; ones the dominant female dies, the most dominant male will become a female. Making use of modern molecular, neuroanatomical and physiological approaches, we study visual adaptation on the level of photoreceptors and their intrinsic light absorbing visual pigment genes (opsins) by comparing different anemonefish species:

- We use Illumina RNA sequencing to compare sequence and expression variation in opsins.
- We combine retinal mapping with fluorescent *in situ* hybridization (FISH) to visualize the topographic distribution of photoreceptors and expressed opsins throughout the retina.
- We measure the spectral absorbance of visual pigments using microspectrophotometry.

Previous studies revealed that damselfish possess one rhodopsin (RH1) used for scotopic, and four cone opsins used for photopic vision being short- (SWS1 and SWS2B), medium- (RH2A and RH2B) and long-wavelength (LWS) sensitive. Using RNA sequencing, we compared the cone opsin expression in five species of anemonefish (*Amphiprion akindynos*, *A. melanopus*, *A. percula*, *A. perideraion*, and *Premnas biaculatus*): all species expressed SWS1, RH2B, RH2A, and LWS. Interestingly, the expression of LWS was prominent in all species but low in *A. akindynos*, possibly reflecting different foraging styles. By comparing a wider range of damselfish species, we could already demonstrate that an increased LWS-expression is correlated to herbivory. More strikingly, we found that the UV-sensitive SWS1 has duplicated in anemonefish with different species expressing different or both copies. Analysis now underway indicate that this duplication event is widespread across the damselfish phylogeny underpinning the importance of UV-vision in damselfish. Making use of FISH and state-of-the-art microscopy (Discovery Spinning Disc Confocal) we currently produce whole retinal maps visualizing expressed opsin genes to further examine visual adaptations in anemonefish.

490

Retrieval of ancient mammalian mitochondrial DNA from Middle and Late Pleistocene sediment

Viviane Slon[1], Charlotte Hopfe[1], Marie Soressi[2], Anatoly P. Derevianko[3, 4], Michail V. Shunkov[3, 5], Richard G. Roberts[6], Zenobia Jacobs[6], Bo Li[6], Henry de Lumley[7, 8], Christian Perrenoud[7], John R. Stewart[9], Rebecca Miller[10], Pavao Rudan[11], Željko Kućan[11], Ivan Gušić[11], Sarah Nagel[1], Birgit Nickel[1], Kay Prüfer[1], Janet Kelso[1], Svante Pääbo[1], Matthias Meyer[1]

1. Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany

2. Faculty of Archaeology, Leiden University, Leiden, The Netherlands

3. Institute of Archaeology and Ethnography, Russian Academy of Sciences, Siberian Branch, Novosibirsk, Russia

4. Altai State University, Barnaul , Russia

5. Novosibirsk National Research State University, Novosibirsk, Russia

6. Centre for Archaeological Science, School of Earth & Environmental Sciences, University of Wollongong, Wollongong, New South Wales, Australia

7. Centre Européen de Recherches Préhistoriques de Tautavel, Tautavel, France

8. Institut de paléontologie humaine, Paris, France

9. School of Applied Sciences, Bournemouth University, Poole, Dorset, UK

10. Service de Préhistoire, Université de Liège, Liège, Belgium

11. Anthropology Center of the Croatian Academy of Sciences and Arts, Zagreb, Croatia

Sediment constitutes a ubiquitous feature of archaeological sites. Previous studies have shown that mitochondrial (mt) or chloroplast DNA can be recovered from ancient sediment by PCR [1, 2]. Here we describe the retrieval of mammalian mtDNA from sediments by hybridization capture [3].

We tested the ability of six DNA extraction methods to retrieve DNA of different sizes bound to 150 mg of clay or lime. We then extracted DNA from 83 sediment samples from six sites: Caune de l'Arago (France), Chagyrskaya Cave (Russia), Denisova Cave (Russia), Les Cottés (France), Trou Al'Wesse (Belgium) and Vindija Cave (Croatia). Aliquots of each DNA extract were converted into DNA libraries and subjected to hybridization capture using probes spanning the complete mitochondrial genomes of 242 mammals, and the isolated DNA fragments were sequenced.

We used simulated datasets composed of different proportions of mammalian and bacterial DNA sequences of different lengths, where varying levels of nucleotide substitutions typically present in ancient DNA [4] had been introduced, to test our ability to identify taxa. By comparison to a database of mammalian mtDNA sequences using BLAST [5] and assignment to taxa using MEGAN [6], taxon compositions at the family level were accurately reconstructed even in the presence of high levels of ancient DNA-like nucleotide substitutions. Biases potentially introduced by the capture procedure and the effect of sequencing depth on taxon composition were also investigated.

DNA sequences were retrieved from all archaeological sediments tested. Cytosine to thymine substitutions at terminal positions, an indication that DNA fragments are of ancient origin [4], were seen at five of the sites. A variety of taxa from twelve families were identified, including mammoths and woolly rhinoceroses. We conclude that molecular analysis of sediment by hybridization capture is a highly parallelizable and non-destructive approach to identify the past presence of animals at archaeological sites.

1. Haile, J., et al., Ancient DNA reveals late survival of mammoth and horse in interior Alaska. Proc Natl Acad Sci U S A, 2009. 106(52): p. 22352-7.
2. Willerslev, E., et al., Diverse plant and animal genetic records from Holocene and Pleistocene sediments. Science, 2003. 300(5620): p. 791-5.
3. Slon, V., et al., Mammalian mitochondrial capture, a tool for rapid screening of DNA preservation in faunal and undiagnostic remains, and its application to Middle Pleistocene specimens from Qesem Cave (Israel). Quat Int, 2016. 38: p. 210-8.
4. Briggs, A.W., et al., Patterns of damage in genomic DNA sequences from a Neandertal. Proc Natl Acad Sci U S A, 2007. 104(37): p. 14616-21.
5. Altschul, S.F., et al., Basic local alignment search tool. J Mol Biol, 1990. 215(3): p. 403-10.
6. Huson, D.H., et al., MEGAN analysis of metagenomic data. Genome Res, 2007. 17(3): p. 377-86.

491

The landscape of mitochondrial genetic diversity in chimpanzees from Gombe National Park and across the genus *Pan*

Anne C Stone[1, 2], Maria A Nieves Colon[2], Andrew T Ozga[1], Charlotte E Till[2], Kat Fowler[2], Rebecca Nockerts[3], Michael Wilson[3], Ian Gilby[1, 2], Anne Pusey[4]

1. Institute of Human Origins, Arizona State University, Tempe, Arizona, United States

2. School of Human Evolution and Social Change, Arizona State University, Tempe, Arizona, United States

3. Anthropology, University of Minnesota, Minneapolis, Minnesota, United States

4. Anthropology, Duke University, Durham, North Carolina, United States

We examine the relationship between geographic location and population sub-structure among chimpanzees (*Pan troglodytes*) and bonobos (*Pan paniscus*) across Africa using a set of new and previously published mitochondrial DNA data. We analyzed 1011 HVR1 and 45 mitochondrial genome (mtgenome) geo-referenced sequences. Included in these data are nine new mtgenome sequences obtained from DNA extracted from the dentin and dental calculus of chimpanzees buried at Gombe National Park. Genetic analyses of the Gombe chimpanzees did not begin until the early 1990s, and our analyses show poor DNA preservation in dentin from individuals buried there prior to that time. However, mtgenome sequences were recoverable from dental calculus. Median network analyses identify distinguishable local clusters of mtDNA lineages within all chimpanzee subspecies as well as bonobos. Mismatch distribution analyses and estimates of Tajima's D and Fu's F indicate historical population expansion in *P.t.schweinfurthii* but not in the other groups, corroborating previously published research. Our preliminary insights are being expanded by Spatial Principal Components analyses and Bayesian inference on both the HVR1 and complete mtDNA datasets. Samples of unconfirmed geographic origin are being added to these analyses to determine if the structures observed remain unchanged. If successful, this research may allow the estimation of the geographic area of origin of individuals with unknown histories beyond a simple sub-species classification. This is the first attempt at a genus-wide spatial analysis with a large sample size across all members of the genus *Pan*.

492

DNA capture of mitogenomes from Sacred Ibis Mummies of Ancient Egypt

Sally Wasef[1], Sankar Subramanian[1], David Martin Lambert[1]

1. EFRI, Griffith University, Nathan, QLD, Australia

The ancient animal catacombs throughout the Egyptian deserts harbour millions of well-preserved mummified Sacred Ibises (*Threskiornis aethiopicus*), the source of which remains unknown. It is thought that since 600 BC, pilgrims presented the mummified Ibises as 'votive' offerings to Thoth, the Egyptian God of wisdom. We have recently estimated the radiocarbon ages of mummies provenience from Saqqara, Roda and Thebes in Egypt age between 2220–2430 yr. BP . The demand for Sacred Ibis was so great it was estimated that each year ~10,000 mummies were interred at one catacomb alone. Such massive numbers suggest that ancient Egyptians kept and reared Ibis on an industrial-scale. However, there is limited evidence in ancient writings that support this suggestion. Sacred Ibis were once prevalent in Egypt but were driven to extinction as early as the mid 1800's. To investigate the likely farming methods employed by these ancient people, we used ancient DNA technology and targeted hybridization enrichment methods to retrieve complete mitochondrial genomes from ancient Sacred Ibis mummies, together with the genomes of contemporary individuals from widespread African populations. Unexpectedly, we show a remarkably high level of mitochondrial genetic variation among ancient Egyptian Sacred Ibis, very similar to that found for all modern wild African populations. Mitochondrial haplotype phylogenies and network analyses showed no genetic evidence to support the existence of a single large centralised Sacred Ibis farm in Ancient Egypt, nor for the presence of smaller localised farms. However, our results support the hypothesis that the ancient Egyptians likely complemented captive Sacred Ibis stocks held in localised enclosures, via multiple wild sources. This method would have helped maintain the health of these populations and facilitated the high Ibis mummies supply in response to the year-by-year demand for sacrificial birds.

493

Historical phylogeography of mainland and Tasmanian thylacines (*Thylacinus cynocephalus*) using ancient mitochondrial genomes

Lauren C White[1], **Jeremy J Austin**[1]

1. Australian Centre for Ancient DNA, University of Adelaide, Adelaide, SA, Australia

Before it became extinct in 1936, the Tasmanian tiger or thylacine (*Thylacinus cynocephalus*) was the world's largest carnivorous marsupial. Its extinction from mainland Australia approximately 3000 years ago, and survival on the island of Tasmania, is a mystery that has attracted much debate and research. The timing of the thylacines' mainland extinction implicates the introduction of the dingo, human intensification and climate change as possible causes. However, extinction time alone is insufficient to resolve the relative importance of these extinction drivers, and determining past population dynamics is essential for this debate to move forward. For example, recent genetic research on Tasmanian and mainland devils (*Sarcophilus harrisii*, which also became extinct on the mainland 3000 years ago), suggests that their mainland extinction was synchronous with a severe decline in the Tasmanian population, implicating a common driver. Here we present the first ancient DNA sequences from mainland thylacines and compare these to the recently extinct Tasmanian population. We used hybridization capture and next generation sequencing to generate mitochondrial genomes from ancient mainland (>3290 years before present), ancient Tasmanian (>500 years before present), and historic Tasmanian (1800-1936) thylacines. We use these data to examine the demographic history of thylacines on Tasmania and the mainland and gain insight into what led to the mainland extinction and Tasmanian survival.

494

Functional responses of salt marsh microbial communities to long-term nutrient enrichment

Christopher Graves[1]

1. Brown University, Providence, RI, United States

Environmental nutrient enrichment from human agricultural and waste runoff could cause changes to microbial communities that allow them to capitalize on newly available resources. Currently, the response of microbial communities to nutrient enrichment remains poorly understood, and while some studies have shown no clear changes in community composition in response to heavy nutrient loading, others targeting specific genes have demonstrated clear impacts. In this study we compared functional metagenomic profiles from sediment samples taken along two salt marsh creeks, one of which was exposed for more than 40 years to treated sewage effluent at its head. We identified strong and consistent increases in the relative abundance of microbial genes related to each of the biochemical steps in the denitrification pathway at enriched sites. Despite fine-scale local increases in the abundance of denitrification-related genes, the overall community structure based on broadly-defined functional groups and taxonomic annotations was similar and varied with other environmental factors, such as salinity, which were common to both creeks. Homology-based taxonomic assignments of nitrous-oxide reductase sequences in our data show that increases are spread over a broad taxonomic range, thus limiting detection from taxonomic data alone. Together, these results illustrate a functionally targeted yet taxonomically broad response of microbial communities to anthropogenic nutrient loading, indicating some resolution to the apparently conflicting results of existing studies on the impacts of nutrient loading in sediment communities.

500

Yeast populations adapted to periodic stress trade fast growth for stress resistance

Christopher J Graves[1], **Chris Bauer**[2], **Mark L Siegal**[2], **Gregory I Lang**[3], **Daniel M Weinreich**[1]

1. Department of Ecology and Evolutionary Biology, Brown University, Providence, RI, United States

2. Center for Genomics and Systems Biology, New York University, New York, NY, United States

3. Department of Biological Sciences, Lehigh University, Bethelehem, PA, United States

Natural populations evolve in unpredictable environments where rare and devastating events such as drought, fire, and starvation pose the risk of extinction. Episodes of severe stress are presumed to underlie traits like plasticity and bet-hedging, that help mitigate environmental risk. In this study we experimentally evolved yeast populations in an environment characterized by intermittent episodes of heat-induced stress. Stress events initially killed nearly 99% of the population. However, after 400 generations all experimental populations had evolved significantly reduced mortality rates, along with a reduction in the time taken to recover after heat shock. More notably, stress-adapted populations also showed significantly reduced rates of growth in benign conditions relative to the ancestral strain. Indeed, a tradeoff between growth rate and stress tolerance is predicted from prior work on the yeast stress response, which shows that reduced growth, regardless of its cause, is correlated with increased stress resistance. To dissect the various components of fitness in our experimental regime, we collected time-lapse microscopy data on thousands of individual cells during exponential growth and heat shock. We show that the time-averaged geometric mean fitness of stress-adapted populations is far greater than the ancestor, despite the short-term cost of reducing growth during favorable conditions. Still, these results imply the establishment and fixation of mutations that reduce fitness at all times except during the brief instance of heat stress. We are currently investigating the dynamics of adaptive mutations arising in our experiment using whole-genome whole-population sequencing from time points across the evolution experiment. Taken together, our results support the classical prediction that selection in variable environments acts to increase geometric mean fitness, but also highlight evolutionary constraints that arise in environments where the fitness effects of a mutation are variable across time.

501

Changes in clock-gene expression, activity, and fecundity in the mosquito *Culex pipiens* f. molestus exposed to artificial light at night

Ann-Christin Honnen[1], **Janina L Kypke**[2], **Michael T Monaghan**[1]

1. Leibniz-Institute of Freshwater Ecology and Inland Fisheries (IGB), Berlin, BERLIN, Germany

2. Zoological museum, Natural History Museum of Denmark, Copenhagen, Denmark

Artificial light is a feature accompanying the vast majority of human settlements. Together with growing urbanisation it presents a global phenomenon of increasing importance.

Light stimuli are directly used to inform the circadian clock and are hence a key component of timekeeping. The internal clock regulates, directly or indirectly, a large number of physiological downstream processes, among them are daily activity and seasonal behaviours. Some of these systems are well studied, i.e. the circadian clock, and yet we lack knowledge on whether and how artificial light influences downstream, processes, such as daily movement patterns and reproductive output.

Some mosquito species, among them *Culex pipiens*, occur in close proximity to humans and artificial light at night becomes therefore part of their habitat. *Culex pipiens* is distributed globally and acts as a principal vector for West Nile Virus in many parts of the world.

We examined the effect of extended periods with artificial light on the mosquito *Culex pipiens* f. molestus by means of gene expression analysis of circadian clock genes, daily activity patterns and egg production. There are pronounced differences in the behaviour and the physiology of male and female mosquitoes. Therefore, we investigated sex specific differences as this will have implications at population level adding to effects present at an individual level.

All areas under study were influenced by our experimental set up and showed sex-specific responses indicating potentially large effects on population level. Gene expression was affected in four out of five analysed genes (mainly down-regulation) and activity was significantly reduced throughout the day. Egg production was also negatively affected suggesting that the forthcoming generation could also be influenced.

502

What can the gut microbiota community of eastern water dragons (*Intellagama lesueurii*) tell us about adaptation to urbanisation?

Bethan Littleford-Colquhoun, Celine Frere

Urbanisation is accelerating at unprecedented rates across the planet, causing rapid environmental change and posing a serious threat to global biodiversity. In order to minimise urbanisation's impact on biodiversity we require an increased understanding about how organisms are adapting to city life. Many studies have provided evidence that animals can, and are, adopting a number of novel strategies to living within the city with growing evidence that many species are changing their diet under urban conditions. However, we still know very little about the extent to which modifications in diet influence the gastrointestinal microbiota community of urban organisms, which is critical to an individual's health. Here, I will present results from isotopic analysis and gut microbiota community profiling across urban and non-urban populations of eastern water dragons to show how their diet, and in turn their microbiota, is adapting to city life.

503

Evidence for genetic adaptation to pollution

Alessandra AL Loria[1], Melania MEC Cristescu[1], Andrew AG Gonzalez[1]

1. McGill University, Montreal, QUEBEC, Canada

Adaptation to pollution has been studied intensively since the first observations of heavy metal tolerance in plants a few decades ago. In order to document micro-evolutionary changes and to fully demonstrate the occurrence of adaptation, researchers should, ideally, show evidence of a phenotypic change (i.e. resistance) with an underlying genetic basis, and a response to selection for tolerance to pollution. Moreover, it should be proven that the population has a positive growth rate. These aspects of the adaptive process can be considered evidentiary criteria for demonstrating adaptation and have been already used in relation to climate change. The purpose of this semi-quantitative review was to assess how well past studies meet these criteria and thus to provide a broad perspective of the extent of knowledge that we have on each step of the adaptive process in relation to soil and water pollution. We reviewed 253 articles published between 1992 and 2014 by searching *genetic adaptation to pollution* and *micro-evolution AND pollution OR contaminants* that focused on invertebrates, vertebrates (fish and amphibians) plants and algae. The studies were classified based on the fulfillment of the criteria. We found that the most compelling pieces of evidence come from studies focused on phenotypic responses (53%) and to a lesser extent, from studies focused on selective processes (almost 30%). However, only 6.7% of the studies fulfilled the first three criteria and only 10% included measurements of population fitness in their investigations. This, together with other potential bias (i.e. publication bias) could lead to an overestimation of the adaptive potential of populations to persist in polluted environments. Our study emphasizes the need to increase the number of studies that look at more than one evidentiary criteria and to increase the number of studies at the population level embracing approaches that integrate demography, ecology and evolution.

504

HMOX2 functions as a modifier gene for high-altitude adaptation in Tibetans

yi peng[1], deying yang[1], bing su[1]

1. kunming institute of zoology, china academy of science, Kunming, YUNNAN, China

Tibetans are well adapted to high altitude environments. Among the adaptive traits in Tibetans, the relatively low hemoglobin level is considered a bl unted erythropoietic response to hypoxic challenge. Previously, *EPAS1* and *EGLN1*, the major upstream regulators in the hypoxic pathway were re portedly involved in the hemoglobin regulation in Tibetans. In this study, we report a downstream gene (*HMOX2*) involved in heme catabolism, whic h harbors potentially adaptive variants in Tibetans. We first re-
sequenced the entire genomic region (45.6 kb) of *HMOX2* in Tibetans, which confirmed the previously suspected signal of positive selection on *HM OX2* in Tibetans. Subsequent association analyses of hemoglobin levels in two independent Tibetan populations (a total of 1,250 individuals) showe d a male-
specific association between the *HMOX2* variants and hemoglobin levels. Tibetan males with the derived C allele at rs4786504 displayed lower he moglobin level as compared to the T allele carriers. Furthermore, our *in vitro* experiments indicated that the C allele of rs4786504 could increase the expression of *HMOX2*, presumably leading to a more efficient breakdown of heme that may help maintain a relatively low hemoglobin level at high altitude. Collectively, we propose that *HMOX2* contributes to high altitude adaptation in Tibetans by functioning as a modifier in the regulation of he moglobin metabolism.

505

Evolutionary trajectory of a heat sensor TRPV1 in clawed frogs inferred from multispecies comparison and ancestral protein reconstruction

Shigeru Saito[1, 2], Claire T Saito[1], Toshio Ohta[3], Makoto Tominaga[1, 2]

1. Division of Cell Signaling, Okazaki Institute for Integrative Bioscience (National Institute for Physiological Sciences), Okazaki, Japan

2. SOKENDAI (The Graduate University for Advanced Studies), Okazaki, Japan

3. Faculty of Agriculture, Tottori University, Tottori, Japan

Species inhabiting in different thermal niches must have acquired thermal sensitivity suitable for the respective niches. Functional changes of thermal sensors could directly influence thermal perception, thus may have played crucial roles in thermal adaptation. Here we compared thermal responses of two species of clawed frogs (*Xenopus laevis* and *Xenopus tropicalis*) inhabiting different thermal niches. We first compared behavioral responses and found that *X. laevis* is much more sensitive to heat stimulation than *X. tropicalis*. Primary cultured sensory neurons also exhibited similar difference between the two species. Thus, we compared thermal responses of an ion channel TRPV1, which serves as a heat sensor, by electrophysiological experiments. Clear species difference in TRPV1 was observed with repeated heat stimulation. *X. laevis* TRPV1 exhibited almost full activity in the first heat stimulation and its responses gradually deceased with repeated heat stimulation (desensitization). On the other hand, *X. tropicalis* TRPV1 exhibited only a partial response in the first heat stimulation and its responses gradually increased (sensitization). In order to estimate the evolutionary trajectory of TRPV1 channel property, we then compared thermal responses of TRPV1 from three additional clawed frog species. TRPV1 from all three species exhibited desensitization property to repeated heat stimulation. To infer the ancestral states of TRPV1 channel properties, we reconstructed the TRPV1 ancestral proteins and examined their heat responses. Ancestral TRPV1 exhibited desensitization properties to repeated heat stimulation, indicating that the heat responses of TRPV1 changed from desensitization to sensitization in the lineages leading to *X. tropicalis*. Moreover, we identified three amino acid substitutions that are largely responsible for the species difference of TRPV1 heat responses. These results suggest that subtle amino acid substitutions can cause functional changes in thermal sensors and may have served as a driving force for the evolutionary changes in thermal perception.

507

Comparison of adaptive mechanism between sexual and asexual reproduction in *Tetrahymena thermophila* based on the experimental evolutionary genomics

Guangying Wang[1], Jie Xiong[1], Kai Chen[2], Wentao Yang[1], Wei Miao[1]

1. Institute of Hydrobiology, Chinese Academy of Sciences, Wuhan, Hubei, China

2. Northwest Normal University, Lanzhou, Gansu, China

The evolution and maintenance of sex is an important problem in evolutionary biology. The advantage of sexual reproduction was thought that it increases the rate of adaptation to the novel environment comparing to the asexual reproduction, but the molecular basis of this adaptive advantage and the similarities/differences between sexual and asexual reproduction are still unclear. *Tetrahymena thermophila*, as a classical unicellular model organism, has made outstanding contributions to the discovery of foundational biology, such as telomere, telomerase and RNA self-splicing. Using its advantages including possessing sexual and asexual stages during its life cycle, culture in the lab, complete genomic and transcriptomic database and techniques of omics and bioinformatics, we dynamically detect the changes of DNA, transcriptional regulation and genome organization during the continuous culture in the lab. Five parallel cell lines were set up. Currently, 1100 asexual fissions and 11 times sexual reproduction had been completed. The fitness tests (growth rate) indicated sexual populations grew significantly faster than asexual populations. Moreover, base on the re-sequencing genome/transcriptome data, we found sexual populations can produce much more genetic variation and accumulate more beneficial mutations than those of asexual populations. In addition, as for sexual group, all the populations had losed mating ability after at most six rounds of sexual reproduction because of the purification of mating types, which resulted in them switching from sexual reproduction to asexual reproduction. Followed by this switching, the fixation of beneficial mutations produced by sex was completed in short 100-200 fissions through asexual reproduction.

508

HIV epitope prediction provides functional link between HLA genotype and viral load

Jatin Arora[1], Paul J. McLaren[2], István Bartha[3, 4], Nimisha Chaturvedi [3, 4], Jacques Fellay[3, 4], Tobias Lenz[1]

1. Department of Evolutionary Ecology, Max Planck Institute for Evolutionary Biology, Plön, Germany

2. National HIV and Retrovirology Laboratory, Public Health Agency of Canada, JC Wilt Infectious Diseases Research Center, Winnipeg, Canada

3. Global Health Institute, School of Life Sciences, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland

4. Swiss Institute of Bioinformatics, Lausanne, Switzerland

Human Leukocyte Antigen (HLA) class I genes mediate the cytotoxic T-cell response against HIV by presenting viral epitopes at the surface of infected cells. Due to polymorphism in their binding groove, individual HLA alleles bind to different arrays of epitopes, resulting in differential responses to HIV infection; e.g. HLA-B*57:01 confers protection while HLA-B*35:02 confers susceptibility to HIV progression. Research over the last decades has robustly established a strong association between HLA variants, particularly in HLA-B, and spontaneous HIV control as well as progression to AIDS. However, the functional basis for this association and more specifically the extent to which HLA-bound HIV epitopes explain this association are largely elusive. Using an unprecedented clinical dataset of 6,311 chronically HIV infected patients we computationally predicted the binding affinities of each patient's HLA class I alleles for all possible HIV epitopes. We show that set point viral load (spVL), an established correlate of HIV disease progression, is negatively associated with the breadth of the epitopes bound by a patient's HLA alleles. In linear regression models, specific HLA-bound epitopes explain more variation in spVL than the established HLA genotype associations. The model was further improved when HLA-bound epitopes were predicted from autologous HIV sequences, available for a subset of the patients, rather than from the HIV reference sequence. In conclusion, our findings provide a novel functional explanation for the well-established association between HLA class I variation and HIV control. The specific epitopes that we found to have strong impact on spVL could provide new targets for antiretroviral drugs.

509

The role of LINE-1 retrotransposition in Parkinson's disease

Gabriela-Oana Bodea[1], Geoffrey J Faulkner[1, 2]

1. Mater Research Institute-University of Queensland, Brisbane, QLD, Australia

2. Queensland Brain Institute, University of Queensland, Brisbane, QLD, Australia

Parkinson's disease (PD) is a complex neurodegenerative condition that affects more than 7 million people worldwide and often occurs after the age of 65. The main hallmark of PD is the selective loss of dopaminergic neurons from substantia nigra, which control voluntary movement. Despite recent advances, current PD treatments only ameliorate symptoms but do not prevent neuronal loss and cannot cure the disease. PD aetiology is multifactorial, with genetic and environmental factors interacting via as yet unclear mechanisms to induce PD pathology. Recent studies have proposed that environmental and genetic factors may trigger hyperactivation of DNA mobile elements. These elements can alter the genome by insertional mutagenesis, recombination and deletion, potentially contributing to the susceptibility and pathophysiology of neurological disorders. Long interspersed element-1 (L1) is the only active and autonomous mobile element in the human genome, and accounts for about 17% of human DNA. L1 is active in somatic cells and can 'jump' from one place in the genome to another by first copying itself into RNA and then reversing the process, thus potentially altering the activity of genes were they relocate. We aim to investigate the role of L1 activity at the intersection of environmental and genetic factors known to contribute to PD aetiology and deepen our understanding of how PD develops. Currently, we are establishing the core parameters of L1 mobilisation in PD and also test whether these somatic variants are likely to alter dopaminergic neuron phenotype. To achieve these aims, the project is using imaging techniques in a mouse model combined with genomics in mouse and human samples.

510

Food and pathogen adaptations: tracing the spread of lactase persistence and human African trypanosomiasis resistance into southwestern Africa

Joana C Pinto[1], Sandra Oliveira[1, 2], Sérgio Teixeira[1], Dayana Martins[3], Anne-Maria Fehn[4], Teresa Aço[5], Magdalena Gayà-Vidal[1], Jorge Rocha[1, 2]

1. CIBIO-InBIO- Research Center in Biodiversity and Genetic Resources, Vairão, Portugal

2. Departamento de Biologia, Faculdade de Ciências, Universidade do Porto, Porto, Portugal

3. ISCED/Huíla: Instituto Superior de Ciências da Educação, Lubango, Angola

4. Max Planck Institute for the Science of Human History, Jena, Germany

5. CEDO: Centro de Estudos do Deserto, Namibe, Lubango, Angola

Introduction: We investigated the frequency distribution and haplotype diversity of APOL1 and LCT variants associated with human African trypanosomiasis (HAT) resistance and lactase persistence (LP), respectively, in populations from southern Angola to trace the spread of these genetic adaptations into southwestern Africa.

Materials and Methods: We resequenced two fragments of the LCT-enhancer and the APOL1 gene and genotyped flanking STRs in six groups from the Angolan Namib with different subsistence traditions, and in other populations from Africa and Europe for comparative purposes. The age and selection coefficient of these variants were estimated.

Results: LP in the Angolan Namib is represented by the -14010*C allele, which is associated with a predominant haplotype shared with other southern and eastern African populations. While LP was more frequent in foragers than in pastoralists, the frequencies of the two APOL1 variants (G1 and G2) did not differ between the two groups. The G1 allele is mostly associated with a single widespread haplotype. The G2 allele is linked to several haplotypes that are related to haplotypes found in African Bantu-speaking populations. The putatively archaic G3 variant displayed more intra-allelic diversity in Africa than in Europe.

Conclusions: The LP adaptation was carried to southern Africa from eastern Africa, probably by non-Bantu speaking pastoralists, although we could not confirm a direct link with groups speaking Khoe-Kwadi family languages. The presence of APOL1 variants G1 and G2 is linked to the Bantu expansions. Our results suggest that the G3 variant was retained in modern humans by incomplete lineage sorting.

511

Insights on Array Design and Genotyping of over 50,000 Diverse Individuals for the Next Generation of Association Studies

Christopher R Gignoux[1], Genevieve L Wojcik[1], Henry Rich Johnston[2], Christian Fuchsberger[3], Suyash S Shringarpure[1], Alicia R Martin[1], Stephanie Rosse[4], Niha Zubair[4], Daniel Taliun[3], Ryan Welch[3], Carsten Rosenow[5], Noura S Abul-Husn[6], Gillian Belbin[6], Hyun M Kang[3], Goncalo Abecasis[3], Michael Boehnke[3], Zhaohui S Qin[2], Christopher Carlson[4], Kathleen C Barnes[7], Carlos D Bustamante[1], Eimear E Kenny[6], on behalf of the PAGE Network[8]

1. Genetics, Stanford University, Stanford, CA, USA

2. Biostatistics and Bioinformatics, Rollins School of Public Health, Emory University, Atlanta, GA, USA

3. Biostatistics, School of Public Health, University of Michigan, Ann Arbor, MI, USA

4. Fred Hutchinson Cancer Research Center, Seattle, WA, USA

5. Illumina, Inc, San Diego, CA, USA

6. Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA

7. Medicine, Johns Hopkins University, Baltimore, MD, USA

8. Population Architecture using Genomics and Epidemiology Consortium,

We have seen numerous successes in genomewide association studies (GWAS) underlying complex traits over the past decade. However much of this work has only been performed in populations of European descent. To address this disparity, we developed the Multi-Ethnic Genotyping Array (MEGA), a single platform designed for balanced GWAS coverage across the globe incorporating a catalog of functional variation.

To maximize trans-ethnic utility we designed the GWAS backbone to be informed by whole genome sequences across 26 populations of the 1000 Genomes Project and be bolstered by tag SNPs from 642 high-coverage whole genomes from individuals of African descent in the CAAPA consortium. We developed a novel cross-population tag SNP selection strategy to capture low frequency variants across the diverse populations in Phase 3 of the 1000 Genomes Project (TGP). Importantly, by optimizing imputation accuracy rather than pairwise LD, the performance of the array is high across all continental TGP super-populations (>90% imputation accuracy for MAF >=1%). We deconvolved admixture to evaluate per-ancestry imputation performance, and devised a whole genome sequencing panel to balance existing reference datasets. A reference panel of several thousand individuals, including the Human Genome Diversity Panel and a large panel of indigenous Americans, will be available on MEGA to aid in rare variant calling, ancestry characterization, and admixture analyses.

Currently we have genotyped >50,000 African-American, Hispanic/Latino, Asian American and Native American and Hawaiian individuals from PAGE cohorts. From these diverse populations we can infer an extraordinary breadth of population structure, admixture, and differential relatedness with important implications for complex trait association studies within and across ethnicities. Here, we highlight the need for methods tothat can capture and model such high levels of diversity, both to optimize statistical power and improve biological interpretation.

512

Understanding schizophrenia: the role of LINE-1 retrotransposition in disease aetiology

Marie-Jeanne H.C. Kempen[1], Gabriela-Oana Bodea[1], Sandra Richardson[1], Francisco Sanchez Luque[1, 2], Chunyu Liu [3], Elliot Gershon[3], Geoffrey J. Faulkner[1, 4]

1. Mater Research Institute - University of Queensland, Brisbane, QLD, Australia

2. Pfizer-Universidad de Granada-Junta de Andalucia Centre for Genomics and Oncological Research (Genyo). P.T. Ciencias de la Salud, Granada, Spain

3. Department of Psychiatry and Behavioral Neuroscience, The University of Chicago, Chicago, Illinois, USA

4. Queensland Brain Institute - University of Queensland, St. Lucia, Queensland, Australia

Long interspersed nuclear element 1 (LINE-1 or L1) is an autonomous retrotransposon that uses a "copy and paste" mechanism to retrotranspose in mammalian genomes. L1 makes up ~17% of the human genome and occur most frequently in the brain. L1 mobilization has the potential to impact the genome structure and alter gene activity. Environmental factors, including stress, were found to enhance L1 activity, suggesting a potential role for L1 in the pathophysiology of disorders in which gene-environment interactions play a major role such as Schizophrenia (SCZ). Recently, Bundo et al. (2014) found an increased copy number variation (CNV) of L1 in the cortex of SCZ patients, compared to controls, as well as in the cortex of two animal models of SCZ. However, the mechanism and extent of L1 mobilization impacting the SCZ phenotype remains largely unexplored. The present study aims to deepen our understanding of how L1 contributes to the SCZ using cutting edge genomics and molecular biology approaches. Firstly, human SCZ and control samples were subjected to retrotransposon capture sequencing, a novel high-throughput sequencing approach, to identify the genomic location and structural characteristics of L1 insertions. Using this technique, we were able to distinguish putative somatic, polymorphic and fixed L1 insertions. Further, we analysed the L1 content of different brain regions by means of qPCR L1 CNV assay in a SCZ mouse model which recapitulates the influence of environmental factors, such as viral infection, *in utero* on SCZ development. Developmental timing of *in utero* exposure to such environmental factors has been shown to differentially impact SCZ symptoms. We are currently investigating whether the established distinction in SCZ phenotype can be related to changes in L1 levels. In this way we aim to shed more light on the involvement of L1 in the development of SCZ symptoms.

513

Effects of rare gene knockouts in a highly endogamous population

Hilary C Martin[1], Vagheesh M Narasimhan [1], Karen A Hunt[2], Dan Mason[3], Christopher L Baker [4], Konrad J Karczewski [5], Eamonn R Maher[6], Richard Trembath[7], Daniel G MacArthur [5], John Wright[3], David A van Heel[8], Richard Durbin[1]

1. Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, United Kingdom

2. Blizard Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London, United Kingdom

3. Bradford Institute for Health Research, Bradford Teaching Hospitals NHS Foundation Trust, Bradford, United Kingdom

4. Center for Genome Dynamics, The Jackson Laboratory, Bar Harbor, USA

5. Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, USA

6. Department of Medical Genetics, University of Cambridge, Cambridge, United Kingdom

7. Faculty of Life Sciences and Medicine, King's College London, London, United Kingdom

8. Blizard Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London, United Kingdom

Studying individuals who are complete knockouts for a gene can be informative about its function. We exome-sequenced 3222 British Pakistani adults with high parental relatedness from Birmingham and the Born-in-Bradford study, and discovered rare homozygous loss-of-function (rhLOF) mutations in 781 genes, a substantial enrichment compared to discovery rates in outbred populations. We observed a 13.7% depletion of homozygous knockout genotypes, implying selection against deleterious recessive variants, and estimated that each adult has an average load of 1.6 recessive loss-of-function lethal-equivalent variants. We also observed a much smaller but significant depletion of homozygous missense variants predicted to be highly deleterious. Our findings emphasize the value of cataloguing human knockouts in healthy adults; we found 52 genes with rhLOFs which are lethal when knocked out in mice, and 32 which have been reported to cause Mendelian disease but for which the carriers of the rhLOF do not fit the reported phenotype. Additionally, the discovery of a healthy *PRDM9*-knockout mother demonstrated unexpected redundancy of this gene in humans, highlighting how the discovery of natural LoF variants can help characterise gene function. We found individual gene knockouts significantly associated with lipid and glycaemic traits. If validated in a larger cohort, these associations could point to potential new drug targets. Finally, we are expanding this study as part of the East London Genes and Health project. We aim to exome-sequence 25,000 healthy British South Asian adults (enriched for consanguinity) with linked electronic health records, in whom we expect to find ~4,800 unique genes with rhLOFs if the individuals all have an inbreeding coefficient of 0.0625. Additionally, we are identifying individuals with high autozygosity in the UK Biobank for sequencing. We will present preliminary results from these projects, which are expected to greatly expand the catalogue of human knockouts, shedding light on gene essentiality.

514

Target enrichment as a tool for wide-scale genome-wide comparison, intra-strain variation and stratification, in genome-integrated Human Herpesvirus 6

Marco Telford[1], Gabriel Santpere[2], David Hughes[1], Arcadi Navarro[1, 3, 4, 5]

1. Pompeu Fabra University, Barcelona, SPAIN, Spain

2. Department of Neurobiology, Yale University, New Haven, CT, USA

3. Centre de Regulació Genòmica (CRG), Barcelona, Spain

4. National Institute for Bioinformatics (INB), Barcelona, Spain

5. Institució Catalana de Recerca I Estudi Avançats (ICREA), Catalonia, Spain

The two strains of Human herpesvirus 6 (HHV-6A and HHV-6B) infect >90% of the adult human population worldwide, and are linked with an increasing number of central nervous system and blood pathologies. These associations are debated because of the difficulty of sequencing the viruses. This led to very few complete published genomes, and lack of information on its intra-strain variation. The possible geographic stratification is of particular relevance, since many of the proposed associations with the virus are with diseases that show geographical patterns themselves.

Starting from the 1000 Genome Project data we scanned for HHV-6 in its genome-integrated form, and found 9 infected individuals. We retrieved biological samples from those individuals, and performed kit-based target enrichment and sequencing. This approach allowed us to obtain deep-coverage sequences to analyse variability and stratification.

The two viral strains show significantly different genome-wide variability, and different variability patterns along the genomes and among the different genomic features. HHV-6A sequences show clear separation of an Asian subgroup compared to the virus from individuals of others geographical origins. HHV-6B seems to have poor stratification, unless recombination is taken into consideration, which allows for an African and a European subgrouping to become detectable.

The overall results show that HHV-6A and B may have, as their sister taxa, geographical stratification. However, the resolution is low due to the low sample size. A higher number of samples, and thus a higher resolution, will allow us to create a better study design for disease association studies on these viruses, and shed light on this controversial field. We have therefore developed an in-house protocol that will allow us to perform target enrichment at low cost, giving us the possibility to strongly improve our data set.

515

Bayesian multivariate analysis of large genetic studies identifies novel associations

Michael C Turchin[1], **Matthew Stephens**[1, 2]

1. *Department of Human Genetics, University of Chicago, Chicago, IL, United States*
2. *Department of Statistics, University of Chicago, Chicago, IL, United States*

Genome-wide association studies (GWAS) are now a common tool among human geneticists to identify genetic variants that significantly affect traits of interest. To date, the NHGRI GWAS Catalog has over 24,000 SNP-phenotype associations. However, the vast majority of these GWAS are conducted in univariate frameworks, ie when genetic variants are only tested against a single phenotype one at a time. This is in contrast to multivariate frameworks where genetic variants are tested against different combinations of traits simultaneously. Multivariate frameworks are of interest because it is well known that under certain biological scenarios these approaches significantly increase power. Additionally, by testing combinations of traits, researchers are able to investigate more complex biological hypotheses. Despite these clear advantages though, there are often recurring reasons why multivariate analyses are not conducted. Univariate GWAS already involve a large computational and statistical burden; performing an extra, exponentially greater number of tests appears highly intractable. Furthermore, it is often unclear how to properly compare different multivariate models even when they can be efficiently conducted. Here, we present a framework and R package that aims to alleviate these obstacles -- Bayesian multivariate analysis of association studies, or bmass. bmass runs on univariate GWAS summary statistics and can quickly conduct all possible multivariate analyses given a set of up to 8 phenotypes. bmass also provides Bayes factors for each multivariate analysis, thus allowing models to be directly compared. Running bmass on various publicly available GWAS datasets consistently show an increase in power up to 40% over univariate approaches while keeping FDRs as low as 15%. bmass also provides novel biological insight at a more intricate level than previously seen, revealing phenotypic combinations that often drive signals of genetic associations. Overall, bmass is a powerful and tractable tool that now allows researchers to effectively conduct multivariate GWAS.

516

Mitochondrial heteroplasmy in Australian hylaeine bees and its association with the parasite *Wolbachia*

Olivia Davies[1], **Michael Gardner**[1], **Mark Stevens**[2], **Michael Schwarz**[1]

1. *School of Biological Sciences, Flinders University, Adelaide, SA, Australia*
2. *South Australian Museum, Adelaide, South Australia, Australia*

Mitochondrial heteroplasmy is the existence of two or more mitogenomes within a single individual. Generally, it is considered to be a rare, transient condition in natural populations that arises due to spontaneous mutations in mitochondrial DNA. However, using Sanger sequencing, cloning and NGS we have identified widespread heteroplasmy in an Australian hylaeine bee, *Amphylaeus morosus*. Not only was every sequenced individual heteroplasmic from its southern-most to northern-most latitudinal limit (Victoria to southern Queensland), but every individual possessed the same two mitochondrial haplotypes, with no variation between individuals despite the large range of this species. A consistent, double infection of the reproductive parasite *Wolbachia* was also identified in these bees. *Wolbachia* has recently been receiving a lot of attention for its ability to affect sex-determination of its hosts and how this mechanism can be harnessed in the application of a bio-control agent. We discuss the possibility that the two conditions in *A. morosus*, mitochondrial heteroplasmy and a double *Wolbachia* infection, are not coincidental in this species, and whether mitochondria heteroplasmy may be driven by the persistent double infection. Hylaeinae is a highly diverse sub-family of bees that have extraordinary rates of species radiation within Australia and throughout the world. We suggest that this genus may be a good candidate for examining the role of *Wolbachia* in speciation and also the maintenance of intra-specific mitochondrial diversity.

517

Investigating the evolutionary pathways towards extremely AT rich mitochondrial genomes

Weilong Hao[1]

1. *Wayne State University, Detroit, MI, USA, United States*

Genomic nucleotide content (usually measured as GC content) varies widely among species, the most extreme (AT-rich) of which is in the mitochondria of yeasts. These extreme genomes provide a unique opportunity to study the evolution of genomic nucleotide landscape. In this study, we sequenced six complete mitogenomes of the *Saccharomycodes ludwigii* yeast, all of which have <10% GC content. Our comparative genomics analyses observed variable intron presence/absence patterns in the large ribosomal subunit (*rnl*) gene and cytochrome c oxidase subunit I (*cox1*) gene, and variable lengths of AT-rich tandem repeats. The whole genome alignments among these mitogenomes showed mosaic sequence patterns, suggesting perhaps frequent mitochondrial DNA recombination. We found no evidence of accelerated substitution rates in these *Saccharomycodes ludwigii* mitogenomes, when compared against other yeast mitogenomes. Thus, mutational pressure and reduced recombination, both

of which can lead to increased AT content, are unlikely the main driving force leading to the extreme AT mitogenomes in *Saccharomycodes ludwigii*. We tend to believe that the proliferation of AT-rich tandem repeats via replication slippage and/or unequal crossing-over plays an important role in driving the extreme AT richness in these mitogenomes.

518

Mitochondrial membrane potential: a trait involved in organelle inheritance?

Liliana Milani[1]

1. University of Bologna, Bologna, Italy

How is the delivery of functional mitochondria across generations achieved? In Drosophila, mitochondria driven by microtubules reach an evolutionarily conserved structure of the egg, the Balbiani body (Bb), which supplies mitochondria to the primordial germ cells (PGCs) of the new individual. In zebrafish, a selective accumulation of mitochondria with high inner membrane potential (Δψm) in the Bb was recently documented: the presence of high Δψm would indicate mitochondrial genome integrity and allow Bb mitochondria to be preferentially transported through the microtubule network and inherited. These are examples of what happens when mitochondria are transmitted through the egg. Mitochondria carried by sperm are commonly not inherited, but spermatozoa have high energy demand for swimming support, and numerous studies on different organisms report that sperm mitochondria have high Δψm.

So, how are mitochondrial activity and segregation linked? Based on what observed across multiple taxa, I propose that Δψm determines which mitochondria reach the PGCs, and how: the more ATP produced, the higher chance to be transported. In animals with an early germ line specification (preformation), the material determining the cell fate is sequestered early on into gonadic presumptive blastomeres along with the most active mitochondria, and it can happen them being sperm mitochondria. This hypothesis is supported by observations in some bivalve species, in which mitochondria are stably transmitted through sperm because of their early delivery into PGCs during embryonic divisions through microtubule dynamics. Instead, when germ line specification happens at a later stage of development (epigenesis), spermatozoon mitochondria would have been already degraded when germ cell precursors form.

In summary, Δψm can be a simple and effective system allowing the most active mitochondria to reach specific locations, but the different timing of action of germ line specification influences the outcome of the segregation mechanism.

519

Detection of ultra-rare mitochondrial variants

Anton Nekrutenko[1]

1. Penn State, University Park, PA - PENNSYLVANIA, United States

Duplex sequencing was originally developed to detect rare nucleotide polymorphisms normally obscured by the noise of high-throughput sequencing. While the experimental portion of duplex sequencing requires robust molecular biology expertise, it is well developed, leaving the data analysis portion of the procedure lagging. Here we describe a new, greatly streamlined, reference-free approach for the analysis of duplex sequencing data. Upon ensuring that the approach precisely reproduces previously published results, we applied it to a newly produced dataset, enabling us to type low-frequency variants in human mitochondrial DNA. Finally, we attempted to democratize the data analysis for duplex sequencing by providing all necessary tools as stand-alone components as well as integrating them into the Galaxy platform. All analyses performed in this manuscript can be repeated exactly as described at http://usegalaxy.org/duplex.

520

Mitochondrial selfish elements and the evolution of biological novelties

Liliana Milani[1], **Fabrizio Ghiselli**[1], **Marco Passamonti**[1]

1. University of Bologna, Bologna, BO, Italy

RPHM21 is a recently discovered protein encoded by a male-specific mitochondrial genome, and with a putative role in the paternal inheritance of sperm mitochondria in the Manila clam *Ruditapes philippinarum*, a species characterized by the doubly uniparental inheritance of mitochondria (DUI). The available evidence suggests a viral origin of RPHM21 and supports its activity during spermatogenesis: RPHM21 is progressively accumulated in mitochondria and nuclei of spermatogenic cells, and we hypothesize it can influence mitochondrial inheritance and sexual differentiation.

We propose a testable model that describes how the acquisition of selfish features by a mitochondrial lineage might have been responsible for the emergence of DUI, and for the evolution of separate sexes (gonochorism) from hermaphroditism.

The appearance of DUI in a species most likely entailed the invasion of at least one selfish element, and the extant DUI systems can be seen as resolved conflicts. It was proposed that hermaphroditism was the ancestral condition of bivalves, and a correlation between DUI and gonochorism was documented. We hypothesize that DUI might have driven the shift from hermaphroditism to gonochorism, with androdioecy as a transition state. The invasion of sex-ratio distorters and the evolution of suppressors can prompt rapid changes among sex-determination mechanisms, and DUI might have been responsible for one of such changes in some bivalve species. If true, DUI would represent the first animal sex-determination system involving mtDNA-encoded proteins.

521

SmithRNAs: could mitochondria 'bend' nuclear regulation?

Andrea Pozzi[1], **Fabrizio Ghiselli**[1], **Marco Passamonti**[1]

1. Dipartimento di Scienze Biologiche Geologiche e Ambientali (BiGeA), University of Bologna, Bologna, Italy

The genome content of metazoan mtDNA is not limited to 37 genes. Besides additional genes and other unusual features, small non coding RNAs (sncRNA) transcribed by mtDNA have been also found, but their occurrence and role are poorly understood. Are mitochondria able to influence nuclear gene expression through these sncRNAs? We sequenced small RNA libraries from gonadal tissues and isolated mitochondria of *Ruditapes*

philippinarum, a species characterized by the Doubly Uniparental Inheritance (DUI) of mitochondria. This clam has two highly different sex-linked mtDNAs, each inherited uniparentally: one through females (F-type) the other through males (M-type). Since males are heteroplasmic for both types, and females are homoplasmic for the F-type, DUI provides an internal control without using Rho 0 cells, at least for sncRNAs transcribed by the M-type. We identified several putative sncRNAs of mitochondrial origin and we predicted their targets *in silico*. The most transcribed sncRNA we found is also differentially expressed in males and females, and its predicted target is the 3' UTR of the Nuclear Receptor Subfamily 0, known to be involved in sex determination of many animals. Actually, a role of mitochondria in sex determination has been hypothesized in the DUI system. Furthermore this sncRNA is transcribed by a non-coding mtDNA region that can be folded in a stable stem-hairpin structure, which makes it a good candidate for functional studies. Many other sncRNAs of mitochondrial origin are likely to be discovered soon, and we here propose to name them 'small mitochondrial highly transcribed RNAs' (smithRNAs).

1. Passamonti and Ghiselli, Doubly uniparental inheritance: two mitochondrial genomes, one precious model for organelle DNA inheritance and evolution. DNA Cell Biol. 28, 79-89 (2009).
2. Ro S. et al. The mitochondrial genome encodes abundant small noncoding RNAs. Cell Res 23, 759–74 (2013).
3. Morin R. D. et al. . Application of massively parallel sequencing to microRNA profiling and discovery in human embryonic stem cells. Genome Res 18, 610–621 (2008).

522

Evolutionary Dynamics of the Mitochondria in Dwarf and Giant Rattlesnakes

Sarah Schaack[1], Jesse Meik, Jeffrey Streicher, Longson Pang, Daniel Dashevsky, Michael Song

1. Reed College, Portland, OR, United States

In a continued effort to understand the rapid evolution of insular dwarfism and gigantism, we have sequenced the complete mitochondrial genomes of 14 specimens of *Crotalus mitchellii*, the Speckled Rattlesnake, from across its range. The *Crotalus* species complex is distributed throughout the SW US and NW Mexico, including a number of islands in the Sea of Cortez where mean body size ranges ten-fold between populations on the smallest and largest islands. Although there may be ecological explanations for some part of this dramatic phenotypic divergence, initial sequencing efforts of three loci revealed unexpected phylogenetic relationships among populations of giants and dwarves. To further explore the puzzling phylogeographical pattern within the complex, as well as to identify the genetic underpinnings of a rarely observed case of rapid body size evolution in vertebrates, we sequenced the whole genome and analyzed the locus-specific substitution rates and phylogenetic relationships among mitochondrial genes. Given the critical role of mitochondria in metabolism and growth, this genome is of particular interest in this study system. Substitution rates vary substantially among genes and individuals and there is a signature of hybrid introgression between distant mainland populations. Both the power and risk of using mitogenomic datasets for understanding short- and long-term dynamics in this species complex, as well as others, will be discussed.

523

Splits and merges of minichromosomes shaped the complex and dynamic mitochondrial genome organization of the sucking lice (Anoplura, Insecta)

Renfu Shao[1], Hu Li[2], Stephen C Barker[3], Simon Song[1]

1. Genecology Research Centre, Faculty of Science, Health, Education and Engineering, University of the Sunshine Coast, Sippy Downs, QLD, Australia

2. Department of Entomology, China Agricultural University, Beijing, China

3. School of Chemistry and Molecular Biosciences, University of Queensland, Brisbane, QLD, Australia

Fragmented mitochondrial (mt) genomes have been reported for 11 species/subspecies of sucking lice (suborder Anoplura) that infest humans, chimpanzees, pigs, horses and rodents. There is substantial variation among these lice in mt karyotype: the number of minichromosomes of a species/subspecies ranges from 9 to 20; the number of genes in a minichromosome ranges from 1 to 8; arrangement of genes in a minichromosome differs between species, even in the same genus. We sequenced the mt genome of the guanaco louse, *Microthoracius praelongiceps*, which is from a major branch of sucking lice not sampled previously. We aim to establish the ancestral mt karyotype for sucking lice, and to understand how mt genome organization evolved in sucking lice.

The 37 mt genes of the guanaco louse are on 12 minichromosomes; each minichromosome is 2,274 to 2,940 bp in size, and has 2 to 5 genes and a non-coding region. The guanaco louse shares many features with the rodent lice in mt karyotype, more than with other sucking lice. Phylogenetic analysis of mt genome sequences, however, showed that the guanaco louse is more closely related to the human lice, chimpanzee lice, pig lice and horse lice than to the rodent lice. By analysis of shared features in mt karyotype, we infer that the most recent common ancestor of sucking lice, which lived ~75 MYA, had 11 minichromosomes; each minichromosome has 1 to 6 genes and a non-coding region. As sucking lice diverged, split of mt minichromosomes occurred many times in the lineages leading to the lice of humans, chimpanzees and rodents whereas merge of minichromosomes occurred in the lineage leading to the lice of pigs and horses, resulting in a complex and dynamic mt genome organization not seen in any other animals.

524

Mitochondrial DNA genetic diversity within Australian domestic pigs and their role in determining reproductive capacity.

Tesha Tsai[2, 1], Justin St. John[2, 1]

1. Department of Molecular and Translational Science, Monash University, Melbourne, Victoria, Australia

2. Hudson Institute of Medical Research, Clayton, VIC, Australia

Mitochondrial DNA (mtDNA) present in the metaphase II oocyte serves as a template for all mtDNA that is replicated during development and distributed amongst all tissues. In mammals, mtDNA is strictly maternally inherited and individuals with the same lineage cluster into groupings known as mtDNA haplotypes. In livestock, certain mtDNA haplotypes are associated with improved milk and meat quality, whilst other species have shown increased longevity, growth and susceptibility to diseases. In this work, we set out to determine the degree of mitochondrial genetic diversity within the Australian domestic pig population, and to find whether different maternal lineages influence reproductive capacity. Five mtDNA haplotypes (A to E) were identified from sequencing the D-loop region of 368 pigs. Whilst A, B and C were of Asian origin, D and E were of European origin. We then asked whether mtDNA haplotypes influence oocyte maturation, fertilization and development to blastocyst. We found that

there were significant differences for maturation and fertilization rates amongst the haplotypes. Moreover, we found that haplotypes C, D and E produced significantly larger litters. To determine the relationship between mtDNA sequence variants and reproductive capacity, we performed next generation sequencing. Amongst the mtDNA haplotypes, the number of mtDNA variants harbored at >25% correlated with oocyte quality. MtDNA copy number for developmentally competent oocytes positively correlated with the level of the 16383delC variant. This variant is located in conserved sequence box II, which is a regulatory region for mtDNA transcription and replication. We conclude that the genetic diversity within Australian domestic pigs is restricted, and that mtDNA haplotypes affect reproductive capacity.

525

Insight into the evolution of the partitioned mitochondrial genome of Myxozoa (Metazoa, Cnidaria).

Dayana Yahalomi[1], Arik Diamant[2], Huchon Dorothee[1, 3]

1. Zoology, George S. Wise Faculty of Life Sciences, Tel Aviv University, Ramat Aviv , Israel

2. National Center for Mariculture, Israel Oceanographic and Limnological Research, Eilat, Israel

3. The Steinhardt Museum of Natural History and Israel National Center for Biodiversity Studies, Tel Aviv University, Ramat Aviv , Israel

Myxozoans form a large group of poorly characterized parasites, which are nested within cnidarians. Complete myxozoan mitochondrial (mt) genomes have been characterized only from two genera: *Kudoa* and *Enteromyxum*. Species from the genus *Kudoa,* possess a rather standard mt genome: a single circular mt chromosome of ~18 kb. In contrast, *Enteromyxum leei*, possesses a fragmented genome divided into seven circular chromosomes of ~23 kb, making it the largest described animal mt genome. Each *Enteromyxum* chromosome harbors one coding gene region and a large non-coding region (~15 kb), nearly identical between chromosomes. To better understand the evolution of the partitioned mt genome in Myxozoa we sequenced the complete mt genome of another myxozan genus, *Sphaeromyxa*.

Preliminary results indicate that the mt genome of *Sphaeromyxa zaharoni* is organized unlike any hitherto sequenced myxozoan genome. It is formed of two circular chromosomes (~15 kb), which share the same non-coding region (~ 700 bp). Interestingly, these two chromosomes were found to recombine to form a single large (~30 kb) chimeric circular molecule, which includes two copies of the non-coding region.

Since *Sphaeromyxa* is known to have diverged before *Enteromyxum* and *Kudoa*, our results suggest that genome fragmentation either occurred several times independently or that the presence of a single molecule is a secondary character in *Kudoa*. These observations highlight the remarkable evolutionary plasticity of mt genome organization within Myxozoa.

526

Population differentiation and local adaptation in the Australian house sparrow.

Samuel C. Andrew[1], Simon C. Griffith[1], Lee A. Rollins[2]

1. Biological Sciences, Macquarie University, North Ryde, NSW, Australia

2. School of Life and Environmental Sciences, Deakin University, Geelong, VIC, Australia

The house sparrow (*Passer domesticus)* was introduced to Australia in the 1860's by Acclimatisation societies. Over the next hundred years the species spread across all of eastern Australia from multiple introduction points. Because the house sparrow is an obligate commensal species, they have jumped from settlement to settlement creating a fragmented meta-population across the sparsely populated areas of Eastern Australia. Independent founder events and successive bottle necks as the species has jumped from one settlement to the next, has resulted in genetic population structure. We have used microsatellite data to show the genetic differentiation across the Australian distribution (26 sample populations, n = 1248). This population structure can mostly be explained by founder effects, Isolation by Distance and independent human introductions. However, the effects of selection on population differentiation have never been tested in this species across a broad climate range. Our findings relating to genetic and morphological differentiation are consistent with results on this species in North America, South America and Europe. We are expanding on this work by using a landscape genomics approach (genome wide SNP data) to identify the effects of selection on functional loci, potentially resulting from local adaptation to the varied climate conditions across the Australian meta-population. Multiple introduction events allow us to validate our results with repeated observations to compare the effects of demographic processes and selection. Ultimately we aim to identify molecular mechanisms that make this species one of the best avian climate generalists.

527

Tracing the presence of an enzyme essential for de-novo biosynthesis of NAD in the avian lineage: A case study for missing sequences in bird genomes

Toni I Gossmann[1], Ines Heiland[2], Falk Hildebrand[3], Marc Niere[4], Veronika N Laine[5], Nicola Hemmings[1], Mathias Ziegler[4]

1. APS, University of Sheffield, Sheffield, UK

2. Arctic University of Norway, Tromsø, Norway

3. EMBL, Heidelberg, Germany

4. University of Bergen, Bergen, Norway

5. NIOO-KNAW, Wageningen, Netherlands

The crucial role of NAD as a cofactor in redox reactions has long been known, but in recent years the involvement of this dinucleotide in a multitude of important regulatory processes and its direct medical implications for cancer and diabetes has led to a vast number of NAD-related studies in molecular biology and medical research. Quinolinate (QA) phosphoribosyl transferase (QAPRT) is the enzyme that is essential for the de-novo synthesis of NAD. In contrast to other vertebrate species, to date there is no evidence for the presence of QAPRT in any bird, even though all required up- and downstream enzymes of the NAD metabolic network are present, as deduced from available avian genomes. There are at least two hypotheses that could explain this observation. (1) QAPRT encoding genes are absent from avian genomes - and QAPRT function is compensated either directly by another enzyme or absent due to regulatory plasticity. (2) QAPRT has not been identified in the avian lineage because it is encoded in a genomic region that is difficult to assemble. To address the latter hypothesis, we used available bird whole-genome and transcriptome resources to extensively scan for sequences potentially coding for QAPRT. We found evidence that indicate the genomic presence and expression of the gene. We validated our results by an *in-vitro* functional essay. We further identified more than two thousand human genes for which no counterpart in most bird genomes has been identified to date and show that these genes are characterized by an extreme GC sequence composition. Taken together these results suggest that there is an annotation bias of bird genomes, potentially caused by the heterogeneous karyotype and sequence features common to all bird genomes.

528

Frequent Inactivation of *MOXD2* Genes in Birds

Chul Jun Goh[1], Dong-Bin Park[1], Hye In Kim[1], Yoonsoo Hahn[1]

1. Department of Life Science, Chung-Ang University, Seoul, South Korea

Vertebrate MOXD2 encodes a monooxygenase DBH-like 2 protein that could be involved in neurotransmitter metabolism, potentially during olfactory transduction. Loss of MOXD2 in apes and whales has been proposed to be associated with evolution of olfaction in these clades. We analyzed 57 bird genomes to identify MOXD2 sequences and found frequent loss of MOXD2 in 38 birds. Among the 57 birds, 19 species appeared to have an intact MOXD2 that encoded a full-length protein; 32 birds had a gene with open reading frame-disrupting point mutations and/or exon deletions; and the remaining 6 species did not show any MOXD2 sequence, suggesting a whole-gene deletion. Notably, among 10 passerine birds examined, 9 species shared a common genomic deletion that spanned several exons, implying the gene loss occurred in a common ancestor of these birds. However, 2 closely related penguin species, each of which had an inactive MOXD2, did not share any mutation, suggesting an independent loss after their divergence. Distribution of the 38 birds without an intact MOXD2 in the bird phylogenetic tree clearly indicates that MOXD2 loss is widespread and independent in bird lineages. We propose that widespread MOXD2 loss in some bird lineages may be implicated in the evolution of olfactory perception in these birds.

529

Local adaptation of the red-browed finch *Neochmia temporalis* across climatic zones

Kerensa McElroy[1], Leo Joseph[1]

1. CSIRO, Canberra, ACT, Australia

The red-browed finch *Neochmia temporalis* occupies diverse habitats, from monsoonal tropics to temperate forests. To study the genomics of local climate adaptation, we performed whole genome sequencing of 60 *N. temporalis*, to 2x per specimen. We justify our use of WGS by demonstrating that it is a powerful population genomics tool, with the potential to also reveal functional targets of selection.

Entire mitochondrial genomes were reconstructed from WGS data for all specimens. Phylogenetic analysis distinguished the northern subspecies, *N. t. minor*, from its southern counterpart, *N. t. temporalis.* Within *N. t. temporalis*, mitochondrial divergence was low (0.27%), mostly comprising singletons, and exhibited a star-like phylogeny, suggesting a recent rapid expansion in temperate habitats. Mitochondrial divergence between the two subspecies was high (2%). We found no evidence for mitochondrial selection.

To facilitate genome wide scans for selection, reads from each subspecies were pooled separately and population-level SNPs called. A parallel analysis of two other widespread finches, *Stizoptera bichenovii* and *Lonchura castaneothorax,* was also performed. Concordant islands of differentiation between the three species may represent genomic regions underlying local climate adaptation.

We also attempted individual-level SNP analysis, arguing that due to biases in read generation, some genomic regions are likely to have sufficient coverage to call variants in most specimens. Around 500 nuclear SNPs were obtained in this proof-of-principle analysis. Clustering of these SNPs identified*N. t. minor* as per the mitochondrial analysis, verifying the utility of this approach.

Intriguingly, the individual-level SNPs also resolved a second subspecies, *N. t. loftyi*, endemic to South Australia. Discordance with the mitochondrial results suggests a role for selection, rather than neutral processes, in the differentiation of *N. t. loftyi* and *N. t. temporalis*. The divide between *N. t. loftyi*and *N. t. temporalis* also represents the first direct molecular evidence of the Murravian biogeographical barrier.

530

Genomic divergence in allopatry vs. parapatry through the speciation process

Joshua V Penalba[1], Leo Joseph[2], Craig Moritz[1]

1. Australian National University, Acton, ACT, Australia

2. Australian National Wildlife Collection, CSIRO, Canberra, ACT

The geographic mode of speciation (allopatry vs. parapatry) has been shown to influence the accumulation of genomic divergence through the speciation process. The outcomes of divergence in allopatry in contrast to parapatry are likely influenced by differences in the effect of genetic drift and linked selection, which in turn are affected by the history of connection and gene flow. To understand the different ways that genomic divergence can accumulate as speciation progresses, we take a comparative using 10 Australian bird species co-distributed across these 4 geographic regions: Northern Territory, Cape York Peninsula, central Queensland, and Papua New Guinea. The populations are divided by well-known biogeographical barriers and population pairs are either allopatric or parapatric relative to one another. Thousands of SNPs from ddRADseq data were collected to characterize population divergence parameters and gene flow. Preliminary analysis shows wide variation in levels of genomic divergence and relationships between populations within each species. Findings depict the differences in the accumulation of genomic divergence when populations diverge in allopatry vs. parapatry. Understanding the process by which genome-wide divergence accumulates in these scenarios will allow us to further understand potential speciation histories of diverged taxa as well as better predict potential outcomes of diverging populations.

531

Delineating the gene birth and death for Mx homologs confers divergent antiviral activity to Influenza Virus

Yali Hou[1], Furong Qi[1], Xuemei Lu[1]

1. Beijing Institute of Genomics, Chinese Academy of Sciences, Chao Yang, Beijing 100101, China

Genes have perpetually been experiencing birth and death during the process of comprehensively genomic evolution. After a systematical investigation of the evolutionary trajectories of Mx homologs across Chordata, we restored the exquisite history of Mx development. The current findings strongly suggested that Mx genes have been undergoing dramatic expansions and losses in long-term evolution. What should be noteworthy is loss of Mx copy in all observed avian and re-duplication in the majority of mammals. We next confirmed the evolutionary profiling of Mx conferred divergent antiviral activity to Influenza A Virus, where mammals including human and mouse were entitled with strong antiviral activity, whereas, Aves consisting chicken and duck were completely deprived of this capability, implying a lineage-specific selection driven by an ever-changed pathogen environment. Functional differentiation of Mx1 and Mx2 in human and mouse provide specific mechanism for anti- host specific

pathogens. Finally, we partially revealed the putative genetic mechanisms by which chicken Mx barely presented anti-virus capability based on the systematic structure and pathway analysis, which may be partially explained by evolution strategies of key genes responsible for influenza A virus replication. Our findings highlight that the evolutionary dynamics of Mx genes across species in Chordata change the genetic systems and take responsibility of phenotypic evolution/diversity.

532

The evolution of RNA recognizing Toll-Like Receptors in migratory waders

Nynke Raven[1], Simeon Lisovski[1], Marcel Klaassen[1], Bethany Hoye[1], Nathan Lo[2], Thomas Madsen[1], Simon Ho[2], Beata Ujvari[1]

1. Center for Intergrated Ecology, School of Life and Environmental Sciences, Deakin University, Waurn Ponds, VIC, Australia

2. School of Life and Environmental Sciences, The University of Sydney, Sydney, NSW, Australia

Migratory birds have a lifestyle that creates exposure to a wide range of pathogens, making them ideal model systems to study the evolution of immune genes, including toll-like receptors (TLRs). TLRs are part of the innate immune system and recognize conserved patterns of pathogens, including viruses. We investigated the evolution of the ectodomain (ECD) of two TLRs (TLR3 and TLR7) involved in virus recognition in three migratory wader species (*C. alba, C. ruficollis, A.interpres*), as well as across other avian groups. Our results revealed that the inferred relationships among avian TLR3 and TLR7 ECDs do not match the phylogeny of birds. Furthermore, we showed that although both loci are mostly under purifying selection, the evolution of avian TLR3 ECD is also driven by episodic diversifying selection.

We discovered that TLR7 is duplicated in all three wader species, as well as in other Charadriiformes, Cuculiformes and Passerines. The duplication appears to be ancestral for each order, and the duplicated copies appear to be undergoing concerted evolution.

Significantly higher proportions of non-synonymous mutations were detected in TLR7 than in TLR3 across the three wader species. In addition, while the phylogenetic relationships of wader TLR7 matched those of the three species, initial analysis of TLR3 showed potential associations with the species' ecology, such as exposure to pathogens like avian influenza virus.

533

Evolutionary dynamics of old, homomorphic sex chromosomes in Paleognathous birds.

Tim Sackton[1], Luohao Xu[2], Scott Edwards[1]

1. Harvard University, Cambridge, MA, United States

2. Dept of Evolutionary Biology, Uppsala University, Uppsala

In species with stable systems of chromosomal sex determination, sex chromosomes typically evolve recombination suppression, which is followed by degeneration of the heteromorphic chromosome. In most birds, which have a ZW sex determination system, the endpoint of this process is a highly degenerate female-specific W chromosome with a typically very small pseudo-autosomal region (PAR), leading to heterogametic (ZW) females and homogametic (ZZ) males. However, many paleognathous birds (including the flightless ratites and volant tinamous), have old, homomorphic sex chromosomes, with large PARs and largely non-degenerate W chromosomes. The mechanisms by which old, homomorphic sex chromosomes can be maintained remain mysterious. To better understand the evolutionary dynamics of Paleognath sex chromosomes, we have identified Z-linked sequence from 10 newly sequenced Palaeognaths, including 7 flightless ratites and 3 tinamous. We annotated the PAR in a subset of these species, revealing considerable variation in the extent of the PAR and sex chromosome degeneration across this group. Transcriptome sequencing in Chilean tinamou and emu reveals surprising, albeit highly preliminary, evidence for complete dosage compensation in adult emus via down-regulation of male genes in the degenerated region, in contrast to Chilean tinamous and other previously studied birds which lack complete dosage compensation. Consistent with many other taxa, we find evidence for faster evolutionary rates of proteins on the Z chromosome compared to autosomes; unexpectedly, we find that in Palaeognaths genes in both the PAR and the degenerated region show elevated rates of protein evolution compared to autosomes. This pattern is consistent with reduced recombination rates in the PAR (reducing the efficacy of selection), and suggests that palaeognath sex chromosomes may experience reduced recombination rates without subsequent degeneration. Taken together, our work suggests that the evolution of old, homomorphic sex chromosomes in palaeognaths may be an exception to standard models of sex chromosome evolution.

534

A genomic approach to examining two avian hybrid zones

Ashlee Shipham[1], Daniel J Schmidt[1], Leo Joseph[2], Jane M Hughes[1]

1. Australian Rivers Institute, Griffith School of Environment, Griffith University, Nathan, Queensland, Australia

2. Australian National Wildlife Collection, CSIRO National Facilities and Collections, Canberra, ACT, Australia

Hybrid zones provide a unique opportunity to examine the strength and nature of species boundaries. The occurrence of hybridization can have wide-ranging implications for the taxa involved, depending on the strength of pre- and post-zygotic barriers to gene flow between them. Rosellas are a group of colourful and charismatic Australian parrots, with a number of morphologically distinct species and subspecies exhibiting zones of intergradation where their ranges overlap. Two zones that have been examined morphologically, but which lack an analysis of their underlying genetic composition are studied here. The first is between two non-sister species, the pale-headed (*Platycercus adscitus*) and eastern (*P. eximius*) rosellas, which hybridize in the border region between Queensland and New South Wales. The second is a broad zone between the two currently recognized *P. adscitus* sub-species, located in Queensland. Here, we have taken a genomic approach to advance the understanding of these two zones. We aim to distinguish whether they represent hybrid zones or geographic clines in morphology, and examine the presence and pattern of gene flow across them. We employ restriction site-associated DNA (RAD) sequencing to generate thousands of loci. Using a combination of assignment tests and genomic cline analyses, we find different patterns of variation in the *P. adscitus /P. eximius* hybrid zone and that between *P. adscitus* subspecies. The former appears to be largely bimodal, with some support for introgression of *P. eximius* alleles into *P. adscitus*. The latter appears less structured, and we consider (1) that this may represent a geographic cline in plumage rather than a hybrid zone, or (2) that weak/absent barriers to reproductive isolation may have resulted in genomic introgression extending beyond that which is readily identifiable based on plumage characteristics. The potential causes and implications of our findings are discussed.

535

Sand dunes drive asymmetric gene flow between arid-zone Thick-billed grasswren subspecies

Amy L Slender[1], Marina Louter[1], Tessa Bradford[1,2], Michael G Gardner[1,2], Sonia Kleindorfer[1]

1. Flinders University, Adelaide, SA, Australia

2. South Australian Regional Facility for Molecular Evolution and Ecology, Adelaide, SA, Australia

Gene flow is a useful tool for understanding evolutionary processes such as genetic drift and natural selection. This is because gene flow is limited across geographic barriers that cause genetic drift or by genetic incompatibilities that develop due to natural selection. The Thick-billed grasswren (*Amytornis modestus*) is one species which show evidence of limited gene flow. Two arid zone subspecies of this threatened species have a parapatric distribution but show 1.7% mitochondiral divergence on either side of a 50km wide divide that runs between the South Australian salt lakes, Lake Eyre and Lake Torrens. Thick-billed grasswrens do not occur on sand dunes yet there are many sand dunes within this divide. Therefore this study investigates whether sand dunes are a barrier to gene flow between the two subspecies, *A. m. indulkanna*, which occurs to the west of the divide and *A. m. raglessi,* which occurs to the east. We identified approximately 3000 SNPs using genotype by sequencing (GBS) methodologies and then measured gene flow between 65 individuals to the east of the sand dunes and 37 individuals to the west. Preliminary results showed that the genotypes of individuals within the divide but to the east of the sand dunes were predominantly *A. m. raglessi* but were introgressed with genotypes of *A. m. indulkanna* indicating that limited gene flow is occurring across the sand dunes between Thick-billed grasswren subspecies. Introgressed individuals were only found to the east of the sand dunes suggesting that gene flow is asymmetric across the sand dunes from the west (*A. m. indulkanna*) into the east (*A. m. raglessi*). In the past, sand dunes were likely to be more abundant and may have been a substantial geographic barrier to gene flow that meant isolated populations were effected by genetic drift.

536

Molecular evolution of dietary diversification in birds

Kai Wang[1], Yi Wang[1], Huabin Zhao[1]

1. College of Life Sciences, Wuhan University, Wuhan, HUBEI, China

Dietary diversification is affected by both sensory perception and nutrient metabolism. We are studying taste receptor genes and digestive enzyme genes to recover the evolutionary history of dietary diversification in birds, which are rich in species and diets. First, we characterized the evolution of avian bitter taste receptor genes (Tas2rs) based on 48 genomes representing all but 3 avian orders. We found that birds generally carry a small repertoire of Tas2rs, and uncovered a positive correlation between the number of putatively functional Tas2rs and the abundance of potential toxins in avian diets. Second, we examined a digestive enzyme gene CHIA, which encodes an enzyme degrading chitin, across the 48 birds as mentioned above. We found that this gene has undergone multiple independent duplication events, driven by positive selection, in most insect-eating birds. Because chitin is abundant in insects, we are testing whether this gene has experienced adaptive functional diversification in insect-eating birds, by measuring enzyme activities of the recombinant enzyme in E.coli. Taken together, we attempt to uncover the genetic evidence in response to dietary changes in birds by examining the evolution of sensory genes and digestive-related genes. Our future work will continue to focus on molecular evolution and ecology in vertebrates, particularly in birds that are iconic in China or the Tibetan Plateau. (Contact Dr. Huabin Zhao by email: huabinzhao@whu.edu.cn)

537

Averaging over alternative multiple sequence alignments increases the accuracy of phylogenetic tree reconstruction

Haim Ashkenazy[1], Itamar Sela[1], Giddy Landan[2], Tal Pupko[1]

1. The Department of Cell Research and Immunology; George S. Wise Faculty of Life Sciences, Tel Aviv University, Tel Aviv, Israel

2. Genomic Microbiology Group, Institute of Microbiology, , Christian-Albrechts-University of Kiel, Kiel, Germany

The classic methodology of inferring a phylogenetic tree from sequence data is composed of two steps. First, a multiple sequence alignment (MSA) is computed. Then, a tree is reconstructed assuming the MSA is correct. However, inferred MSAs have been shown to be inaccurate and errors in them reduce tree inference accuracy. It was previously proposed that filtering unreliable alignment regions can increase tree accuracy inference. However, it was also demonstrated that the benefit of this filtering is often obscured by the resulting loss of phylogenetic signal.

In this work we explored an ad-hoc approach, in which instead of relying on a single MSA, we generate a large set of alternative MSAs using the GUIDANCE2 methodology and concatenate them into a single super-MSA. By doing so, we account for phylogenetic signals contained in columns that are not present in the single MSA produced by alignment algorithms. Using simulations, we demonstrate that this approach results in more accurate trees compared to (1) using an un-filtered alignment; (2) using a single alignment with weights assigned to columns according to their reliability. Next, we explore in which regions of the MSA space our approach is expected to be beneficial. Finally, we provide a simple criterion for deciding whether or not the extra effort of computing a super-MSA and inferring a tree from it is beneficial. We expect our methodology to be useful for many cases in which relatively diverged sequences are analyzed and applying the more computationally intensive statistical alignment approach is not feasible.

538

Alignment-free networks: One step further into the next generation phylogenomics

Guillaume Bernard[1], Paul Greenfield[2], Mark A Ragan[1], Cheong Xin Chan[1]

1. Institute for Molecular Bioscience, The University of Queensland, Brisbane, QLD 4072, Australia

2. CSIRO, Oceans and Atmosphere, North Ryde, NSW 1670, Australia

Genome evolution in microbes involves highly dynamic molecular mechanisms such as genome rearrangement and lateral genetic transfer (LGT). These mechanisms violate the implicit assumption of full-length contiguity in multiple sequence alignment (MSA), commonly used in phylogenetic analysis. Furthermore, MSA-based approaches necessitate heuristic methods e.g. Bayesian inference in reconstructing phylogenies, which are not scalable to the quantity of existing and forthcoming genome data. An alternative strategy is to infer evolutionary relatedness based on shared subsequences of fixed length, known as *k*-mers, i.e. *alignment*-free (AF) methods. Here using 143 complete genomes, we reconstruct a phylogenomic network using an AF approach (based on 25-mers). This AF network showcases the extent of shared genomic fragments among diverse phyla, e.g. the high extent of genetic exchange among Proteobacteria versus the low extent between Chlamydophilia and Cyanobacteria. Our observations largely agree with published studies, but also highlight the inadequacy of representing microbial phylogeny as a tree. AF

approaches provide exact solutions (i.e. pairwise distance between genomes based on shared *k*-mers) which can be directly used in deriving a phylogenomic network. Using different distance thresholds, we can easily capture changes in the network structure, e.g. cliques, that can reflect evolutionary dynamics of microbial genomes. Functional relevance of the different evolutionary scenarios, e.g. *k*-mers implicated in the formation of a clique of interest, can be inferred based on correlation of *k*-mer positions to gene annotation. To investigate the impact of plasmids and highly conserved genes in phylogenomic inference, using 2,774 complete bacterial genomes we reconstructed AF phylogenomic networks using (a) all genome data including plasmids, (b) plasmid-only sequences, (c) chromosomal sequences without ribosomal RNAs, and (d) only ribosomal RNAs. Here I present our recent findings from these analyses, and demonstrate how AF approaches can be applied to understand the evolutionary dynamics of microbial genomes using large-scale data.

539

Estimating Mutation Parameters and Population History Simultaneously from Temporally-Spaced Genome Data

Arman Bilge[1], Tanja Stadler[2], Alexei Drummond[1]

1. Centre for Computational Evolution, The University of Auckland, Auckland, New Zealand

2. Department of Biosystems Science and Engineering, ETH Zürich, Basel, Switzerland

Next generation methods have made it increasingly feasible to sequence entire genomes. Furthermore, the possibility to recover ancient genomes also augments the amount of available data in another dimension. The analysis of temporally- spaced sequences dramatically increases our statistical power to make inference about evolutionary dynamics, as frequently demonstrated in epidemiology. However, current methods that handle sampling-through-time do not scale well to the large genomes of diploid organisms due to the computational intractability of searching tree space. Instead, we use a closed-form expression that analytically integrates over all coalescent phylogenies at a constant or biallelic site. By assuming the genealogical independence of sites, we can express the likelihood in a form that is computationally tractable. Additional speedups are attained by pre-processing the genome alignment and an efficient parallel implementation. This approach utilises similar mathematical techniques to the SNAPP software for inferring species trees from genetic markers, but is extended to account for serially-sampled data. Notably, our method considers both biological processes, such as the transition–transversion ratio and site rate heterogeneity, as well as practical problems, including unphased genomes and sequencing error. After placing appropriate priors, we use Markov chain Monte Carlo to sample from the posterior distribution on our parameters. We demonstrate the usefulness of our method for inference on both simulated and real datasets.

1. David Bryant, Remco Bouckaert, Joseph Felsenstein, Noah A. Rosenberg, and Arindam RoyChoudhury. Inferring Species Trees Directly from Biallelic Genetic Markers: Bypassing Gene Trees in a Full Coalescent Analysis. Mol Biol Evol (2012) 29 (8): 1917-1932. doi:10.1093/molbev/mss086

540

Estimating the Effective Population Size from Experimental Evolution Data

Agnes Jonas[1], Thomas Taus[1], Carolin Kosiol[1], Christian Schlötterer[1], Andreas Futschik[2]

1. Population Genetics, University of Veterinary Medicine, Vienna, Austria

2. JKU Linz, Linz, OÖ, Austria

The effective population size (N_e) is an important parameter in population genetics and evolutionary biology telling us about the expected amount of variability due to genetic drift. Although several methods have been developed, this parameter is known to be often difficult to estimate. In particular, appropriate estimators have not been available for experimental evolution experiments producing pool sequencing data. Therefore we propose a new estimator that relies on allele frequency changes in temporal data. Our approach corrects both for the variance due to sampling individuals and the random sampling of reads out of the DNA pool during sequencing. In our simulations, we obtained accurate estimates, as long as the drift variance is not too small compared to the sampling and sequencing variance. We also extend our method using a recursive partitioning approach to estimate $N_e$locally along a chromosome. Since type I error control is available, our method permits the identification of genomic regions that differ significantly in their effective population size. We present an application to Pool-Seq data from experimental evolution with *Drosophila melanogaster* and give recommendations for using our approach on whole-genome data. The estimator is computationally fast and available as an R-package at https://github.com/ThomasTaus/Nest .

541

Evolution and diversity of complement genes in crocodilians

Christine Keo[1], Victoria M Lee[1], Yi Wei[1], Qais ALRawahi[1], Sebastian Duchene[2], Amanda Chong[1], Sally R Isberg[3], Jaime Gongora[1]

1. Faculty of Veterinary Science, School of Life and Environmental Sciences, The University of Sydney, Sydney, NSW 2006, Australia

2. Faculty of Science, School of Life and Environmental Sciences, The University of Sydney, Sydney, NSW 2006, Australia

3. Center for Crocodile Research, Noonamah, NT, Australia

The complement system plays an important role in the innate immune response in higher vertebrates. It is involved in cell lysis and initiation of phagocytosis by opsonisation of pathogens and induction engages the cells of the immune system to trigger processes leading to inflammation. The complement system consists of about thirty proteins grouped in five major gene families. These encode for distinct plasma proteins that react with each other forming three activation cascades (alternative, lectin and classical) which converge in a single terminal pathway. This system appears to be highly conserved in vertebrates. However there is a gap in the knowledge in regards to evolution and diversity of the complement system in crocodilians. To address this, we investigated a large number of complement system genes in the recent available genomes of three crocodilian species (*Alligator mississippiensis, Crocodylus porosus, and Gavialis gangeticus*) and identified 28 genes. These genes were then compared to other vertebrates including reptiles, birds and mammals to further understand the extent of conservation and differentiation of the complement system among the three extant crocodilian families. We surveyed 25 exons, representing seven genes, across 20 species of Alligatoridae and Crocodylidae. Phylogenetic analyses showed that crocodilian complement genes form orthologous clades across species suggesting that they have evolved independently from each other after speciation. As expected, a considerable level of sequence conservation among species was observed. However, we also found a relatively high frequency of DNA substitution in particular among species of Alligatoridae. No correlation between genetic diversity or allele distribution and species' habitat or geographic distribution was identified. These findings advance

our knowledge of the evolution of this innate immune system and the immunogenetic resources generated here are now being used for population genetic and disease association studies.

542

A transcriptome annotation pipeline for non-model organisms

Fabrizio Ghiselli, Mariangela Iannello, Emanuele Procopio, Marco Passamonti

The introduction of high-throughput sequencing technologies allowed researchers to generate large amounts of genomic data at limited cost and time. This opportunity had a groundbreaking impact on the study of non-model organisms: above all, RNA-Seq and *de novo* transcriptome assembly represent a valuable source of information in species for which genomic resources are scarce or absent. However, sequencing and assembly are only the first steps, and an accurate annotation is fundamental for every kind of biological analysis. Annotation of transcriptomes from model organisms and their closely-related species is quite straightforward, and is generally based on simple sequence similarity searches. Conversely, non-model organisms require more complex and integrated procedures in order to infer remote homology and function. We present a pipeline specifically thought for the annotation of transcriptomes of non-model organisms. It consists of an integrated approach that combines different bioinformatics tools to obtain: 1) filtration from contaminant sequences; 2) ORF prediction, identification of pseudogenes and artificially fused transcripts; 3) coding sequence annotation based both on sequence similarity and on the identification of conserved domains by protein signature recognition; 4) functional annotation of coding sequences by the assignment of GO terms; 5) identification of orthologs; 6) annotation of noncoding transcripts. We tested our pipeline by re-annotating the transcriptome of *Ruditapes philippinarum* (Bivalvia, Veneridae).

543

VAST-DB: an Atlas of Alternative Splicing Profiles in Vertebrate Tissues and Cell Types

Javier Tapial[1], Toni Hermoso-Pulido[1], Timothy Sterne-Weiler[2], Kevin Ha[2], André Gohr[1], Benjamin J Blencowe[2], Manuel Irimia[1]

1. Centre for Genomic Regulation (CRG), Barcelona, Spain

2. The Donnelly Centre, University of Toronto, Toronto, ON, Canada

Alternative splicing (AS) is a widespread process facilitating the generation of regulatory and proteomic complexity. However, the function of the vast majority of AS events detected to date is unknown, and landscapes of regulated AS remain to be identified. To contribute to address these challenges, we present VAST-DB, a massive resource of genome-wide quantitative profiles generated using vast-tools (https://github.com/vastgroup/vast-tools) for all main classes of AS events of a wide range of human, mouse and chicken tissues, cell types and developmental stages. The resulting atlas of AS events reveals extensive new intergenic and intragenic regulatory and functional relationships involving different classes of AS events, as well as previously unknown conserved landscapes of tissue-regulated exons. We also report and validate hundreds of AS events that are alternatively spliced in virtually all profiled tissue and cell types. These AS events are highly enriched in genes that encode transcription factors and DNA binding proteins, and single cell RNA-seq data show that they are usually predicted to generate protein isoforms that co-exist in the same cell. Finally, we also provide mapping of these AS events to protein regions and experimentally determined or modeled protein structures. Our AS atlas thus provides a valuable basis for new explorations of splice isoform regulation and function in an evolutionary context.

544

ModelFinder: A new model-selection method that greatly improves the accuracy of molecular phylogenetic estimates

Subha Kalyaanamoorthy[1], Bui Q Minh[2], Thomas KF Wong[3], Vivek Jayaswal[4], Arndt von Haeseler[2], Lars S Jermiin[1]

1. CSIRO, Canberra, ACT, Australia

2. CIBIV/MFPL, Vienna, Austria

3. Research School of Biology, Australian National University, Canberra, ACT, Australia

4. School of Biomedical Sciences, Queensland University of Technology, Brisbane, Queensland, Australia

Model-based molecular phylogenetic analysis plays a key role in most comparisons of genomic data, and model selection is a critical step in all such analyses. We introduce ModelFinder, a novel model-selection method that: (a) incorporates a model of rate-heterogeneity across sites (RHAS) not previously considered in this context, and (b) allows the topology of the tree to be a variable during the search for an optimal model of evolution. Using alignments of amino acids obtained by simulation on a 100-tipped tree, ModelFinder is found to produce accurate estimates of RHAS, where other model-selection methods fail (e.g., when the model of evolution incorporates a bimodal distribution of RHAS). The benefits of ModelFinder are demonstrated using genomic data from a Tree-of-Life study involving Bacteria and Archaea. The results from our analysis of these data reveal that: (a) the optimal model of RHAS is trimodal, (b) the popular Γ-distribution-based model of RHAS provides a poor fit of the data, and (c) the optimal trees inferred under these two models of RHAS are very different. ModelFinder is implemented in the phylogenetic program IQ-TREE.

545

MST -> P: A scalable and accurate method for reconstructing phylogenetic trees

Prabhav Kalaghatgi[1]

1. Max-Planck-Institut für Informatik, Saarbrücken, SAARLAND, Germany

Popular clustering-based algorithms such as neighbor-joining have a worst case time-complexity that is $\Omega(n^3)$. These approaches do not scale well with the large amount of data that is being made available with current sequencing technologies. Additionally, clustering-based methods are not as accurate as methods based on tree search that optimize criteria such as likelihood. Recent advances in this field have been in the direction of dividing the data into smaller overlapping sets, constructing subtrees on each small set, and then combining the subtrees in order to recover the full tree. Aside from having lower time complexity, these approaches reduce reconstruction error, when compared with reconstruction based on the entire dataset.

One notable such supertree method is CLGrouping, in which the relevant sequence sets are defined as the neighborhood of each internal vertex of a minimum spanning tree (MST) that is constructed using all pairwise distances.

Our contributions are as follows. We extend CLGrouping to the case where there are multiple MSTs. We developed a similar MST-based method that builds subtrees using family-joining, a method for constructing *generally labeled trees*. Generally labeled trees are unrooted phylogenetic trees which may have taxa placed at internal vertices, and which may contain polytomies. We demonstrate on simulated trees with 1000 taxa that CLGrouping's method of partitioning sequences yields very small sequence sets, resulting in a significantly higher reconstruction error, when compared with trees built using relatively larger sequence sets.

There seems to be a trade-off between the error of reconstructing subtrees and the error of partitioning sequences using edges in the MST. We are currently developing a method that selects the edges of the MST, for partitioning the sequences in such a way that reconstruction error is minimized. This approach allows accurate reconstruction while maintaining the high degree of parallelism that is inherent to the divide-and-conquer approach.

546

A reversible Polymorphism-aware phylogenetic Model (revPoMo) for species tree estimation

Dominik Schrempf[1, 2], Nicola De Maio[3], Bui Q Minh[4], Arndt von Haeseler[4], Carolin Kosiol[1]

1. Institute of Population Genetics, Vetmeduni Vienna, Vienna, Austria

2. Vienna Graduate School of Population Genetics, Vienna, Austria

3. Nuffield Department of Medicine, Oxford University, Oxford, UK

4. Center for Integrative Bioinformatics Vienna, Max. F. Perutz Laboratories, University of Vienna, Vienna, Austria

The availability of genome-scale inter- and intraspecies data leads to new opportunities in phylogenetics to improve tree accuracy and resolution as well as to take important steps towards understanding the process of speciation.

We present a reversible Polymorphism-Aware Phylogenetic Model (revPoMo) for species tree estimation from genome-wide data. revPoMo enables the reconstruction of large scale species trees for many within-species samples. PoMo expands the alphabet of DNA substitution models to include polymorphic states (De Maio et al., MBE 2013). It is a selection-mutation model which separates the mutation process from the fixation process. Thereby, a Moran process is used to model genetic drift. Although a single phylogeny — the species tree — is considered, PoMo naturally accounts for incomplete lineage sorting because ancestral populations can be in a polymorphic state. A large scale simulation study as well as applications to great ape data (12 populations in total, Prado-Martinez et al., 2013) show that PoMo is fast while being more accurate than coalescent-based methods (De Maio, Schrempf, and Kosiol, Syst. Biol. 2015).

We recently implemented revPoMo in the maximum likelihood software IQ-TREE (Nguyen et al., 2015). The runtimes of our approach and standard substitution models are now comparable on large trees (e.g. 60 species with 10 individuals each) but revPoMo has much better accuracy in estimating trees, divergence times and mutation rates under the scenarios of recent radiation and incomplete lineage sorting. The advantage of revPoMo is that an increase of sample size per species improves estimations but does not increase runtime. Therefore, revPoMo is a valuable tool with several applications, from speciation dating to species tree reconstruction. We also present preliminary results on applying our method to genome-wide data from baboons showing interesting insights into their complex population history and new estimates of divergence times and mutation rates.

547

Phylogeny-guided genome assembly method for short nucleotide reads from deep sequencing of mixed microbial samples

Tommy Tsan-Yuk Lam[1], Guangchuang Yu[1]

1. The University of Hong Kong, Pokfulam, Hong Kong

Background: Co-circulation of different microbial strains in the same host population/region is common (e.g. influenza virus). Although high-throughput sequencing (HTS) is very efficient, its relatively short sequencing reads require assembly to obtain complete/longer biological sequences for downstream analysis, which is a challenge in samples with co-infections/co-existence of multiple genetically similar microbial strains. Typical *de novo* assembly methods, which largely rely on overlapping regions between reads, have a high risk of mis-assembling the short reads from similar strains into recombinant sequences. Conventional reference-based assembly methods rely on pre-selection of correct genome sequences as reference templates, which is often difficult.
Methods: To address this problem, we have proposed and implemented an algorithm to efficiently and accurately assemble short reads into genomes of different strains with the aid of phylogenies built from database sequences. This method is template-selection free and is expected to be less erroneous than *de novo* assembly that relies on overlapping regions. Here, we demonstrated the utility of this novel method for influenza A virus samples. Mock co-infections were generated by mixing two or more sets of HTS reads simulated from different influenza virus strains, which were then subject to assembly by our phylogeny-based method, conventional *de novo* (Velvet) and reference-based (Bowtie2, BWA) methods, for performance comparison.
Results: The coverage and accuracy of the genomes assembled by our algorithm were as high as using reference-based assembly when the correct number and strains of reference genomes were known. Our method outperformed *de novo* methods and reference-based methods when incorrect strains were used as templates. Our algorithm also simultaneously determined phylogenetic positions of the assembled genomes in the global phylogeny. These results show that our phylogeny-based method is a useful alternative to other existing assembly methods.

548

Evolution of the Major Histocompatibility Complex (MHC) class I in wild pigs and peccaries

Carol Lee[1], Marco Moroldo[2], Alvaro Perdomo[1], Núria Mach[2], Sylvain Marthey[2], Jérôme Lecardonnel[2], Per Wahlberg[2], Amanda Chong[3], Simon Ho[4], Jordi Estellé[2], Claire Rogel-Gaillard[2], Jaime Gongora[1]

1. Faculty of Veterinary Science, School of Life and Environmental Sciences, University of Sydney, Sydney, NSW 2006, Australia

2. IGABI, INRA, AgroParisTech, Université Paris-Saclay, 78350 Jouy-en-Josas, France

3. Vertebrate and Health Genomics, The Genome Analysis Centre, Norwich, UK

4. Faculty of Science, School of Life and Environmental Sciences, University of Sydney, Sydney, NSW 2006, Australia

The Major Histocompatibility Complex (MHC) is an important component of developing and regulating host immune response. It is a key genomic model region for understanding the evolution of gene families and co-evolution between host and pathogen. However, from Suidae (pigs) and its' closest living taxa, Tayassuidae(peccaries), which comprise ~30 extant species (mostly wild), only the genome of the domestic pig (*Sus scrofa*) has been extensively sequenced and annotated. Based on the *S. scrofa* MHC, it has been suggested that the non-classical class I gene series (MHC Ib) emerged after Suidae diverged from the rest of the artiodactyls. To test this hypothesis, we used DNA sequence capture to generate MHC resources from 12 wild species of Suidae and Tayassuidae. This was done using the *S. scrofa* MHC as a reference sequence for the design of the capture array and next-generation sequencing. Our results show that: i) the repertoire of MHC non-classical genes MHCIb (and classical MHC Ia) is present in both Suidae and Tayassuidae in contrast with the previous hypothesis; ii) these genes underwent a series of duplications and differentiation before these taxa diverged from the common ancestor more than 35 million years ago; iii) all genes have evolved independently from each other after speciation; and iv), there are genetic patterns of differentiation for most of the MHC Ia and Ib genes between Eurasian and sub-Saharan wild Suidae. These findings improve our understanding of the evolutionary history of these adaptive immune genes and provide genetic resources to investigate the immune response of wild populations to diseases including local adaptation of some taxa to emerging diseases.

549

Haplotype reconstruction from Short Read Sequences using Vector Quantization

Louis Ranjard[1], Allen Rodrigo[1]

1. The Australian National University, Canberra, ACT, Australia

Uncovering the genetic diversity in an unlabelled mixed sample of individuals is a challenging computational problem. However, this task can be of primary importance. For example, reconstructing the true set of viral haplotypes in an infected host is correlated to the clinical outcome and pathogenesis. We present a vector quantization approach to reconstruct the set of haplotype from a mixed sample of short read sequences when the number of haplotypes is unknown and the sequencing reads are unlabelled. Our method consists in mapping the high dimensional short read sequence space to a lower dimensional space representing the reconstructed haplotypes. We propose to encode each position in the nucleotide sequences as a vector where each element represent the contribution of each nucleotide base. Then, an artificial neural network is used to reconstruct the haplotype sequences using competitive learning. The network is tree shaped and can therefore be considered as a short read classification tree. During the learning process, bifurcating tree branches are added in the tree according to a dispersion criterion. Preliminary results show that (i) the true haplotype sequences can be reconstructed and (ii) the true number of haplotypes can be inferred from the size and the structure of the classification tree. In particular, the distance between the short read sequences to the tree node weight matrices is informative of the true set of haplotypes in the sample under study.

550

Comprehensive Annotation of Multigenic Protein-Family Structures (CAMPS)

David F Clarke[1, 2], Charles Robin[1], Lars S Jermiin[2]

1. University of Melbourne, Parkville, Vic, Australia

2. Land and Water, CSIRO , Canberra, ACT, Australia

The automated methods currently used for genomic annotation of gene structure often fall short of the desired degree of accuracy appropriate for analysis of multigenic protein-families within and between species. These methods are restrained by a poor ability to distinguish the relationships of stretches of coding sequence to each other and an over reliance on a single inevitably flawed genome assembly. To increase the accuracy of gene structural annotation the CAMPS annotation methods focus on a single protein-family at a time. In this context 'Comprehensive' refers to incorporating multiple lines of evidence from the target species, including multiple genome assemblies and different sources of transcriptome data.

Genomic loci for the protein family in question are initially identified through sequence similarity with previously identified proteins from the target species or other species. Additional evidence is taken from the transcriptomes of the target species and other species. CAMPS clusters the identified loci and transcripts into 'campsites' through sequence similarity and genomic position. A campsite is divided into 'tents', which include variants of suspected genes. Annotation of the suspected gene is performed using all data within the tent. Sequence variants within the tent are identified and then classified.

These methods are designed to; reduce the concatenation and splitting of genes; identify partial and duplicate gene sequences present within the dataset; and compare data across the available genomic assemblies and transcriptomes to identify where assembly errors have caused problems. Accuracy estimates are generated for individual genes from the evidence used to derive the gene models. CAMPS is also designed to present multiple sequence variants of genes when they are found in the dataset.

551

Time-sampled population genetics: an experimental investigation of drug-induced mutational meltdown as an antiviral treatment strategy

Hyunjin Shim[1, 2], Claudia Bank[1, 2], Nicholas Renzette[3], Sebastian Matuszewski[1, 2], Ping Liu[3], Jennifer P. Wang[3], Jeffrey D. Jensen[1, 2]

1. Swiss Institute of Bioinformatics, Lausanne, Vaud, Switzerland

2. École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Vaud, Switzerland

3. University of Massachusetts Medical School, Worcester, Massachusetts, USA

High mutation rates in non-recombining populations invoke several processes that may lead to extinction. R.A. Fisher argued that an intermediate mutation rate would be optimal, ensuring a steady input of beneficial mutations while avoiding the detrimental accumulation of deleterious mutations. Here, we utilized a time-serial experimental evolution framework, combined with novel population genetic inference methods, to investigate the process of mutational meltdown and the role of error catastrophe in influenza A virus (IAV) populations with artificially increased mutation rates. A novel antiviral drug favipiravir, which increases the genome-wide mutation rate in IAV, was administered with varying dosage strategies. Under low concentration regimes, we report the first evidence to date for the ability of virus populations to adapt to favipiravir. However, under high concentration regimes, we observe extinction in all replicates. We discuss the observed evolutionary dynamics with respect to previous theoretical results pertaining to mutational meltdown and evolutionary rescue.

552

Enrichme – a gene set enrichment tool that naturally corrects for gene length and clustering

Hannes Svardal[1, 2], **Ümit Seren**[2], **Magnus Nordborg**[2]

1. *Wellcome Trust Sanger Institute, Cambridge, United Kingdom*
2. *Gregor Mendel Insitute, Austrian Academy of Sciences, Vienna, Austria*

Gene set enrichment analysis is a vital tool to test the biological relevance of and interpret genomic data such as selection scans and genome-wide association studies. However, many of the current methods are biased by the fact that (i) longer genes are more likely to overlap high scores and (ii) genes with similar function are often clustered in the genome. The latter problem is often corrected for by removing variants in high linkage disequilibrium, but the cutoff for this is arbitrary. Here, we present a method that is not affected by either of these biases. Briefly, it determines significance by comparing scores attributed to genes and gene categories to an empirical null-distribution of scores obtained by randomly rotating the data against its genomic positions, hence, keeping LD-structure and gene-clustering intact. Our method is implemented in the free and open source tool enrichme. The implementation features a classical candidate enrichment mode, where the user can define a threshold on genomic input scores, to test whether genes in the vicinity of scores above the threshold are enriched in particular gene sets, such as gene ontology (GO) terms. Alternatively, rather than defining a hard cutoff for genes to test, the program can evaluate the significance of user defined summaries of scores across genes and across gene sets. As an example, one could test whether the mean across a gene set of average scores across each gene in the set are significantly elevated. This approach increases statistical power in cases where the relative magnitude of scores confers biological signal.

1. https://github.com/feilchenfeldt/enrichme

553

Measuring Natural Selection on Chromatin Accessibility in 1000 Humans From 10 Populations.

Ashley K. Tehranchi[1], **Hunter B. Fraser**[1]

1. *Department of Biology, Stanford University, Stanford, CA*

Cis-regulatory elements such as transcription factor (TF) binding sites or open chromatin regions can be identified genome-wide, but it remains far more challenging to pinpoint genetic variants affecting these elements. We recently developed a pooling-based approach to mapping quantitative trait loci (QTLs) for molecular-level traits (Tehranchi et al., Cell, 2016). We applied this to chromatin immuno-precipitation followed by high-throughput sequencing (ChIP-seq) in five TFs and a histone modification and mapped thousands of cis-acting QTLs, with over 25-fold lower cost compared to standard QTL mapping. Thousands of these QTLs have been implicated in genome-wide association studies, providing candidate molecular mechanisms for many disease risk loci, and suggesting that TF binding variation may underlie a large fraction of human phenotypic variation. We are now applying our pooling method to ATAC-seq (a method to assay accessible chromatin) in cells from 1,000 human individuals, from 10 diverse populations, to identify chromatin-accessible QTLs (caQTLs). Our goal is to better understand how natural selection has shaped the landscape of cis-regulation in humans, and how regulatory differences may give rise to population-specific traits. Although many studies have used the genomes of diverse human populations to investigate patterns of natural selection, this approach has significant limitations—for instance, the statistical burden of testing the entire genome has severely limited the power of most selection scans. With our data, we will be able to focus the search on likely causal variants affecting chromatin accessibility. In addition, we will explore an entirely new type of population-genetic analysis not possible with DNA sequence data, comparing chromatin accessibility within populations to variation between populations. Loci with large differences between populations but little variation within populations will be top candidates for targets of positive selection.

554

Estimating Identical-By-Descent tracts from low coverage NGS data

Filipe G. Vieira[1], **Anders Albrechtsen**[2], **Rasmus Nielsen**[3]

1. *Natural History Museum of Denmark, Copenhagen, NA, Denmark*
2. *Department Biology, University of Copenhagen, Copenhagen, Denmark*
3. *Department Integrative Biology, University California Berkeley, Berkeley, USA*

Genome-wide patterns of Identical-By-Descent (IBD) tracts and their variation across individuals provide a valuable insight into human genetic diversity and evolutionary history. Methods have been developed to infer these tracts but they are based on marker/genotype data, due to the low error rates. Next Generation Sequencing (NGS) technologies have revolutionized research in evolutionary biology by both increasing speed and reducing costs. However, these data typically have high error rates due to multiple factors (from random sampling of homologous alleles, to sequencing or alignment errors) and, furthermore, many studies rely on low coverage sequence data (< 5× per site per individual), causing SNP and genotype calling to be associated with considerable statistical uncertainty.

Recent methods rely on probabilistic frameworks to account for these errors, integrating the base quality score together with other error sources to calculate an overall ”genotype likelihood”. This likelihood function can then be combined with a prior to calculate a posterior probability for the genotype. Here, we present a new Hidden-Markov-Model based method to estimate IBD tracts, specially suited to low coverage NGS data since it takes the uncertainty in the data into account by working with genotype likelihoods. Furthermore, and apart from the IBD tracts, this new method also estimates genome-wide inbreeding coefficients that can be used as priors in other analyses. We assess its performance both on simulated data and a subset of the 1000 genomes, looking into several combinations of sample size and coverage, and show accurate inferences for sequencing coverages as low as 2x.

555

MetaMIS: a metagenomic microbial interaction simulator based on a microbial community profiles

Daryi Wang[1]

1. *Biodiversity Research Center, Academia Sinica, Taipei, Taiwan*

The complex and dynamics of microbial community are major factors in ecology system. With NGS technique, metagenomics data provides a new source to explore microbial interactions. Lotka-Volterra models have been widely used to infer animal interaction of dynamic systems and recently been applied to analyze metagenomic data. In this paper, we presented the first Lotka-Volterra model based tool, **Meta**genomic **M**icrobial **I**nteraction **S**imulator (MetaMIS), to analyze time series data of microbial community profiles. MetaMIS firstly infers the underlying microbial interactions from operational taxonomic units (OTU) abundance tables and interprets interaction models by the use of Lotka-

Volterra model. We also embedded Bray-Curtis dissimilarity method in MetaMIS to evaluate the resemblance of biological reality. MetaMIS was designed to tolerate a high level of missing data, it can estimate the interaction information without the influence from rare microbes. For each interaction model, MetaMIS systematically examines interaction patterns (such as mutualism (+/+), competition (-/-), parasitism or predation (+/-), commensalism (+/0), amensalism (-/0), and no effect (0/0)) and refines the biotic role inside microbes. The output of MetaMIS can be exported as Gephi or Cytoscape format for advanced analysis. In a test case, we collected the human female gut microbiota which contained 124 time points of 88 OTUs at the family level. Through the test, MetaMIS generated 55 interactions in around 5 minutes on a standard desktop computer, the results also revealed that rare species may play important roles in the microbial dynamic system. MetaMIS provides an efficiency and user-friendly platform and may reveal new insights from metagenomics data.

556

HAL-HAS 2: A new algorithm that estimates evolution process for heterogeneity across lineages as well as convergent evolution

Thomas K F Wong[1], Vivek Jayaswal[2], Subha Kalyaanamoorthy[3], John Robinson[4, 5], Leon Poladian[4, 5], Lars S Jermiin[3]

1. *Research School of Biology, Australian National University, Canberra, Australia*
2. *School of Biomedical Sciences, Queensland University of Technology, Brisbane, Australia*
3. *Land & Water, CSIRO, Canberra, Australia*
4. *School of Mathematics and Statistics, University of Sydney, Sydney, Australia*
5. *Centre for Mathematical Biology, University of Sydney, Sydney, Australia*

Model-based phylogenetic studies of homologous sequences of nucleotides often assume that the underlying evolutionary process was stationary, reversible, and homogeneous (SRH) over the tree. However, an increasing body of data suggests that evolution under these conditions is an exception, rather than the norm. Moreover, there is growing evidence of convergent evolution, not only at the phenotypic level but also at the genotypic level. In 2014, we introduced a family of mixture models (HAL-HAS) that approximate heterogeneity in the substitution process across lineages and across sites. Subsequently, this model was found to return biased results when convergence had occurred during the evolution. Here we present a new algorithm (HAL-HAS 2) that overcomes these issues. Based on simulation-based analyses of alignments of nucleotides generated on a 4-tipped tree with convergent evolution, the accuracy of phylogenetic estimates improved from 26% to 98%. Results obtained using data sets with 20 or 30 sequences showed that the algorithm is also accurate when a larger number of models are considered. When applying HAL-HAS 2 to a real data set obtained from eight yeast genomes, a model of evolution that includes convergence and rate heterogeneity across lineages as well as sites provides a better fit to these data than other models of evolution do.

557

Evolution of the Shine-Dalgarno (SD) motif in prokaryotes

Deepa Agashe[1], Gaurav D Diwan[1]

1. *National Centre for Biological Sciences (NCBS), Bangalore, KARNATAKA, India*

In prokaryotes, translation initiation typically depends on complementary binding between the Shine-Dalgarno (SD) motif in the 5' untranslated region of mRNAs and the 3' tail of the 16S ribosomal RNA. In coding regions, "internal" SD-like motifs can cause ribosomal pauses and are proposed to face strong negative selection. However, these motifs may also evolve under neutral or positive selection in species that do not use SD-based translation initiation; in AT-rich genomes where the G-rich SD motif is rare; and when translational pauses are beneficial for protein folding or targeting. To determine the nature and consequences of selection acting on internal SD-like motifs, we analysed their frequency in 284 prokaryotes. After accounting for genome GC content, we found that internal SD-like motifs are relatively rare in 227 species, but also in three of seven species that do not use the SD mechanism. The depletion was stronger in GC-rich genomes, mesophiles, and in N-terminal gene regions. In contrast, 50 species either showed no signature of avoidance or were enriched in internal SD-like motifs. Further, C-terminal regions of genes in operons or those closely followed by a downstream gene were relatively enriched in such motifs. Finally, we found no empirical correlation between the predicted and observed fitness impact of altering internal SD-like motifs in a bacterium. Together, our results show that selection against internal SD-like motifs is neither consistently strong nor predictable across genes and species. The evolution of internal SD-like motifs is governed by multiple factors including GC content and genome organization, and other motifs in prokaryotic genomes may evolve under similarly complicated evolutionary pressures.

558

Population structure of widespread *Anopheles* mosquitoes in New Guinea, Australia and the Solomon Islands

Luke Ambrose[1], Robert D Cooper[2], Neil Lobo[3], Nigel W Beebe[4]

1. *The University of Queensland, Ferny Hills, QUEENSLAND, Australia*
2. *Army Malaria Institute, Brisbane, Queensland, Australia*
3. *Eck Institute for Global Health, University of Notre Dame, Notre Dame, Indiana, United States of America*
4. *School of Biological Sciences, The University of Queensland, Brisbane, Queensland, Australia*

The Australo-Papuan region is an area containing high levels of endemism and biodiversity. New Guinea in particular is geologically and biogeographically very complex yet there are only a limited number of studies on population structure that have been performed in this region. I will present a comprehensive microsatellite data set from a widespread and biologically interesting mosquito species, *Anopheles hinesorum,* occurring throughout in this area. I show that there are a number of genetically distinct populations of this species and that some of these populations may in fact be separate species or subspecies as there is no evidence of hybridization or introgression between populations. This work builds on previous phylogeographic work done on the species that revealed strong population structure in this species and evidence that two distinct mitochondrial lineages of *An. hinesorum*are present on the Solomon Islands. Thus there is evidence for two separate dispersals of this species to the archipelago. Interestingly this species bites humans and transmits malaria in New Guinea but does not in the Solomon Islands, raising questions as to whether this potentially medically significant trait has evolved independently in each of the mitochondrial lineages or whether it was transmitted from one to the other via nuclear gene flow.

559

Nonmetric ANOVA for Assessing Phylogenies

Ali Amiryousefi[1], **Jarkko Salojärvi**[2]

1. *Department of Biosciences, University of Helsinki, Helsinki, Finland*
2. *Department of Biosciences, University of Helsinki, Helsinki, Finland*

Classic and Kruskal-Wallis ANOVA methods test the explanatory power of a partitioning by an independent categorical variable on the associated response variables, either using their actual values or ranks. Nonparametric ANOVA extends the classic methodology to a case where a distance metric between data points can be defined, rather than response values themselves. Even though considerably widening the applicability of the ANOVA, it still does not provide a principled framework for the case where the responses do not follow a metric, such as BLAST similarities and distances in non-ultrametric trees. Here we consider ANOVA models for non-metric spaces. We consider cases where metric properties (identity, symmetry, and subadditivity) are each relaxed in turn to derive the resulting nonparametric ANOVA, and then focus on the special case where ranks of the similarity between responses are used. We show that nonparametric ANOVA follows as a special case if the response variable follows a proper metric. As a practical example, the methodology is applied to assess the confidence of bipartitions in a consensus species tree of 149 land plants estimated from pairwise BLAST scores of plastid genes.

561

Experimental evolution with recombining and non-recombining *Acinetobacter baylyi*

Beth J Brittain[1], **Chowdhury Mehedi Hasan**[1], **Daniel Ortiz-Barrientos**[1], **Jan Engelstaedter** [1]

1. *The School of Biological Sciences, The Univeristy of Queensland, St Lucia, QLD, Australia*

Recombination may enhance adaptation for bacterial populations by alleviating competition among clonal lineages carrying different beneficial mutations (clonal interference), and by freeing beneficial mutations from deleterious genetic backgrounds. This benefit of recombination, known as the Fisher-Muller model, is a leading explanation for the prevalence of sexual reproduction in eukaryotes and has been proposed as an evolutionary benefit of genetic exchange in bacteria.

To test for an evolutionary benefit of recombination in bacteria adapting to new environments, we used experimental evolution and whole-genome sequencing with *Acinetobacter baylyi* (which exchange genes via natural transformation). Naturally competent and genetically constructed, non-competent populations were experimentally evolved for 800 generations, in novel media conditions. The extent of adaptation was assessed using competitive fitness assays and beneficial mutations were sought by comparing ancestral and evolved genomes.

Results show that both competent and non-competent bacterial populations adapted to similar extents, and no benefits of recombination were found. Subsequent whole genome sequencing revealed only few candidate adaptive mutations, indicating that adaptation may have been mutation limited (for both types of bacteria). If few beneficial mutations were available to evolving populations, this may explain why there was no apparent benefit to recombination, as the benefits of recombination require that multiple mutations be present concurrently within evolving populations. While previous studies have reported enhanced adaptation of recombining populations, the current results indicate that the conditions under which recombination can enhance bacterial evolution are limited.

562

Inferring symbiotic innovation through a shotgun metagenomics lens: progress and pitfalls of the latest techniques

Amanda M.V. Brown[1], **Dana K. Howe**[1], **Dee R. Denver**[1]

1. *Oregon State University, Corvallis, OR, United States*

Genome and transcriptome sequencing provide a window into the world of beneficial interactions between unculturable microbes and their hosts. Yet, we still face major roadblocks when trying to extract meaning from this mass of data. Missassembly and a large proportion of uncharacterized genes are some of the biggest challenges. These problems are especially common in host-associated symbionts sampled from nature, where divergent strains may exist. Researchers must often choose between expensive high-coverage low-sample sequencing, shallow-coverage limited-diversity sequencing, or smaller-target (fewer loci) sequencing approaches. Using data from several microbial communities associated with below- and above-ground plant pests (specifically, plant-parasitic nematodes and sap-feeding insects), we explored several new approaches to gather meaningful insights into symbiosis from shotgun genome and transcriptome data. Results show larger library inserts and longer reads increase coverage evenness to offset the cost in coverage magnitude, helping with both assembly fragmentation and chimeric assembly, but have limited benefit for untangling strains. Combatting strain diversity by two opposite methods, collapse versus separation, illuminates the need for a sophisticated clustering algorithm to bin strains meaningfully before functional analysis. Finally, one of the most promising facets of this study was the benefit of a comparative approach using shared gene sets and pangenomes to better-handle the proportion of shared, conserved, and sometimes co-expressed but functionally unannotated genes. In summary, we outline a promising, simplified series of steps to bridge some of the formidable rifts in shotgun metagenomics for host-associated microbes.

563

The relative role of positive and background selection in shaping genetic variability at linked sites in *Drosophila melanogaster*

Jose L. Campos[1], **Brian Charlesworth**[1]

1. *University of Edinburgh, Edinburgh, MIDLOTHIAN, United Kingdom*

Selection at one site shapes patterns of genetic variability of nearby sites. Both the fixation of advantageous mutations (selective sweeps) and the removal of deleterious mutations (background selection) can reduce genetic diversity at nearby sites. However, the relative importance of recurrent positive selection and background selection is not well understood. The observation of a negative correlation between levels of synonymous site variability (π_S) and the rate of protein evolution (K_A) in genes across the genome of *Drosophila melanogaster* has been interpreted as support for a major role of recurrent positive selection. However, background selection can also cause such a pattern when recombination within a gene is sufficiently low. This means that genes under lower selective constraints are more susceptible to background selection and may show less genetic variation. Here we analyse the patterns of genetic variation across the autosomal genes of an African *D. melanogaster* population to test these predictions, taking into account the effects of other genomic features that covary with π_S and K_A. The results suggest that both background selection and recurrent positive selection determine the negative relation between synonymous diversity and protein sequence divergence.

564

Comprehensive Annotation of Multigenic Protein-Family Structures (CAMPS)

David Clarke, Lars S Jermiin[1], Charles Robin[2]

1. CSIRO, Canberra, ACT, Australia

2. The University of Melbourne, Parkville, VIC, Australia

The automated methods currently used for genomic annotation of gene structure often fall short of the desired degree of accuracy appropriate for analysis of multigenic protein-families within and between species. These methods are restrained by a poor ability to distinguish the relationships of stretches of coding sequence to each other and an over reliance on a single inevitably flawed genome assembly. To increase the accuracy of gene structural annotation the CAMPS annotation methods focus on a single protein-family at a time. In this context 'Comprehensive' refers to incorporating multiple lines of evidence from the target species, including multiple genome assemblies and different sources of transcriptome data.

Genomic loci for the protein family in question are initially identified through sequence similarity with previously identified proteins from the target species or other species. Additional evidence is taken from the transcriptomes of the target species and other species. CAMPS clusters the identified loci and transcripts into 'campsites' through sequence similarity and genomic position. A campsite is divided into 'tents', which include variants of suspected genes. Annotation of the suspected gene is performed using all data within the tent. Sequence variants within the tent are identified and then classified.

These methods are designed to; reduce the concatenation and splitting of genes; identify partial and duplicate gene sequences present within the dataset; and compare data across the available genomic assemblies and transcriptomes to identify where assembly errors have caused problems. Accuracy estimates are generated for individual genes from the evidence used to derive the gene models. CAMPS is also designed to present multiple sequence variants of genes when they are found in the dataset.

565

Evolution of Tumor Necrosis Factor Superfamily (TNFSF) genes TNFSF 12 and 13: Phylogenetic clues for the emergence of *in genomic* fusion of *TNFSF12-TNFSF13*

Sabyasachi Das[1], Yoichi Sutoh[1], Masayuki Hirano[1], Qifeng Han[1], Jianxu Li[1], Max D. Cooper[1], Brantley R. Herrin[1]

1. Emory University, Atlanta, GA, United States

BAFF and APRIL (TNFSF13 members) are important regulatory factors for lymphocyte activation and survival. Another TNFSF13 gene that encodes a BAFF and APRIL-like molecule (BALM) was found in fish. Here we report the molecular characterization of the first TNFSF13 homolog in lampreys, a jawless vertebrate representative. In an investigation of the evolution of *BAFF*, *APRIL*, *BALM* and a closely related gene called *TWEAK* (*TNFSF12*), we identified an ancestral TNFSF13 gene in jawless vertebrates, but not the *TWEAK* gene. Hence, while *TWEAK* evolved in jawed vertebrates the*TNFSF13* gene appeared before the divergence of jawed and jawless vertebrates. Considering the encoded protein of ancestral *TNFSF13* gene in lamprey possess more BAFF like features than that of APRIL, we could notice that *BAFF* is present in all vertebrates, but *APRIL* and/or *BALM*independently lost in different lineages. For example, *BALM* is absent in all tetrapod genomes, and *APRIL* is lost in birds and several fish species. *TWEAK* is also lost in bird lineage suggesting that the genetic network of immune related genes have greatly reconstructed in birds genome. The comparative genome and transcriptome analyses suggest that an *in-genomic* fusion between *APRIL* and closely related *TWEAK* genes that produce a hybrid molecule called TWE-PRIL originated in mammalian lineage. Like mammalian *BAFF* and *APRIL* the ancestral TNFSF13 in lamprey exhibits a wide range of tissue and cellular expression including innate lymphoid cells and T cell-like (VLRA and VLRC) and B cell-like (VLRB) lymphocytes in early stage of vertebrate evolution.

566

Mechanism and fitness benefits of pyruvate kinase: a recurrent target of evolution

Kristina Duan[1], Fen Peng, Timothy F Cooper[1]

1. Biology and Biochemistry, University of Houston, Houston, TX, USA

Beneficial mutations in *pykF*, which encodes pyruvate kinase, a major control-point enzyme of glycolysis, occurred in 12 independently evolved experimental populations of *Escherichia coli* propagated in a low glucose environment. We found that the evolved *pykF* mutations cause higher levels of its substrate, phosphoenolpyruvate (PEP), to be available at the start of each daily growth cycle. We hypothesize that higher PEP levels allow a quicker restart of the PEP-dependent glucose uptake system, providing a benefit by shortening lag times. Consistent with this hypothesis, competitive fitness assays show that selection favors *pykF* mutants most strongly in lag phase. This advantage is not general, being exclusive to resource environments in which PEP is necessary for sugar transport. We also found that the *pykF* mutations increased growth rates on gluconeogenic substrates, as expected if their changed function reduced consumption, and thus futile cycling, of PEP. Lastly, we manipulated expression of the ancestral *pykF* to test if fitness benefits could be reproduced through changes only in overall enzyme activity. Up- or down-regulating *pykF* did not mimic the fitness benefit conferred by mutant enzymes, suggesting that this benefit depended on a change in the enzyme's functional properties. Our study represents a rare comprehensive test of a proposed metabolic basis of the benefit of a naturally selected beneficial mutation. The mechanism of benefit is likely to be general and may be relevant to understanding the benefit conferred by changes in pyruvate kinase that are common in human cancer cells.

567

Genetic basis of Lactase Persistence in Ethnically diverse African Populations

Elizabeth Eyermann[1], Alessia Ranciaro[1], Simon Thompson[1], Dawit Woldemeskel[2], Stephanie Fagbemi[1], William Beggs[1], Gurjia Belay[2], Thomas Nyambo[3], Sarah A Tishkoff[1, 4]

1. Department of Genetics, University of Pennsylvania, Philadelphia, PA, USA

2. Department of Biology, Addis Ababa University, Addis Ababa, Ethiopia

3. Department of Biochemistry, Muhimbili University of Health and Allies Sciences, Dar es Salaam, Tanzania

4. Department of Biology, University of Pennsylvania, Philadelphia, PA, USA

In most humans, levels of lactase, an enzyme encoded by *LCT*, decrease after weaning. This leads to a reduced ability to digest lactose, referred to as "lactase non-persistence" (LNP). However, some individuals maintain high expression of lactase and are able to digest lactose into adulthood, referred to as "lactase persistence" (LP). LP is a recent adaptive trait in humans and has evolved in populations that practice milk production and consumption. We sequenced 594 bp in intron 13 of *MCM6,* a candidate enhancer region for *LCT* where LP-associated variants have been identified in European and African populations. Our sample consists of 641 individuals from 4 Tanzanian populations and 650 individuals from 19 Ethiopian populations that practice diverse subsistence methods. In addition to genetic data, we used phenotypic data from a lactose tolerance test to conduct a genotype-phenotype association analysis in 209 Tanzanians and 150 Ethiopians. The association between LP and the C-14010 variant in Tanzania was confirmed. For the first time we identified the presence of the C-14010 and G-13915 variants in the Hadza of Tanzania, who are traditionally hunter-gatherers. The G-13915 variant had not been identified in Tanzanian populations previously studied. The G-13907 variant was not found in the Tanzanian populations, though we identified it in the Ethiopians, consistent with work done by previous researchers indicating that this variant originated in Northeast Africa. Network analyses and tests of natural selection were performed to reconstruct the evolutionary history of this locus and to trace historic migration events in East Africa.

568

Genetic origins of bushpigs from Madagascar

Jaime Gongora[1], Steven M Goodman[2, 3], Miguel Pedrono[4], Guillaume Besnard[5], Carol Lee[1], Laurent Frantz[6], Peter J Taylor[7]

1. *Faculty of Veterinary Science, School of Life and Environmental Sciences, The University of Sydney, Sydney, NSW 2006, Australia*
2. *Field Museum of Natural History, Chicago, Illinois 60605, USA*
3. *Association Vahatra, Antananarivo 101, Madagascar*
4. *CIRAD, U.R. AGIRs, Antananarivo 101, Madagascar*
5. *Laboratoire Evolution et Diversité Biologique, Université Toulouse III Paul Sabatier, 31062 Toulouse, France*
6. *School of Archaeology, University of Oxford, Oxford OX1 3QY, UK*
7. *School of Mathematical and Natural Sciences, University of Venda, Thohoyandou, Limpopo Province, South Africa*

The island of Madagascar, situated off the east coast of Africa, was settled by Austronesians (proto-Malagasy) ~1500–2000 years ago and immediately after by native East African groups. Bushpigs of the genus *Potamochoerus* are suggested to have been introduced to Madagascar from eastern mainland Africa and/or offshore islands. The earliest archaeological evidence for bushpigs in Madagascar dates to the 10th-13th centuries and possibly on the Comoro Islands in the 9th-10th centuries. Although the circumstances of the translocation are unclear, it has been proposed that the specific identification of Malagasy bushpigs is *P. larvatus* from sub-Saharan Africa, which could have been transported directly into Madagascar across the Mozambique Channel or through a corridor via the neighbouring islands by early sea navigators who settled in Madagascar. Furthermore, two subspecies/populations of Malagasy bushpigs have been nominated, *P. l. hova* and *P. l. larvatus* from eastern and western Madagascar, respectively. It has been proposed that the former population of bushpigs may have originated from the southern African populations of *P. l. koiropotamus*, which range from mid-Tanzania southwards. However, genetic evidence to make definitive conclusions on the taxonomic status and geographical origins of these Malagasy wild bushpigs is currently not available. To contribute to this debate, we investigate the phylogenetic position of Malagasy bushpigs in relation to other species of African and Eurasian Suidae and assess their relationships with other bushpig populations from mainland Africa using mitochondrial DNA. Our preliminary results show that the Malagasy bushpigs cluster within the genus *Potamochoerus*. Analyses of further samples from mainland Africa and biparental DNA markers are underway to better pinpoint the evolutionary relationships of Malagasy bushpigs with the recognised species of this genus and the geographical source of the populations.

569

COMBINE, bringing together Australian students and early-career researchers in bioinformatics and computational biology

Jane Hawkey[1]

1. *The University of Melbourne, Parkille, VIC, Australia*

COMBINE is a student-run Australian organisation for researchers in computational biology and bioinformatics. COMBINE is the official International Society for Computational Biology (ISCB) Regional Student Group (RSG) for Australia.

We aim to bring together students and early-career researchers from the computational and life sciences for networking, collaboration, and professional development.

Australia has many research institutes, each with their own cohorts of students. Aside from conferences, there are few opportunities that bring these students together, allowing them to discover the different kinds of research going on at other institutes.

COMBINE aims to bridge this institutional divide by organising seminars, workshops and social events, and the yearly COMBINE Student Symposium. Together, these events allow students to connect with each other and build a professional network in a casual environment.

Find out more and get involved at http://combine.org.au/

570

Biogeography of sex reversal and the effects of climate on sex determination

Clare Holleley[2, 1], Stephen Sarre[2], Arthur Georges[2]

1. *CSIRO, Canberra, ACT, Australia*
2. *University of Canberra, Canberra, ACT, Australia*

Recent advances have demonstrated that reptiles can rapidly switch from a sex determination mode that is predominantly genetic (using sex chromosomes) to a system where sex is determined by egg incubation temperature (without sex chromosomes)[1]. This is achieved through the occurrence of sex reversal in wild populations – where chromosomal males are feminised at high incubation temperature. Sex reversal can rapidly trigger the evolution of new sex determining modes, by facilitating the loss of the W sex chromosome and thus a transition from genetic sex determination to temperature dependent sex determination. Here we use a continental-scale data set to characterize the genotypic and phenotypic sex of bearded dragon populations (*Pogona vitticeps*) and identify populations with sex reversal events. We then explore whether these events occur in locations experiencing a rapid increase in diurnal temperature range, an increase in mean temperatures, or record-breaking warming

events. We also explore the possibility of using specimens from Australia's National collections to track temporal changes in the rates of sex reversal over the last century.

1. Holleley CE, O'Meally D, Sarre SD, Graves JAM, Ezaz T, Matsubara K, Azad B, Zhang X, Georges A. 2015. Nature 523 (7558), 79-82

571

The Hsp90 chaperone protein is essential in the *Shewanella oneidensis* bacteria under stress condition

Flora Honoré[1], Vincent Méjean[1], Olivier Genest[1]

1. CNRS, Marseille, FRANCE, France

Molecular chaperones are proteins that assist the folding of other proteins. In stress conditions, they are crucial as proteins become misfolded. The Hsp90 chaperone is present in almost all living organisms and is essential in eukaryotes. In contrast, Hsp90 is dispensable in proteobacteria and few client proteins have been identified for bacterial Hsp90. Therefore, to uncover the role of Hsp90 in bacteria, new bacterial models are needed.

Bacteria from the *Shewanellae* genus are gamma-proteobacteria widely found in aquatic environments. *Shewanellae* have the astonishing ability to adapt to many stress conditions: large range of temperatures and salt concentrations, extremely high hydrostatic pressure since they can be found in deep-sea, and growth in very poor media. These environmental conditions being known to affect protein folding, it suggests the *Shewanellae* possess a large diversity of powerful chaperone proteins.

Our research aims to elucidate the function of Hsp90 in bacteria by using *Shewanella oneidensis* as a model organism. We first constructed an *hsp90* deletion strain in *S. oneidensis* and found that in contrast to wild type, the mutated strain was unable to grow at high temperature. Taking advantage of this phenotype, we identified proteins that can suppress the absence of Hsp90 at high temperature by using a *S. oneidensis* library of plasmids. This genetic selection revealed several protein candidates that are currently under study. It is anticipated that our study using *S. oneidensis* as a new bacterial model will lead to a better understanding of the role of Hsp90 in bacteria.

572

Illuminating *Limbodessus*: speciation of sister species within a genus of subterranean dytiscid beetles in Western Australia.

Josephine C A Hyde[1], Andrew D Austin[1], Pablo Munguia[2], William F Humphreys [3], Remko Leijs[4], Steven J B Cooper[1, 4]

1. Australian Centre for Evolutionary Biology and Biodiversity and School of Biological Sciences, The University of Adelaide, Adelaide, SA, Australia

2. School of Biological Sciences, The University of Adelaide, Adelaide, South Australia, Australia

3. Western Australian Museum, Perth, Western Australia, Australia

4. Evolutionary Biology Unit, South Australian Museum, Adelaide, South Australia, Australia

The Yilgarn Region of Western Australia contains a rich diversity of subterranean invertebrates and comprises hundreds of physically isolated calcrete aquifers, resembling a subterranean archipelago. Each calcrete has a unique combination of aquatic species including diving beetles (Dytiscidae). Two different genera of diving beetle are found in the aquifers: *Paroster* and *Limbodessus* within which ~100 species have now been described. Phylogenetic analyses based on mitochondrial DNA (mtDNA) suggest that most (~85%) of the species have independently evolved from surface ancestors, but also provide evidence for the existence of 11 sister species pairs and two sister species triplets living in sympatry within the same calcrete. This finding raises the possibility that speciation occurred underground from a common ancestor already pre-adapted to subterranean life. Here we focused on*Limbodessus* and aimed to test the sister species status of key taxa using multiple independent nuclear DNA markers. A further aim was to provide a robust phylogeny for the genus using seven genes, three mitochondrial (COI, 12S, 16S) and four nuclear (CN, WG, TOPO, ARK). These genes were sequenced for 55 species of *Limbodessus*, three species of *Paroster* and one species of *Allodessus*; the latter two genera were used as outgroups. Phylogenetic analyses of individual genes for the 55 *Limbodessus* species included 13 that had previously been suggested to be part of a sister pair or triplet. These analyses supported the sister species status of five pairs of taxa, but not the status of the triplet, suggesting that potentially there has been some hybridisation in the past leading to introgression of mtDNA between these three species. Overall, our study shows that *Limbodessus* contains multiple examples of independent speciation underground in either sympatry or parapatry. Further research is required to determine the selective forces that are operating to drive these patterns of speciation.

573

Testing the parent-of-origin hypothesis in dasyurid telomere length dimorphism

Emory Ingles[1], Janine Deakin[1]

1. Institute for Applied Ecology, University of Canberra, Canberra, ACT, Australia

The marsupial family Dasyuridae is characterised by an unusual and non-random pattern of telomere length distributions. On any given chromosome pair, one homologue will have short and the other long telomeres. In males, it is always the Y chromosome that has long telomeres, leading to the hypothesis that the distribution of telomere length is based on parental origin of chromosomes (the parent-of-origin hypothesis). This project aims to test this hypothesis, in particular using Tasmanian devils, antechinus, quoll and dunnart species. Because marsupials inactivate the paternally-derived X chromosome in females as part of a dosage compensation mechanism, it is possible to identify the parent origin of sex chromosomes in females using epigenetic immunofluorescence. Specifically, the maternally-derived active X is hypermethylated, the paternally-derived inactive X hypomethylated, and there is differential staining of the histone active marks H3K9ac, H4Kac, H3K4me2 between the active and inactive X chromosomes. In conjunction with telomere quantitative fluorescence *in* situ hybridisation (qFISH) to determine telomere length at the chromosome level, this will allow us to test if telomere length is correlated with chromosome parent origin in female individuals. Preliminary results for methylation staining on devil and quoll support the parent-of-origin hypothesis, with the paternally derived, hypomethylated X being the homologue with long telomeres. To test for any male specific telomere elongation prior to fertilisation, telomere qFISH will further be used to investigate changes in telomere length throughout the male germ line, as well as in sperm. Taken together, these experiments will provide insight into the workings and regulation of telomere length dimorphism in dasyurid species.

574

An evolutionary perspective of AMPK-TOR signaling in the three domains of life

Arpit Jain[1], Valentin Roustan[2], Markus Teige[2], Ingo Ebersberger[2], Wolfram Weckwerth[2]

1. Goethe University, Frankfurt Am Main, HESSEN, Germany

2. University of Vienna, Vienna, Austria

AMPK and TOR protein kinases are the major control points of energy signalling in eukaryotic cells and organisms. They form the core of a complex regulatory network to coordinate metabolic activities in the cytosol with those in mitochondria and plastids. Despite its relevance, it is still unclear when and how this regulatory pathway was formed during evolution, and to what extent its representations in the major eukaryotic lineages resemble each other. Here we have traced 153 essential proteins forming the human AMPK-TOR pathways across 413 species representing all three domains of life, and subsequently through time. Further, we have characterized these traced proteins based on their feature architecture similarities to obtain a functionally meaningful interpretation of the phylogenetic profiles. The resulting phylogenetic profiles indicate the presence of primordial core pathways including 7 proto-kinases in the last eukaryotic common ancestor. The evolutionary roots of the oldest components of the AMPK pathway, however, extend into the pre-eukaryotic era, and descendants of these ancient proteins can still be found in contemporary prokaryotes. The TOR complex in turn appears as a eukaryotic invention, possibly to aid in retrograde signalling between the mitochondria and the remainder of the cell. Within the eukaryotes, the two pathways display beyond the conserved cores, a considerable plasticity. Most notably, KING1, the protein originally assigned as the gamma subunit of AMPK in plants, is more closely related to the yeast SDS23 gene family then to the gamma subunits in animals or fungi. This suggests that also its functionality differs from that of a canonical AMPK gamma subunit.

575

Life history evolution in response to condition-dependent juvenile selection

Chris Kimber[1], Urban Friberg[1]

1. Linköping University, Linköping, ÖSTERGöTLAND, Sweden

Why organisms age the way they do remains a fundamental question in evolutionary biology. Ageing was traditionally attributed to alleles that are deleterious only in late life, but it has recently become clear that this explanation is incomplete. Mutations affecting both early reproduction and late life survival in the same direction are also common, likely because they affect overall condition. This positive pleiotropy potentially plays an important and unappreciated role in the evolution of lifespan, but how it affects the evolution of ageing itself remains undocumented. Furthermore, little is known about the link between selection on pre-adult condition and survival rate later in life. To investigate how selection acting on mutations that affect organismal condition at the juvenile stage indirectly influences survival, ageing rate and the trajectory of life history in general, we employ an experimental evolution approach in Drosophila melanogaster. Beginning with a large outbred laboratory population, we imposed two distinct novel selection environments that were confined to the juvenile stage of the life cycle. Each environment was designed to impose selection that should reduce mutational load affecting a wide range of traits that capture overall organismal condition. We can then study how reduction of mutational load at the juvenile stage influences adult life history pleiotropically. We will present data that addresses the effect of the selection environments on mutational load and the evolution of juvenile life history as a result of selection, and explores pleiotropic changes in adult life history traits as a consequence of selection on juvenile condition.

576

Historical biogeography of the ancient genus *Selaginella* – Early adaptation to xeric habitats on Pangea

Kristina Vanessa Klaus[1], Christian Schulz[1], Dennise Stefan Bauer[1], Thomas Stützel[1]

1. Ruhr Universität Bochum, Bochum, NORTH RHINE-WESTPHALIA, Germany

Selaginella is an ancient genus of the lycophytes with fossils dating back to the Early Carboniferous (Rowe 1988). Plants of this genus are creeping to erect growing, herbaceous plants. The genus *Selaginella* has a cosmopolitan distribution with main distribution areas in the tropics and subtropics of Asia and South America (Jermy 1990). The genus comprises approximately 700 species (Jermy 1990) and 6 subgenera, according to the most recent classification (Zhou and Zhang 2015). Here, we present a global time tree for the genus *Selaginella* based on a concatenated sequence dataset containing DNA marker regions *rbc*L and ITS1 5.8S ITS2. The taxon sampling comprises 200 species from all major distribution areas. The time calibrated phylogeny was estimated using the BEAST package (Drummond et al. 2012). Calibration was performed using age records of the fossils*Lepidosigillaria* (Bateman 1992) and *Selaginella resimus* (Rowe 1988) which could be assigned to specific nodes within the phylogeny. The R package BioGeoBears (Matzke 2013) was used to estimate ancestral areas and to test for altering reconstruction models. The biogeographic reconstructions suggest that *Selaginella* originated in Laurussia (Euramerica), the second largest land mass of the Devonian and Early Carboniferous. The divergence of the two major lineages within *Selaginella* and an ecological adaptation to xeric habitats already took place in the Early Permian, when the two major land masses Laurussia and Gondwana merged together and formed the supercontinent Pangea. Ancestral areas of the subgenera, as well as historical long distance-dispersal versus vicariance events, were discussed under the aspect of plate tectonic and historical climate conditions.

1. Bateman, R.M. (1992) Morphogenetic reconstruction, palaeoecology, and phylogeny of Oxroadia gracilis Alvin emend., and O. coniferata sp. nov.: anatomically preserved lycopods from Oxroad Bay, SE Scotland. Palaeontographica Abteilung B, 228, 29-103.
2. Drummond, A.J., Suchard, M.A., Xie, D. & Rambaut, A. (2012) Bayesian phylogenetics with BEAUti and the BEAST 1.7. Molecular biology and evolution, 29, 1969-1973.
3. Jermy, A.C. (1990) Selaginellaceae. Pteridophytes and gymnosperms (ed. by K.U. Kramer and P.S. Green). Springer, 39-45.
4. Matzke, N. (2013) BioGeoBEARS: BioGeography with Bayesian (and Likelihood) Evolutionary Analysis in R Scripts. University of California, Berkeley, Berkeley, CA.
5. Rowe, N.P. (1988) A herbaceous lycophyte from the Lower Carboniferous Drybrook Sandstone of the Forest of Dean, Gloucestershire. Palaeontology, 31, 69-83.
6. Zhou, X.-M. and Zhang, L.-B. (2015) A classification of Selaginella (Selaginellaceae) based on molecular (chloroplast and nuclear), macro-morphological, and spore features. Taxon, 64, 1117-1140.

577

Trait-dependent dispersal models for phylogenetic biogeography, in the R package BioGeoBEARS

Nicholas Matzke[1]

1. The Australian National University, Turner, ACT, Australia

Organism traits must be important in historical biogeography. In particular, rates of dispersal (both range-expansion dispersal, and jump dispersal leading to founder-event speciation) must depend to some degree on traits such as flight and its loss, and seed dispersal mechanisms and the dispersal abilities of animals that transport seeds. However, to date no probabilistic historical biogeographical models have been available that allow geographic range and traits to co-evolve on the phylogeny, with traits influencing dispersal ability. In purely continuous-time Markov models, adding a trait is just a matter of doubling the size of the rate matrix; however, biogeographical models also include a much more complex discrete-time model describing how geographic range can change during cladogenesis. Traits might also influence this process. I present an addition to the R package BioGeoBEARS that enables an evolving discrete trait to influence dispersal ability for both anagenetic and cladogenetic range change. This model can be freely combined with models adding jump dispersal (e.g., DEC+J), distance as a predictor of dispersal (+x models, with dispersal rate multiplied by distance^x), and other variants. I test the model against simulations and datasets where large evolutionary changes in dispersal ability are highly likely (e.g., Pacific rails, which have repeatedly lost flight).

578

Comparison of Neanderthal and modern human Y chromosomes: implications for reproductive isolation

Fernando L Mendez[1], David Poznik[2], Sergi Castellano[3], Carlos D Bustamante[1, 4]

1. Department of Genetics, Stanford University, Stanford, California, United States

2. 23andMe, Mountain View, California, United States

3. Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany

4. Depatment of Biomedical Data Science, Stanford University, Stanford, California, United States

The sequencing of archaic humans has transformed our understanding of human origins. Neanderthals have contributed genetic variants to modern populations, potentially influencing multiple phenotypes. However, large sections of the Neanderthal genome are not represented in the genomes of modern populations, including the Y chromosome. It has been suggested that genetic incompatibilities played a role in the isolation of archaic and modern humans. We present the analysis of ~120 kb of exome-captured Y-chromosome DNA from a Neanderthal individual from El Sidrón, Spain. Comparing this sequence with those of the human and chimpanzee references, and with those of two A00 Y chromosomes, we: 1) determined the branching order of the human lineages, 2) estimated the time to the most recent common ancestor (TMRCA) between modern human and Neanderthal lineages, and 3) assessed differences in the coding sequence of their Y chromosomes. We determined that the Neanderthal Y chromosome lineage branches out before modern human lineages diversified. The TMRCA estimate of ~588 thousand years ago, approximately doubles that of the human reference and the A00 lineage. We observed nine protein-coding changes between Neanderthal and modern human lineages, four of which (in genes *PCDH11Y*, *TMSB4Y*, *USP9Y*, and *KDM5D*) are predicted to have functional importance. Three of these four genes are known to result in male-specific minor histocompatibility antigens, and might be implicated in the reproductive isolation of Neanderthals and modern humans. We discuss the implications of our finding and our future steps.

579

Evolutionary insights from the marine metagenomics in the Red Sea

Katsuhiko Mineta[1], Yoshimoto Saito[1], Intikhab Alam[1], Hayedeh Behzad[1], Mohamed Alarawi[1], Muhammad S Amini [1], Hajime Ohyanagi[1], Magbubah Essack[1], John Archer[1], Vladimir Bajic[1], Takashi Gojobori[1]

1. King Abdullah University of Science and Technology, Thuwal, Saudi Arabia

The Red Sea is the world's northernmost tropical sea, positioned between Africa (African continent) and Asia (Arabian Peninsula). The Red Sea has several unique features; high temperature, high salinity, and low nutrient level. These features are characteristic of harsh environments that compel organisms' to adapt in order to survive or become extinct. Not much is known about the functional gene content and evolutionary process of the organisms in this harsh Red Sea environment. To better characterize the gene content and variability of its microbial community, we took an approach using metagenomics. Metagenomics is one of the technologies used to produce a large number of sequences from organisms in the environment, providing a comprehensive view into the marine microbial communities and their ecosystems. We have now been conducting the Red Sea metagenome project since 2014. Seawater and sediment samples were collected and sequenced on a monthly basis. We utilize these data to see the dynamics of the microbial community as well as to examine the species and genes in relation to the evolution. In this presentation, we will provide an overview of our Red Sea metagenome project and detail the outcome of our metagenomics analyses with the evolutionary insights.

580

Adaptive Landscapes of Resistant Genes change as Antibiotic Concentration Changes

Portia Mira[1], Miriam Barlow[1], Juan Meza[1]

1. UC Merced, Merced, C, United States

Most studies about the evolution of antibiotic-resistance are focused on selection for resistance at lethal antibiotic concentrations (Hughes, Andersson 2012), which has allowed the detection of mutant strains that show strong phenotypic traits. However, solely focusing on lethal concentrations of antibiotics narrowly limits our perspective of antibiotic resistance evolution. New high-resolution competition assays have shown that resistant bacteria are selected at relatively low concentrations of antibiotics (Hughes, Andersson 2012). This finding is important because, sub-lethal concentrations of antibiotics are found widely in non-medical conditions, such as wastewater treatment plants, and food and water used in agriculture and farming. To understand the impacts of sub-lethal concentrations on selection we computed thirty adaptive landscapes for a set of TEM β-lactamases containing all combinations of the four amino acid substitutions that exist in TEM-50 for 15 β-lactams at multiple concentrations. We found that there are many evolutionary pathways within this collection of landscapes that lead to nearly every TEM-genotype that we studied. While it is known that the pathways change depending on the type of β-lactam, this study demonstrates that the landscapes also change dramatically as the concentrations of antibiotics change. Based on these results we conclude that the presence of multiple concentrations of β-lactams in an environment likely accelerates the selection of numerous TEM variant genotypes within that environment.

581

Host response to highly pathogenic avian influenza infection in quails

Katrina Morris[1], Jacqueline Smith[1], Ian R Paton[1], Angela F Denner[2], Robert G Webster[2], David W Burt[1]

1. Roslin Institute, University of Edinburgh, Easter Bush, MIDLOTHIAN, United Kingdom

2. Virology Division, Department of Infectious Diseases, St. Jude Children's Research Hospital, Memphis, USA

Highly pathogenic influenza A viruses (HPAI), such as H5N1, are responsible for enormous economic losses in the poultry industry and pose serious threat to public health. Quail is a popular domestic poultry species raised for meat and eggs in Asia and Europe. While quails can survive infection with Low Pathogenic influenza viruses (LPAI), they experience high mortality when infected with strains of HPAI. Previous research has shown that quail may play a key role as an intermediate host in evolution of avian influenza, allowing viral strains to spread from wild birds to chickens and mammals. While aquatic reservoir species such as duck are resistant to most HPAI strains and act as natural reservoirs, Galliformes, including quails and chickens, are highly susceptible. To better understand the effect this disease has on quails we performed differential analysis of gene expression in quails infected with low and high pathogenic strains of Influenza A. We compare this to previous findings in ducks and chickens. As in chickens, quails lack a key intracellular receptor for viral single stranded RNA, RIG-I. In addition they do not show strong upregulation of IFITM proteins, which are thought to be key to HPAI tolerance in ducks. In comparison to chickens, quails had a sustained response to HPAI infection, however this did not translate into a longer survival time. This data will be important for management of HPAI in quails and other bird species.

582

Risk loci on CFA13 associated with lymphoma in Bullmastiffs

Sally Mortlock[1], Mehar S Khatkar[1], Peter Bennett[1], Peter Williamson[1]

1. School of Life and Environmental Sciences, Faculty of Veterinary Science, The University of Sydney, Camperdown, NSW, Australia

Limited diversity within dog breeds has resulted in the accumulation of deleterious alleles predisposing breeds to diseases. Lymphoma is the one of the most common malignancies seen in dogs, affecting 1.2% of the population. Research has shown that Bullmastiffs have a high incidence of lymphoma, suggesting a genetic predisposition. The aim of this study was investigate the incidence of lymphoma in the Australian Bullmastiff population and identify risk loci predisposing the breed to lymphoma. Survey data and data analysed from clinical and pathology databases revealed a high incidence of lymphoma in dogs' ≤6 years. A total of 194 Bullmastiff dogs were genotyped using the 170,000 SNP Illumina CanineHD Beadchip. Information on genetic structure of the population was incorporated into a mixed linear model in GCTA (18 cases, 29 controls) to detect loci associated with lymphoma risk in dogs ≤ 6 years. An association was detected across a 5.4 Mb region on CFA13, 67 SNPs reaching chromosome-wise significance (FDR < 0.05). The region was fine mapped to ~1.2 Mb using both haplotype association and homozygosity analysis. Five risk haplotypes were identified that were significantly associated with lymphoma, 77% of cases were homozygous for risk haplotypes compared to <11% of controls. The associated 1.2 Mb region accounted for over 23% of disease liability in Bullmastiff dogs. One 3'UTR variant and 380 intergentic variants were identified in the associated region using DNA sequence data. Potential functional candidates in the region include *MYC*, predicted precursor coding regions for *miR-1204, miR-1205 and miR-1206* and a region downstream of *MYC* syntenic to human *PVT1,* all have ontologies related to cell cycle progression, cell proliferation and cancer. Through our analysis we have identified a region on CFA13 as a risk locus for lymphoma in Bullmastiff dogs. Validation of variants in the CFA13 region is underway and investigation into functional implications.

583

Experimental and theoretical approaches to elucidate isochore evolution by ENU mutagenesis

Saotoshi Oota[1], Ryutaro Fukumura[1], Naruya Saitou[2], Yoichi Gondo[1]

1. RIKEN, Tsukuba, IBARAKI, Japan

2. National Institute of Genetics, Mishima, Japan

One of the main issues of the mammalian genome evolution is the "isochore:" a queer spatial structure of the genome in GC content. So far, no flawless explanations were achieved due to contradictory observations against the proposed models: e.g., selection, biased gene conversion, and mutation bias models. This indicates that the conventional frameworks of the molecular evolution may be inadequate to understand the fundamental mechanisms harbored in the isochore evolution. The typical approach to study molecular evolution is the "backward" analysis by using extant molecular data: we infer the past by extrapolating relatively recent evolutionary patterns based on observable data. In the isochore evolutionary study, however, this conventional approach has at least two kinds of issues: (1) critical evolutionary signal for the isochore evolution might be easily eroded during the evolutionary process; (2) to directly analyze the ishcore evolution, we need to handle non-coding regions, for which conventional evolutionary models may be inappropriate. To overcome those problems, we took advantage of ENU mutagenesis as a tool for the experimental evolution, accelerating evolutionary rate to make it possible to observe ongoing evolution through a large number of de novo mutations. In other words, we attempted "forward" analysis on the genome evolution. We found bidirectional (diverged) mutation pressures that support the legendary Sueoka's mutation bias theory. Our finding has potential to explain the enigmatic isochore evolution.

584

Molecular co-evolutionary insights into the HIV-1 pre-integration complex

Madara Hetti Arachchilage[1], Helen Piontkivska[1]

1. Kent State University, Kent, OH, United States

The pre-integration complex (PIC), mediated by the viral proteins reverse transcriptase (RT) and integrase (IN), plays an important role in HIV-1 life cycle. The formation and stabilization of the PIC involve direct and indirect interactions between the RT and IN proteins, although the mechanisms of the IN-RT interactions within the PIC remain to be poorly understood due to the transient nature of the protein complex and the intrinsic flexibility of its components. In this study we used the coevolutionary analysis of amino acid sequences of these two proteins to identify potentially interacting regions between the IN and RT proteins within the PIC. Our results enabled us to identify a set of regions with strong coevolutionary signatures, indicating likely direct and prolonged interactions between them that require high affinity and/or specificity, and hence, higher degree of coevolution. Likewise, such regions may participate in interactions mediated by dynamic conformational changes, including both direct and indirect interactions. Other set of regions was found to have relatively weak, albeit positive correlations, indicating that such regions likely participate in interactions that are transient and/or have low affinity. Overall, for proteins with poorly resolved three-dimensional structures, such as IN, coevolution analysis provides important insights for better understanding of functional interactions and enables identification of specific segments and/or residues involved in interaction interfaces.

585

Robustness of Bayesian molecular dating to tree prior misspecification

Andrew M Ritchie[1], **Nathan Lo**[1], **Simon YW Ho**[1]

1. School of Life and Environmental Sciences, University of Sydney, University Of Sydney, NSW, Australia

The practice of specifying prior probabilities is a defining feature of Bayesian phylogenetics. In the case of Bayesian molecular dating, the specification of priors on divergence times, also known as tree priors, is especially contentious. Existing tree priors include pure-birth and birth-death priors used for species-level data and coalescent priors designed for population-level data. All tree priors make strong assumptions regarding the nature of the underlying evolutionary process and are likely to be misspecified to some degree for many real data sets. However, the assumption that sufficient sequence data can overcome any biases in the prior has meant that there has been little empirical investigation into the behaviour of real analyses under different tree priors. This could lead to undetected errors being introduced into analyses of non-conforming datasets, such as those including a mixture of between- and within-species relationships.

We tested the robustness of Bayesian analyses to increasing degrees of tree prior misspecification by simulating mixed inter- and intra-species datasets along a continuum from few species with many individuals (coalescent-like) to many species with few individuals (pure-birth-like). We estimated divergence dates under each prior and compared the analyses for accuracy, precision and model fit. We confirmed the applicability of our results to three empirical data sets for cetaceans, phocids and whitefish.

Our results suggest that Bayesian dating is quite robust to the choice of tree prior in most cases, even when it is severely misspecified for the data. However, simpler priors such as the pure birth prior can produce inaccurate results on some heterogeneous data sets even with substantial amounts of sequence data. More complex priors show greater robustness and should be preferred where more extensive model testing is not practicable.

586

Coinfection study reveals immune modulatory effects of a cestode in its vertebrate host

Marc Ritter[1], **Agnes Piecyk**[1], **Martin Kalbe**[1]

1. Max Planck Institute for Evolutionary Biology, Plön, Germany

The highly specific association between three-spined sticklebacks (*Gasterosteus aculeatus*) and the tapeworm *Schistocephalus solidus* is an excellent model system for studies on host parasite interaction. In order to investigate the putative immune manipulation by *S. solidus* we used infection success of the eye fluke *Diplostomum pseudospathaceum* as a proxy for immune competence of the host fish. We used F1 progeny of *S. solidus* from a Norwegian (highly virulent) and a German (low virulence) population and their respective lab bred sympatric stickleback hosts in a fully reciprocal cross infection experiment. Six different fish families were exposed to four different worm sibships or sham exposed as controls. The fish were coinfected with *D. pseudospathaceum* at five different time points (1, 3, 6, 9, 12 weeks post exposure to *S. solidus*). In each infection round, a pool of cercariae from several field-collected snail hosts were used to account for the strong effects of *D. pseudospathaceum* genotype specificities. Fish from every treatment group (i.e. fish family x *S. solidus* sibship combination) were individually exposed to 100 *D. pseudospathaceum* cercariae. The infection success of *D. pseudospathaceum*, i.e. the number of metacercariae in the sticklebacks' eyes, was determined two days post exposure. We analyzed a total of 991 fish and found significant differences in their susceptibility to eye flukes between time points and between treatments within rounds, due to interaction effects of the origin of the tapeworm and the origin and sex of the fish. The main mechanisms of the infection phenotypes are now investigated by expression analysis of candidate genes of the innate and the adaptive immune system. Using this molecular approach, the underlying immune modulations of sympatric and allopatric host-parasite combinations can be studied in further detail.

587

The Mechanistic, Genetic, and Evolutionary Basis of Worker Sterility in the Social Hymenoptera

Isobel Ronai[1], **Vanina Vergoz**[1], **Benjamin P Oldroyd**[1]

1. The University of Sydney, Sydney, NSW, Australia

Extreme reproductive skew towards particular females is a defining feature of the social insects and workers are completely sterile in at least thirteen genera. The evolution of worker sterility is problematic because an individual that has decreased fertility has reduced direct fitness. In order to understand how worker sterility evolved it is essential to identify the mechanistic basis of worker sterility. We show that the developmental mechanisms that underlie worker sterility are 'reproductive control points' that reduce reproductive capacity in workers. We propose that environmental cues (nutritional and social) interact with particular signalling pathways in the worker and regulate worker fertility through reproductive control points both pre- and post-eclosion. We have identified eight gene signalling pathways that are likely to be involved in regulating worker fertility in honey bees: IIS; juvenile hormone; ecdysteroid; mTOR; dopamine; MAPK; Egfr and Wnt. We suggest that the common mechanism underlying all the reproductive control points is programmed cell death, an active process that causes the worker's reproductive organs to degenerate. An example of a reproductive control point is the abortion of oocytes during mid-oogenesis in the honey bee. The presence of the queen affects the expression of *Anarchy* in the ovaries of honey bee workers and during mid-oogenesis *Anarchy* transcripts localise to these degenerating oocytes. These reproductive control points are likely to have been involved in the ancestral emergence of worker sterility from a solitary insect.

588

Sex-biased dispersal in the short-tail stingray

Emily Roycroft[2, 1], **Agnes Le Port**[1, 4, 3], **Shane Lavery**[1, 4]

1. School of Biological Sciences, The University of Auckland, Auckland, New Zealand

2. The University of Melbourne, Melbourne, VICTORIA, Australia

3. Centre for Tropical Water and Aquatic Ecosystem Research (TropWATER), James Cook University, Townsville, QLD, Australia

4. Institute of Marine Science, Leigh Marine Laboratory, The University of Auckland, Auckland, New Zealand

The short-tail stingray (*Dasyatis brevicaudata*) is a large, coastal ray species with a disjunct range across the Southern Hemisphere. Recent work using mitochondrial DNA in this species from New Zealand, Australia and South Africa (Le Port and Lavery 2012) has indicated strong population structure in the maternal line. Given the mounting evidence for sex-biased dispersal in sharks, this finding suggests that female philopatry and male-biased dispersal may also occur in stingrays. In support of this, we present evidence for male-biased dispersal in the short-tail stingray and a nuclear DNA-based measure of population structure for the species around New Zealand and the Southern Hemisphere.

589

Evolutionary implications of gene duplication in plant secondary metabolism

Arunraj Saranya Prakashrao[1], Elisabeth Kaltenegger[1], Dietrich Ober[1]

1. Kiel University, Kiel, SCHLESWIG-HOLSTEIN, Germany

Pyrrolizidine alkaloids (PAs) are secondary metabolites found sporadically in many species of flowering plants. They are synthesized as a part of plant chemical defense against herbivores. The enzyme Homospermidine synthase (HSS) catalyzes the first step of PA biosynthesis. HSS is known to have originated from a gene duplication event from the ubiquitous eukaryotic enzyme deoxyhypusine synthase (DHS) involved in the activation of translation initiation factor. In our study we are focusing on the Convolvulaceae (Morning glory) family, wherein PAs occur only in a few isolated species. The recruitment of HSS to PA biosynthesis in these species has shown to be an efficient model to study the molecular evolution and functional divergence of enzymes after gene duplication. Our recent studies indicate a single duplication event in a common ancestor has led to the HSS genes in PA producing species as well as putative HSS copies in PA-free species. Hence, leading to interesting questions like, did the common ancestor of all PA-producing plants possess the pathway, followed by many independent losses? Or did the PA-biosynthesis evolve repeatedly in different lineages? We are using molecular phylogenetic analyses combined with biochemical characterization to study the factors that influenced the evolution of HSS. On the broader level we are trying to understand the role of gene duplication on the origin of new function and metabolism.

590

Genetic Drift in Variable Size Haploid and Diploid Populations

Helmut Simon[1]

1. Research School of Biology, Australian Nation University, Canberra, ACT, Australia

We derive a basic result for the loss of heterozygosity in a single generation of a haploid population and the limiting heterozygosity for the model after a large number of generations. The model used is based on a Galton-Watson branching process. For populations with a mean growth rate greater than one, heterozygosity does not approach zero, as would occur under the Wright-Fisher model [Kimura 1983]. We show that Sewall Wright's result [Wright 1938] on the effective population size of a diploid population with a range of variances in offspring number can be considered as a special case of a significantly more general result and use this as an example of the care that needs to be taken with the concept of effective population size. Our results provide some indication as to how and under what circumstances questions in coalescent theory for diploid populations can be reduced to problems in haploid coalescent theory. We then introduce mutation in the form of infinite-alleles and finite-alleles models and demonstrate that for populations with a mean growth rate greater than one, the concept of a balance between mutation and drift is not generally applicable during growth. Finally, we consider a limitation on the practical application of single-locus neutral population genetic models that does not appear to have been widely considered, namely the influence of selection on the wider genotype (cf. [Felsenstein 1974]).

1. Wright S 1938. Size of population and breeding structure in relation to evolution. Science 87, 430-431.
2. Kimura M 1983. The Neutral Theory of Molecular Evolution. Cambridge:Cambridge University Press.
3. Felsenstein J 1974. The evolutionary advantage of recombination. Genetics 78: 737–756.

591

Revisiting the plastid phylogenomics of Pinaceae with two complete plastomes of Pseudolarix and Tsuga

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

1. Biodiversity Program, Taiwan International Graduate Program, Taipei, Taiwan

2. Department of Life Science, National Taiwan Normal University, Taipei, Taiwan

3. Biodiversity Research Center, Academia Sinica, Taipei, Taiwan

Phylogeny of the ten Pinaceous genera has long been contentious. Plastid genomes (plastomes) provide an opportunity to resolve this problem because they contain rich evolutionary information. To comprehend the plastid phylogenomics of all ten Pinaceous genera, we sequenced the plastomes of two previously unavailable genera, *Pseudolarix amabilis* (122,234 bp) and *Tsuga chinensis* (120,859 bp). Both plastomes share similar gene repertoire and order. Here for the first time we report a unique insertion of tandem repeats in *accD* of *T. chinensis*. From the 65 plastid protein-coding genes common to all Pinaceous genera, we re-examined the phylogenetic relationship among all Pinaceous genera. Our phylogenetic analyses resulted in an identical tree topology, with the five genera of the Abietoideae subfamily constituting a monophyletic clade separate from the other three subfamilies: Pinoideae, Piceoideae, and Laricoideae. The five genera of Abietoideae were separated into two sister clades consisting of (1) *Cedrus* alone and (2) two sister subclades of *Pseudolarix-Tsuga* and *Abies-Keteleeria*, with the former uniquely losing the gene *psaM* and the latter specifically excluding the 3'*psbA* from the residual inverted repeat.

592

Evolution and genetic basis of *Acropora* species fluorescence

Shiho Takahashi-Kariyazono[1], Jun Gojobori[1], Yoko Satta[1], Yohey Terai[1]

1. SOKENDAI (The Graduate University for Advanced Studies), Hayama, KANAGAWA, Japan

Coral fluorescence is attributed to genome-encoded fluorescent proteins (FPs). This fluorescence is proposed to contribute to coral colors. However, genetic bases of coral color are not well understood. Narrow band laser excitation measurements of live *Acropora digitifera* revealed a clear separation of fluorescence from reflectance, and showed a wide range of fluorescent emission. *FP* cDNA sequences from *A. digitifera* and *A.*

tenuis revealed the presence of a multi-gene family with an unexpectedly large number of genes, separated into middle-wavelength emission (MWE), middle-/long-wavelength emission (M/LWE), and chromoprotein (CP) clades. *FP* gene copy numbers in the genomes of four *A. digitifera* colonies were estimated as 16–22 in the MWE clade, 3–6 in the M/LWE clade, and 8–12 in the CP clade. Fluorescent light produced by recombinant protein products encoded by the newly isolated genes explained the fluorescent range of live *A. digitifera*, suggesting that coral fluorescence is determined by a multi-*FP* gene family. The functionally diverse multi-*FP* gene family must have existed in the *Acropora* species ancestor, suggested by phylogenetic analyses and evolutionary.*Acropora* species have persisted multi-*FP* gene family during their evolution.

593

Eighteen non-visual opsins in the ancestor of vertebrates, astonishing duplication in ray-finned fish and loss in Amniota

Felix EG Beaudry, Tom Iwanicki, John S Taylor

There are eighteen subfamilies of non-visual opsins, when subfamily is defined as a clade with tetrapod and ray-finned fish orthologs. An amphibian sequence occurred in every subfamily, whereas Mammalia was represented in only seven. Fish species representing Holostei, Osteoglossomorpha, Otomorpha, Protacanthopterygii, Cyprinodontoidei, and Pleuronectiformes all had a large number of non-visual opsins (from 22 to 32 genes) as a result of ancient gene duplication events including, but not limited to, the teleost genome duplication (TGD). The non-visual opsin repertoire appears to have stabilized shortly after the TGD event and consequently these distantly related fishes had repertoires of similar size and composition. This differs from the pattern observed for visual opsins, where large fish repertoires were generated by relatively recent lineage-specific duplications. Most non-visual opsins have been named without the benefit of a phylogenetic perspective. We propose major revisions.

594

Dynamic cis-regulatory module recruitment during cardiac evolution

Michael Eichenlaub[1], Markus Tondl[1], Sara Alaei[1], Jose Polo[1], Mirana Ramialison[1]

1. ARMI, Monash University, Clayton, VIC, Australia

The heart evolved from a simple layer of contracting mesoderm in tunicates and contracting vessels in amphioxus to a more complex organ consisting of 2 chambers in teleosts, 3 in amphibians, 3.5 in reptiles, and 4 chambers in crocodiles, birds, and mammals. Although structural complexity has increased over 600 million years of evolution, astonishingly, the master control genes known to drive heart development are extremely well conserved. We hypothesise that non-coding variations trigger changes in the regulation of cardiac gene expression which contributes to distinct morphological phenotypes. To address this, we generated for the first time a genome-wide atlas of CRMs that are active in zebrafish embryonic cardiomyocytes. ChIP-seq experiments with antibodies directed against selected histone modifications were carried out which lead to an extensive description of the zebrafish cardiac-specific *cis*-regulatory landscape. CRM compositions were determined by analysing TFBS and motif occurrences to consolidate a vertebrate “cardiac CRM code”. By ways of comparative genomics, we then compared our zebrafish dataset to the available cardiac *cis*-regulatory repertoire in mouse, in order to investigate the contribution of regulatory elements to the establishment of morphological differences between the 2 speices. We identified only 8 ultra-conserved CRMs between vertebrate species that were conserved in sequence and in function, however the majority of cardiac-specific CRMs were species-specific, suggesting that cardiac evolution is driven by rapidly evolving *cis*-regulation.

595

The evolution of the innate immune system: insights from an early-divergent lineage

Chloe A van der Burg[1], Ana Pavasovic[1], Peter J Prentis[1]

1. Queensland University of Technology, Brisbane, QLD, Australia

Phylum Cnidaria (jellyfish, corals, hydroids and sea anemones), an early divergent eumetazoan lineage, is characterised by a lack of an adaptive immune system and hence must rely on the capacity of the innate immune system to cope with selection pressure associated with rapidly-evolving microbes. Interrogation of early divergent lineages may provide us with insights into the evolution and origin of eumetazoan gene sets. Therefore, cnidarians present an excellent avenue for research investigating both conserved and novel innate immune genes in eumetazoans. Here we report the first comprehensive and comparative study on the cnidarian innate immune system, across 13 transcriptomes from nine actiniarian species, to identify the origins, expansions and contractions of highly-conserved candidate gene families and novel innate immune genes. We characterised five candidate gene families; single copies of *TLR*, *MyD88* and *NF-κB* were found in most species, and several copies of *IL-1R* and *NLR* were found in all. A subset of NLRs possessed multiple transmembrane domains, which have so far only been identified in two cnidarian species, suggesting that membrane-bound NLRs may be a uniquely cnidarian innovation. Multiple novel immune genes were also identified with domain architectures including the Toll/interleukin-1 receptor (TIR) homology domain, which is well-documented as functioning in protein-protein interactions and signal transduction in immune pathways. We hypothesise that these genes may interact as novel proteins in immune pathways. Overall, these results provide an insight into the evolution of the innate immune system and show that cnidarians have a diverse repertoire of conserved and novel innate immune genes.

596

The relationship between coat colour phenotype and equine behaviour: a pilot study

Jessica L Finn[1], Bianca Waud (Haase)[1], Cali E Willet[1], Diane van Rooy[1], Tracy Chew[1], Claire M Wade[1], Natasha A Hamilton[1], Brandon D Velie[2]

1. Faculty of Veterinary Science, Sydney University, Sydney, NSW, Australia

2. Department of Animal Breeding & Genetics, Swedish University of Agricultural Sciences, Uppsala, UPPSALA, Sweden

Mutations in the genes influencing melanocytes not only affect the colour of an animal, but are also believed to impact physiological and behavioural functions. When this is taken into consideration, the common perception among horse owners that the chestnut coat colour is associated with adverse behaviours seems plausible. The aim of this study was to explore this perception by providing insight into any potential genetic associations between coat colour and adverse behaviours in horses. Data were acquired through an internationally accessible online

questionnaire. Respondents provided information on their horse's behaviour during general handling, whilst being exercised, towards different stimuli in their environment and when isolated from other horses. Analyses considered behavioural data on 477 horses that represented a range of breeds, ages, and event disciplines. The breed, sex, and age of the horse significantly ($P < 0.05$) influenced many of the equine behaviours assessed in the questionnaire. Significant differences in behavioural responses between bay and chestnut horses were only present for four questions. No evidence was found to support that chestnut horses are more likely than bay horses to display behaviours often associated with training difficulties. However, chestnut horses were more likely to approach objects and animals in their environment, regardless of their familiarity. This suggests that selection for the chestnut phenotype in horses may have inadvertently involved selection for boldness and altered the way horses interact with their surroundings.

598

Genetic and Protein Study of Alpaca Fibre.

Yen Ting Wong, Ching-Seng Ang, Beata Ujvari, Belinda Appleton

The Australian Alpaca industry is growing and the suri alpaca fleece is in high demand compared to huacaya fleece. By using genetic and protein approaches we investigated the alpaca genome to determine the order of alpaca's scaffolds and to narrow down the regions containing the genes responsible for the suri trait.

Genetic markers were used to map the six scaffolds in the region identified as containing the genes causing the suri trait. The locations of the causative mutations were narrowed to two regions- one on scaffold B (keratin region) at position 11,834 and one on scaffold A at position 2, 734, 593. Our analyses therefore suggest that two genomic regions are associated with the fleece variation in alpaca. Consequently, our results do not support to previous generally accepted single-locus genetic model where the suri trait has been proposed to be dominant but Presciuttini et al. (2010) who concluded that a two-locus gene model explains the genetic underpinnings of the suri fleece. To further investigate the regions, a protein approach was used to identify the difference between suri and huacaya fibre protein. Two keratin protein candidates were identified that showed high association with the suri trait. They were both located in the causative mutation region on scaffold B close to marker 11,834 confirming that the causative mutation of suri trait is located in the scaffold B region. Apart from being important for the alpaca industry, our results also contribute to deciphering the genetics of fleece phenotypes in other commercially important species.

599

Evolution of parent-of-origin effects on complex traits

Ge Zhang[1]

1. Cincinnati Children's Hospital Medical Center, Cincinnati, OHIO, United States

Parent-of-origin effects, in which the expression of phenotype effect of an allele depends on its parental origin, have been observed in many complex traits. Many theories have been proposed to explain the evolutionary origin of parent-of-origin effects. In this study, I examine the evolutionary advantage of parent-of-origin effects using a quantitative genetics model. I consider a quantitative trait under stabilizing selection is influenced by a large number of loci with various degree of unequal expression depending on their parent of origin. I evaluate the adaptive advantage of different parent-of-origin dependent expression patterns of these loci under various selection scenarios using forward-time population genetics simulation. The results show that when there is a positive parental effect, selection favors expression of alleles inherited from the opposite parent. If resemblance between mothers (or fathers) and offspring increases fitness, selection favors expression of alleles from the same parent. Parent-offspring conflict favors bilateral parent-of-origin expression of alleles, with some loci preferentially expressing the maternal alleles and some expressing the paternal alleles. If the strength of selection on maternal alleles is larger than paternal alleles, then better mean fitness is achieved when there are more loci with high expression of paternal alleles. In addition, these selection scenarios only favor strong parent-of-origin effects in a fraction of trait influencing loci. This study indicates that some complex traits loci might have acquired parent-of-origin dependent expression through different evolutionary pathways.

600

Are essential genes more likely to be conserved?

Fatemeh Ashari Ghomi[1], **Paul Gardner**[1], **Lars Barquist**[2]

1. University of Canterbury, Christchurch, New Zealand

2. Institute for Molecular Infection Biology, University of Wuerzburg, Wuerzburg, Germany

In this study, we investigate the relationship between the essentiality of genes and their conservation. Barquist et al. have studied the essentiality of genes in *Salmonella* Typhi and *Salmonella* Typhimurium and their results suggested that essential genes are not necessarily conserved and also conserved genes are not necessarily essential in these two serovars (Barquist et al., 2013).

We have studied the essentiality of genes in 12 strains from Enterobacteriaceae family using transposon-directed insertion-site sequencing (TraDIS). This method uses a large-scale transposon mutagenesis approach to generate a population of mutants. The organisms are then left to live in a rich media and after a few minutes, only the fit mutants survive. The positions of remaining insertions can be identified using next-generation sequencing. The genes that are free of insertions are likely to be essential for the survival of the organism. To investigate the conservation of genes, we have developed a protein clustering tool to identify homologous genes within these strains. We are analysing the genes that are conserved between these 12 strains and those that are not and the relationship between their conservation and essentiality.

1. Barquist, L., Langridge, G. C., Turner, D. J., Phan, M. D., Turner, A. K., Bateman, A., ... & Gardner, P. P. (2013). A comparison of dense transposon insertion libraries in the Salmonella serovars Typhi and Typhimurium. Nucleic acids research, gkt148.

602

Prevalence, evolution and concurrent infection by multiple dengue virus serotypes in patients from Pakistan

Qamar Bashir[1]

1. University of the Punjab, Lahore, Lahore, PUNJAB, Pakistan

During the last two decades, dengue fever has emerged as a serious health problem in Pakistan. We have analyzed the prevalence, the simultaneous occurrence of multiple serotypes in the patients, and phylogenetic analysis of the prevalent serotypes. The study comprised of blood samples of 18 dengue fever patients from Swat and 5 from Lahore. Using Nested Multiplex PCR, these samples were investigated for the dengue virus serotypes and occurrence of concurrent infection by more than one serotype. Domain III of the DENV envelope protein has been reported to be important for binding with the host cell receptors, and potentially a promising candidate for developing recombinant protein vaccine against dengue. We have amplified and sequenced domain III of the prevailing serotype for its phylogenetic analysis.

We have detected the presence of all four DENV serotypes in these samples, either existing solitary or present along with the others. Of the four DENV serotypes, DENV-2 was found to be the most prevalent, existing in 17 samples and DENV-1 was the least common detected only in two patients. We have identified concurrent infection with more than one serotype in 6 out of the 23 samples studied. Four samples showed the simultaneous presence of two serotypes with one sample having the three serotypes. Surprisingly we have observed the concurrent infection with all four serotypes in one sample, which has never been reported earlier. The phylogenetic analysis suggests that the prevalent serotype (DENV-2) circulating in Swat might have travelled from Lahore. It showed maximum homology with the one reported from India. Our study presents the first report of concurrent infection with DENV-1/3/4 and DENV-1/2/3/4 in a single patient. Thus all the four serotypes are circulating in Pakistan and might be taken into consideration in view of the dengue control.

603

Tracing functional protein interaction networks using 'feature-aware' phylogenetic profiling

Holger Bergmann[1], Julian Dosch[1], Vinh Tran[1], Ingo Ebersberger[1]

1. Goethe University Frankfurt am Main, Frankfurt Am Main, HESSEN, Germany

Tracing the phylogenetic distribution and, thus, the evolution of protein interaction networks across hundreds or even thousands of species calls for reliable and scalable methods for functional annotation transfer. Standard homolog or ortholog inferences resulting in so called 'phylogenetic profiles' do not suffice in many cases, as the functional similarity between evolutionary related sequences decays as a function of time. Here, we present HaMStR_OneSeq, a novel method that aids in the search for functionally equivalent proteins even over large evolutionary distances. The program integrates a targeted ortholog search with a subsequent assessment of the feature architecture similarity (FAS) between the proteins. Features comprise, among others, functional protein domains, secondary structure elements, transmembrane domains and low complexity regions. In detail, orthologs are identified in an iterative procedure starting from a single gene of interest - the 'seed protein'. Ortholog candidates are then weighted according to their pairwise FAS when compared to the 'seed protein'. In the cases of overlapping, redundant annotations in the architecture, we obtain the highest scoring linear paths through the graph using, where applicable, a greedy, and alternatively an exhaustive or a heuristic approach. The resulting score of an identified ortholog serves then as a proxy of its functional equivalence to the respective 'seed protein'. A dynamic visualization tool enables the user to visualize and explore the resulting 'feature-aware' phylogenetic profile.

To demonstrate the application of HaMStR_OneSeq, we traced the DNA uptake machineries of five naturally competent bacteria in more than 1,000 species. The aim was to shed light on the distribution and evolution of natural competence in the bacterial domain. The prediction of hitherto unknown naturally competent bacteria with high confidence indicated that the capability of direct DNA uptake is far more common among bacteria than acknowledged to date.

604

Phylogenomics analysis of large bacterial phylogenies using whole genome information

Minh Duc Cao[1], Lachlan Coin[1]

1. The University of Queensland, Brisbane, QUEENSLAND, Australia

Advances in high-throughput sequencing have dramatically transformed microbial research by allowing sequencing whole genomes of thousands of microorganisms in a single study. The evolutionary analysis of these microorganisms is generally a crucial step in any many analyses. Classical phylogenetic analyses based on single genes or parts of the genomes are often inconsistent because of factors such as variable rates of evolution and horizontal gene transfers. Recent phylogenetic studies focus on establishing the phylogeny from core SNPs and hence are limited to the analysis of closely related organisms. Here we present an information theoretic method to estimate the genetic distance between two bacterial organisms using the mutual information of their genomic sequences. The method employs an adaptive local alignment model to identify homologous regions and quantify all the variation into a unified information unit. As a consequence, our method can account for a range of diversity among the taxa as well as different types of variation such as SNPs, indels and rearrangements. We demonstrate the robustness of our method by building a phylogeny of 2000 bacterial organisms in 12 pathogen bacterial species from their draft genome assemblies. We found that all taxa from the same species were correctly grouped together and the placements of these species were in complete agreement with the current bacterial taxonomy. The subtree for each species was also largely congruent with the tree built from the multi-locus strain typing profiles of the species.

605

Can adaptive radiations in the Bemisia tabaci species complex be revealed by co-evolving endosymbionts?

Samia Elfekih[1], Wee Tek Tay[1], Karl Gordon[1], Paul De Barro[2]

1. CSIRO, Canberra, ACT, Australia

2. CSIRO, Brisbane, QLD, Australia

The whitefly Bemisia tabaci is a complex of morphologically indistinguishable species and is one of the most destructive insect pests worldwide. This pest is multivoltine, highly polyphagous and a vector to important plant viruses including the cassava brown streak virus and the cassava mosaic virus. B. tabaci cryptic species are known to harbor various bacterial endosymbionts, some of which are know to confer fitness to the host. Here, we conducted genome-wide single nucleotide polymorphism (SNP) analysis using a Nextera-tagmented, reductively amplified DNA protocol (nextRAD) to investigate the phylogenomics of the Bemisia cryptic invasive species complex and study their endosymbiont metacommunities. We ascertained levels of endobacterial abundances in cryptic host species, identified reads mapping to primary (i.e. Candidatus Portiera aleyrodidarum) and secondary endosymbionts (e.g., *C. Hamiltonella defensa*), and explored patterns of host-endosymbiont co-evolutionary relationships from these diverse classes of primary- and secondary-endosymbionts through the endosymbionts' genome-wide SNP data. Our findings will have implications for further understanding the effect of bacterial endosymbionts on the host evolutionary ecology.

606

Avoidance of stochastic RNA interactions can be harnessed to control protein expression levels in bacteria and archaea

Paul Gardner[1], **Sinan U Umu**[1], **Anthony Poole**[1], **Renwick Dobson**[1]

1. University of Canterbury, Christchurch, New Zealand

A critical assumption of gene expression analysis is that mRNA abundances broadly correlate with protein abundance, but these two are often imperfectly correlated. Some of the discrepancy can be accounted for by two important mRNA features: codon usage and mRNA secondary structure. We present a new global factor, called mRNA:ncRNA avoidance, and provide evidence that avoidance increases translational efficiency. We also demonstrate a strong selection for avoidance of stochastic mRNA:ncRNA interactions across prokaryotes, and that these have a greater impact on protein abundance than mRNA structure or codon usage. By generating synonymously variant green fluorescent protein (GFP) mRNAs with different potential for mRNA:ncRNA interactions, we demonstrate that GFP levels correlate well with interaction avoidance. Therefore, taking stochastic mRNA:ncRNA interactions into account enables precise modulation of protein abundance.

A preprint describing this work can be found here: http://biorxiv.org/content/early/2016/03/23/033613

607

Evolutionary Genomics of Plant Pathogenic Enterobacteria

Jeremy Glasner[1], **Nicole Perna**[1]

1. University of Wisconsin, Madison, Wisconsin, United States

Phytopathogenic enterobacteria cause economically significant losses of ornamental and crop plants worldwide. Abundant genome sequences for the family as a whole provide rich data to investigate the evolutionary events associated with specialization in animal and plant hosts. Genes acquired by horizontal transfer in the ancestral lineages leading to phytopathogens are high priority targets for experimental characterization as potential contributors to plant-microbe interactions.

Ortholog groups of for all proteins from 307 genomes of enterobacteria were predicted with OrthMCL and used for species tree reconstructions using RAxML, ASTRAL, PhyloNet and SNaQ. Ancestral state reconstructions (ML, parsimony and ad hoc) were used to transform ortholog presence/absence data to predictions of gene gains and losses in the most recent common ancestral lineages of the phytopathogen clades. We integrated these analyses with gene expression profiles, protein annotations, and a database of genes involved with host-microbe interactions.

This study reveals a dominant phylogeny with two distinct clades of phytopathogens. The ancestral lineages leading to the soft rot-associated and Yersinia clades appear to have exchanged substantial numbers of homologous genes, providing evidence that horizontal gene transfer is not randomly distributed among taxa throughout the tree. Focused exchange between lineages can be a major barrier to robust species tree reconstruction and may explain inconsistencies observed in previous phylogenetic analyses. Inferences of gene gains/losses identified hundreds or thousands of genes acquired in the most recent common ancestral lineages leading to the soft rot clade. Integrated analyses of gene flux and functional genomics data prioritizes candidates for further experimentation and suggest that expansion of chemotaxis receptor and ABC transporter protein families through both horizontal transfer and duplication preceded diversification of the soft rot clade.

Despite a clearly reticulate evolutionary history, reconstruction of gene losses and gains in a relatively old lineage provides insight into specialization in plant hosts.

608

Testing the role of recombination in the evolution of multidrug resistance in experimental bacterial populations

Chowdhury Mehedi Hasan[1], **Hong Nga Nguyen Vu**[1], **Beth Brittain**[1], **Pål J Johnsen**[2], **Daniel Ortiz-Barrientos**[1], **Jan Engelstaedter**[1]

1. School of Biological Sciences, University of Queensland, Brisbane, Queensland, Australia

2. Faculty of Health Sciences, Department of Pharmacy, UIT - The Arctic University of Norway, Tromsø, Norway

The emergence of multidrug resistant bacteria has become a major cause of therapeutic failure in treating infectious diseases. Multidrug resistance is frequently acquired by horizontal gene transfers, but can also arise *de novo* through mutations. In the latter case, recombination may be important in reducing clonal interference between selected resistance mutations that spread simultaneously within the population. Many bacteria, including important pathogens, regularly undergo recombination via natural transformation (uptake of DNA from the environment), but the role of natural transformation in the evolution of *de novo* multidrug resistance evolution is unclear. Our study aims at characterizing the evolutionary dynamics of *de novo* multidrug resistance through evolution experiments in which the emergence and spread of resistance mutations is monitored and the impact of recombination assessed. We initiated our evolution experiment with populations comprising either naturally competent or non-competent genotypes of *Acinetobacter baylyi*. These populations were then propagated by serial transfer for ~650 generations under sub-lethal doses of rifampicin and streptomycin antibiotic combinations. We then characterized evolved populations by phenotypic assays and whole genome sequencing. Both growth rate and competition assays demonstrated that the populations propagated under drug pressure had evolved higher fitness when tested in same environment, but there was no difference in fitness gain between competent and non-competent populations. Moreover, antibiotic susceptibility assays showed that clones that evolved in presence of drugs had become strongly resistant to rifampicin, whereas resistance to streptomycin was weaker or absent. Consistent with these findings, whole genome sequencing revealed an abundance of different rpoB mutations (indicating target alteration as a resistance mechanism). We have also identified other mutations previously reported to be associated with resistance to other antibiotics. In conclusion, we saw no evidence that recombination by transformation facilitates adaptation to antibiotics, presumably because the low number of mutations that were spreading prevented clonal interference.

609

Geographical distribution influences purifying selection on innate immune genes in *Drosophila*

Rosemarie Herbert[1], **Elizabeth McGraw**[1]

1. Biological Sciences, Monash University, Clayton, Victoria, Australia

There is a wealth of genome material available for the model organism, Drosophila melanogaster. Work in the area has elucidated a range of immune genes under positive selection. However, a broader question is whether there are interactions between *Drosophila* genetics and the environment. Using PAML, we found that purifying selection works most strongly on *Drosophila* spp. found primarily in a narrow geographical distribution, namely in tropical species. This work suggests that there may be more interesting interactions to be plumbed from other data, particularly in vector datasets. If similar trends are seen in virus vectors, this has implications for the spread of disease and vector ability to respond to pathogens as the climate warms.

610

The origin and evolution of the emergent plant pathogen *Pseudomonas syringae* pv. *actinidiae*

Honour McCann[1], Li Li[2], Yifei Liu[3], Elena Colombi[1], Matthew Templeton[4], Paul Rainey[1, 5, 6], Hongwen Huang[3, 2]

1. Massey University, Auckland, NORTH ISLAND, New Zealand

2. Wuhan Botanical Garden, Chinese Academy of Sciences, Wuhan, China

3. South China Botanical Institute, Chinese Academy of Sciences, Guangzhou, China

4. Plant and Food Research, Auckland, New Zealand

5. Max Planck Institute for Evolutionary Biology, Plon, Germany

6. Ecole Superieure de Physique et de Chimie Industrielles, Paris, France

A new lineage of the bacterial plant pathogen *Pseudomonas syringae* pv. *actinidiae* (*Psa*) recently emerged to cause a pandemic of bleeding canker disease on kiwifruit (*Actinidia chinenis*). Evidence of recombination between the pandemic lineage and strains isolated during earlier outbreaks led us to predict the existence of a genetically diverse and recombining source population of *Psa*, from which stochastic sampling events followed by selection in agricultural environments for host specialization occurred. We hypothesized that China was the likeliest origin of the pathogen source population, as it is the origin of *A. chinensis* and the center of kiwifruit abundance and diversity. In order to identify the location of and extent of diversity within the source population of *Psa*, pinpoint the origin of the recent pandemic, and investigate the evolutionary processes leading to its emergence, a phylogenomics study was initiated to sample *P. syringae* from kiwifruit across six provinces in China as well as Japan and Korea. To date, we have found that all Chinese *Psa* are members of the same lineage as the pandemic isolates. Our work reveals there is far greater diversity within this lineage than was previously known, indicating *Psa* was circulating in China before the global outbreak began. While the sublineage responsible for the latest pandemic emerged in China, Japan and Korea harbour strains from multiple lineages of *Psa*, suggesting the center of origin of the pathogen is outside the center of host plant diversity.

611

Ribosomal footprinting allows detecting and analyzing weakly translated, evolutionary young genes in EHEC

Klaus Neuhaus[1], Sarah Hücker[1], Richard Landstorfer[1], Benjamin Ruprecht[2], Harald Marx[2], Bernhard Kuster[2], Svenja Simon[3], Daniel A Keim[3], Siegfried Scherer[1]

1. Lehrstuhl für Mikrobielle Ökologie, Technische Universität München, Freising

2. Lehrstuhl für Proteomik und Bioanalytik, Technische Universität München, Freising

3. Lehrstuhl für Datenanalyse und Visualisierung, Universität Konstanz, Konstanz

Genomes of *Escherichia coli*, including that of the pathogen enterohemorrhagic *E. coli* O157:H7 (EHEC), still harbor undetected protein-coding genes which, apparently, have escaped annotation due to their small size and non-essential function. However such genes might be important for adaptation and evolution. Mass spectrometry (MS) based proteomics is a sensitive high-throughput method that directly measures the presence of polypeptides. An emerging technique to determine translational activity is ribosomal footprinting (Ribo-seq), measuring the ribosome coverage of mRNA, thus constituting the translatome.

A quantitative comparison of 52 well MS-measurable EHEC proteins with Ribo-seq translatome data resulted in a positive Pearson correlation (r_P = 0.84). But the global correlation between MS and Ribo-seq data of all proteins drops substantially. This could be attributed to proteins with a bias in MS detection, caused by “non-standard” protein parameters, e.g., low abundancy, missing or too many tryptic cleavage sites, or hydrophobicity, and differences in protein half-life. Almost all (98%) of the proteins detected by MS display a strong translatome signal, but many proteins not detected by MS showed a significant Ribo-seq signal for their mRNA as well. A number of not-annotated genes were found, some of which are annotated in other enterobacteriaceae. However, several genes constitute novel discoveries. In addition, protein structure and function were predicted computationally and compared between EHEC-encoded proteins and 100-times randomly shuffled proteins. Based on this comparison, about 85% novel proteins exhibit predicted structural and functional features similar to those of annotated proteins.

These findings demonstrate that ribosomal footprinting can be used to detect novel protein coding genes, contributing to the growing body of evidence that such weak genes are not mere artifacts. Ribo-seq opens an additional way to detect and analyze novel genes, since most were taxonomically restricted and, therefore, appear to have evolved relatively recently *de novo*.

612

The basis of the Bartonella radiation: Origin and evolution of a specialized gene transfer agent

Daniel Tamarit[1], Minna Neuvonen[1], Philipp Engel[2], Lionel Guy[3], Siv Andersson[1]

1. Molecular Evolution, Uppsala University, Uppsala, Sweden

2. Fundamental Microbiology, University of Lausanne, Lausanne, Switzerland

3. Medical biochemistry and microbiology, Uppsala University, Uppsala, Sweden

Gene transfer agents (GTAs) are domesticated viral particles that mediate horizontal transfer of small random segments of chromosomal DNA between prokaryotic cells. Recently, it was discovered that pathogens of the genus *Bartonella* contain a conserved phage-derived gene cluster for a novel GTA, which is physically separated from another phage-derived cluster that replicates the surrounding genomic regions in a process called run-off replication. Together, these two loci became a key innovation facilitating the adaptive radiation of *Bartonella* by efficiently transferring adaptive genes between cells. How GTAs evolve from ancestral phages has remained elusive. To study the origin and evolution of the *Bartonella* GTA, we sequenced the closest known relatives to the *Bartonella* group, symbionts of ants and bees, and performed comparative and evolutionary analyses with *Bartonella*genomes. We discovered that the GTA genes co-diversify with the host genomes, in contrast to other

GTA-like lineages, which evolve like phages capable of horizontal descent. Interestingly, we found that the GTA genes were already present in the *Bartonella* last common ancestor, while the gene cluster causing run-off replication was acquired later. Therefore, although transfer of random DNA fragments probably took place early on, it is plausible that the targeted sharing of a genomic island enriched in genes for host association caused the *Bartonella* radiation, leading to the infection of a wide range of mammalian hosts. With this work we propose a model for the gradual transformation of a prophage into a specialised GTA in a process driven by selection from the host bacterial genome.

613

Identification of source and sink populations for the emergence and global spread of the East-Asia clone of Community-Associated MRSA

Melissa Ward[2, 1], Mariya Goncheva[3], Emily Richardson[3], Paul McAdam[3], Emma Raftis[3], Angela Kearns[4], Robert Daum[5], Michael David[5], Tsai Ling Lauderdale[6], Kirsty Girvan[7], Giles Edwards[7], Graeme Nimmo[8], Xander Huijsdens[9], Mark Woolhouse[1], Ross Fitzgerald[3]

1. Centre for Immunity, Infection and Evolution, University of Edinburgh, Edinburgh, UK

2. University of Oxford, Oxford, OXFORDSHIRE, United Kingdom

3. The Roslin Institute, University of Edinburgh, Edinburgh, United Kingdom

4. Public Health England, Colindale, United Kingdom

5. University of Chicago, Chicago, Illinois, USA

6. National Health Research Institutes, Taiwan

7. Scottish MRSA Reference Laboratory, NHS Greater Glasgow and Clyde, Glasgow, United Kingdom

8. Griffith University School of Medicine, Gold Coast, Queensland, Australia

9. National Institute for Public Health and the Environment, Bilthoven, The Netherlands

Background

Our understanding of the factors influencing the emergence, dissemination and global distribution of epidemic clones of bacteria is very limited. ST59 is a major epidemic clone of community-associated MRSA in East Asia, responsible for extensive morbidity and mortality, but has a much lower prevalence in other parts of the world. The geographic origin of the ST59 clone and its international routes of dissemination are unclear and disputed in the literature.

Results

To investigate the origin and spread of the ST59 clone, we obtained whole genome sequences of isolates from four continents, sampled over more than a decade, and carried out a time-scaled phylogeographic analysis. We discovered that two distinct ST59 clades emerged concurrently, in Taiwan and the USA, but underwent clonal expansion at different times. The Taiwan clade was strongly enriched for gene determinants associated with antibiotic resistance, consistent with regional differences in antibiotic usage. Both clones spread independently to Australia and Europe and we found evidence of the persistence of multi-drug resistance following export from Taiwan. Direct transfer of strains between Taiwan and the USA was not observed in either direction, consistent with geographic niche exclusion. Unexpectedly, *in vitro* competitive fitness experiments revealed that ST59 strains from the Taiwan and USA clades were able to out-compete USA300, the dominant community-associated MRSA strain in the USA.

Conclusions

Our results resolve a longstanding controversy regarding the origin of the ST59 clone, revealing the major global source and sink populations and routes for the spread of multi-drug resistant clones. In addition our findings indicate that the diversification of the accessory genome of epidemic clones partly reflects region-specific patterns of antibiotic usage, which may influence bacterial fitness after transmission to different geographic locations.

614

Epidemic success and underlying driving forces of the *Mycobacterium tuberculosis* complex in a low-prevalence setting

JEAN-PHILIPPE RASIGADE[4, 3, 1, 2], MAXIME BARBIER[1, 2], OANA DUMITRESCU[4, 3], JEAN-PIERRE FLANDROIS[4], GERARD LINA[4, 3], THIERRY WIRTH[1, 2]

1. Institut de Systématique, Evolution, Biodiversité, UMR-CNRS 7205, Muséum National d'Histoire Naturelle, Université Pierre et Marie Curie, Ecole Pratique des Hautes Etudes, Sorbonne Universités, Paris, FRANCE

2. Laboratoire Biologie Intégrative des Populations, Ecole Pratique des Hautes Etudes, PSL Research University, Paris, FRANCE

3. CIRI, Centre International de Recherche en Infectiologie, Lyon, Région Lyonnaise, FRANCE

4. Dpt. of Microbiology, Hospices Civils de Lyon, Lyon, Région Lyonnaise, FRANCE

Tuberculosis caused by the *Mycobacterium tuberculosis* complex (MTBC) is a worlwide emergency. A better understanding of the epidemic success of MTBC in various settings is needed to inform control strategies. To this aim, we introduce timescaled haplotypic density (THD), a novel genotype analysis method based on kernel density estimation to derive correlates of epidemic success, endemicity and pathogen transmission between groups of patients over specific timescales. Using mycobacterial tandem repeat sequences as genetic markers, we investigated a retrospective multicentric cohort of 1,641 MTBC-infected patients from France, a low-prevalence country where MTBC cases are frequently imported from abroad. THDs with timescales of 20 and 200y were included in association analyses with pathogen and patient characteristics. From the pathogen standpoint, our results identify the ability to cause pulmonary (hence, transmissible) disease as a major driving force of long-term epidemic success, most notably in the Euro-American and Beijing MTBC lineages. THD discriminated isolates of the regional endemic background from those imported more recently, allowing to identify several socio-demographic factors, such as younger age and student status, independently associated with non-endemic MTBC infection. We also decipher how past contacts between French and foreign populations might have contributed to shape the population structure of MTBC strains currently circulating in French-native patients. Our results highlight a combined influence of contacts with Europe, Northern and Middle Africa over a 200y timescale, and a preeminent influence of contacts with Northern Africa over the more recent 20y timescale, in line with historical and epidemiological evidence. To conclude, we describe how the interplay of MTBC lineage specificities, host risk factors and past human migrations contribute to the large-scale population dynamics of MTBC in a low-prevalence setting. Our approach could be applied to other pathogens, allowing to leverage the increasing wealth of genotypic and clinical data available from infection surveillance databases.

615

Co-evolution of Hepatitis B Virus subgenotype C4 and Indigenous Australians for at least 53,000 years

Lilly Yuen[1], Margaret Littlejohn[1], Jane Davies[2, 3], Sarah Bukulatjpi [2, 4], Paula Binks[2], Kathy Jackson[1], Josh Davis[3, 5], Steven Tong[2, 3], Stephen Locarnini[1]

1. Victorian Infectious Diseases Reference Laboratory, Doherty Institute, Melbourne, VIC, Australia

2. Menzies School of Health Research and Charles Darwin University, Darwin, NT, Australia

3. Department of Infectious Diseases, Royal Darwin Hospital, Darwin, NT, Australia

4. Ngalkanbuy Clinic, Galiwin'ku, NT, Australia

5. John Hunter Hospital, Newcastle, NSW, Australia

Hepatitis B is an ancient human infection characterised by slow disease progression. It has been proposed HBV evolved with modern humans following migration out of Africa and diverged into 10 genotypes (A-J). Indigenous Australians have the oldest continuous living culture outside of Africa. However, entry time and likely access point into Sahul remain controversial. We have shown the strain of HBV infecting Aboriginal people in Northern Territory of Australia is a novel subgenotype HBV/C4. These people are located over vast distances with no known epidemiological connections. We hypothesized HBV/C4 entered Sahul with these First Australians. We obtained genome sequences of 59 HBV/C4 and 216 other publicly available HBV sequences for recombination, time divergence and ancestral state reconstruction analyses, and found HBV/C4 is a recombinant virus, predominantly genotype C (80%) with a genotype J surface (S) gene (20%). Phylogenies showed the HBV/C component clustered with other human HBV/Cs, while the S-gene of HBV/C4 and HBV/J clustered with South-east Asian (SEA) (Sunda) non-human primate HBVs. Time to most recent common ancestor (tMRCA) for HBV/C4 and HBV/J, potentially when the C-J recombination event took place, was inferred ~71K years. tMRCA of all HBV/C4 was ~53K, indicating this unique strain has been in Australia for this long, overlapping the estimated arrival date of Indigenous Australians based on archaeological evidence. Ancestral state reconstruction analysis inferred Daly River was where HBV/C4 originally entered Sahul. Given the Wallace Line separates the fauna between Asia and Australia ecozones, the recombination of HBV/C and HBV/J to generate HBV/C4 would most likely has to have occurred on Sunda. Based on the data obtained, we propose HBV/C4 originated on Sunda ~71kya, has been in Australia for at least 53K years, and likely was brought in by the ancestors of Indigenous Australians entering via the Daly River region.

616

Overexpression of overlapping ORFs in *Escherichia coli* O157:H7 reveals growth phenotype

Barbara Zehentner[1], Richard Landstorfer[1], Svenja Simon[2], Daniel A. Keim[2], Siegfried Scherer[1], Klaus Neuhaus[1]

1. Technical University Munich, Freising, GERMANY, Germany

2. Lehrstuhl für Datenanalyse und Visualisierung, Fachbereich Informatik und Informationswissenschaft, Universität Konstanz, Konstanz

The existence of overlapping genes is commonly accepted for viruses but not generally in other organisms. Bacterial genomes seem to contain a number of overlapping open reading frames (ORFs) which could be a coding reserve for future adaptation. Antiparallel ORFs which show transcriptional and translational signals were found in the genome of *Escherichia coli* O157:H7 (EHEC). However, the potential function of such ORFs remains unknown.

A high-throughput phenotyping method was developed to investigate the effect of overlapping ORFs on bacterial growth. Therefore, the candidate ORFs were cloned and simultaneously overexpressed in 20 different growth conditions. To detect phenotypes within the different growth conditions, the fraction of each clone before and after growth was analyzed using next-generation sequencing. For each overlapping ORF tested, a 'phenotypic signature' across all stress conditions was determined. In two biological repeats several of the tested candidates showed clear growth disadvantages or advantages. At least two candidates were found to improve growth under acidic conditions. Single competitive assays verified this phenotype. However, most overlapping ORFs do not influence the bacterial growth significantly under the conditions tested. Thus, these sequences might be useful in specific habitats only, despite their formerly detectable expression, which would corroborate the hypothesis that such genes may form a coding reserve potentially useful under stress conditions.

617

Re-evaluating the target of selection within *FOXP2* suggests functional divergence among diverse human populations

Elizabeth G Atkinson[1], Gregory T Smith[1], Brenna M Henn[1]

1. Stony Brook University, Stony Brook, NEW YORK, United States

Identified for its critical role in the development of human speech, *FOXP2* is a canonical language gene showing a Mendelian pattern of inheritance. Earlier research proposed a recent selective sweep in *Homo sapiens* targeting two derived amino-acid changes. However, these substitutions were also found in ancient hominin DNA suggesting that the selective sweep was not unique to modern humans. Here, we comprehensively re-analyze *FOXP2* with a high-resolution dataset comprising hundreds of next-generation sequenced genomes from globally distributed human populations. We test for fine-scale selection patterns both within the gene and between various human populations in order to resolve a hypothesis of recent positive selection. Intriguingly, haplotype networks and window-based Tajima's *D* calculations indicate balancing selection in African populations. Specifically, we identify*in silico* three major, common *FOXP2* haplotypes segregating in modern humans: two within Africa, and one fixed in individuals whose ancestors underwent the Out-of-Africa migration. These three haplotypes span a narrow intronic region that is significantly (P<0.001) evolutionarily conserved across vertebrates based on PhyloP and Genomic Evolutionary Rate Profiling (GERP) scores. Additionally, this region harbors high-GERP SNPs that are derived in humans compared with chimps and archaic hominins. This region of interest is a statistically significant GERP outlier compared to comparably sized genic windows in the same datasets. Strong evolutionary constraint amongst taxa but variability within *Homo sapiens* is compatible with this locus having a major functional role unique to humans. We examine the effect that these haplotypes might have on the FOXP2 transcription factor's function (i.e. production of alternative coding isoforms and/or differential expression of target genes) using data from the CommonMind Consortium, which contains matched SNP array and RNA-seq data obtained from fresh cortical brain tissue from hundreds of ethnically diverse individuals.

618

Reconstructing human diffusion and collapse in the Americas

Chiara Barbieri[1], Silvia Ghirotto[2], Leonardo Arias[3], José Sandoval[4], Federica Sevini[5], Sara De Fanti[6], Zelda Franceschi[7], Claudio Franceschi[5], Donata Luiselli[6], Mark Stoneking[3], Ricardo Fujita[4], Paul Heggarty[1]

1. Department of Linguistic and Cultural Evolution, Max Planck Institute for the Science of Human History, Jena, Germany

2. Department of Life Sciences and Biotechnologies, University of Ferrara, Ferrara, Italy

3. Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany

4. Centro de Genética y Biología Molecular, Universidad San Martin de Porres (USMP), Lima, Peru

5. C.I.G. Interdepartmental Centre L. Galvani for Integrated Studies on Bioinformatics, Biophysics and Biocomplexity, University of Bologna, Bologna, Italy

6. Department of Biological, Geological and Environmental Sciences, University of Bologna, Bologna, Italy

7. Dipartimento di Storia Culture Civiltà, University of Bologna, Bologna, Italy

The Americas, with their relatively recent history of population diffusion and collapse after European contact, represent an ideal case-study for testing the genetic footprints left by demographic changes. Native American mtDNA genomes have been successfully employed to reconstruct the timing and magnitude of the initial population expansion into the continent. However, previous studies were based on collections of individual lineages and were missing actual population samples. This limited perspective lacked the resolution to explore regional population spread and diversification. Furthermore, the recent collapse was not adequately tested in populations with different prehistories and in different regions.

In this study we focus on Meso- and South America to explore 1) the diversification of populations who crossed the Isthmus of Panama into South America and 2) the traces of a recent demographic collapse in various sets of population mitogenomes. Our dataset comprises 320 full mtDNA genomes from 13 populations from Mesoamerica and from different ecogeographic regions of South America: the Andes, Amazonia and the Gran Chaco. A Bayesian approach is employed to reconstruct population demographies and to test different scenarios of expansion and collapse. Our results suggest different prehistories for the populations studied, irrespective of their geographic location. Similarities between Mesoamerican and Andean populations indicate a possible connection on the Pacific coast, which is tested with spatially-informed simulations. Coalescent simulations support the recent collapse in most of our populations, helping us to understand the impact of recent events on population mitogenomes.

Populations in the Americas experienced different demographic trajectories and strong diversification, which possibly drove the high cultural and linguistic diversity reported today. Including mtDNA genomes from modern populations along with ancient samples promises to shed light into the past of the Americas and into the demographic dynamics behind population divergence.

619

Human population genetics of Papua New Guinea

Anders Bergstrom[1], Stephen J Oppenheimer[2], Alexander J Mentzer[3], Kathryn Auckland[3], George Koki[4], William Pomat[4], Peter Siba[4], Yali Xue[1], Manjinder S Sandhu[1], Chris Tyler-Smith[1]

1. Wellcome Trust Sanger Institute, Cambridge, United Kingdom

2. School of Anthropology and Museum Ethnography, Oxford University, Oxford, United Kingdom

3. Welcome Trust Centre for Human Genetics, Oxford University, Oxford, United Kingdom

4. Papua New Guinea Institute of Medical Research, Goroka, Papua New Guinea

Human presence in Sahul, the ancient continent encompassing Australia, Tasmania and New Guinea, dates back to ~50 kya. Papua New Guinea has a varied, mountainous geography and has attracted long-standing anthropological interest because of its great human cultural and linguistic diversity. Genetic studies so far have however been small and mostly limited to a few markers and/or the uniparental chromosomes. We have genotyped 382 individuals from Papua New Guinea on the 1.7 million sites of the Infinium Multi-Ethnic Global array. Our sample includes individuals from 18 of the 22 provinces and covers dozens of geographically and linguistically distinct groups within the highlands and the northern and southern lowlands. We find that most groups have remained unaffected by external gene flow into the region since the initial settlement of Sahul. We find strong population structure and high genetic differentiation – many F_{ST} values between groups within the highlands are larger than those between major populations within the continents of Europe or East Asia. Using relative cross coalescence rates between high-coverage genome sequences from selected populations we estimate that this differentiation has formed mostly within the last 10 ky. This study provides the first comprehensive overview of human population history and structure in this historically and anthropologically important part of the world.

621

Characterizing introgressed sequences in *S. cerevisiae*

Anne Clark[1], Josh Akey[1]

1. University of Washington, Seattle, WASHINGTON, United States

Hybrid yeast are of interest both in basic research and in economically important fermentation processes. Little is known, however, about how often hybridization occurs in wild yeast populations, or what role the sequences transferred between species through these hybridization events play in evolutionary adaptation. With advances in sequencing technology, a much greater diversity of budding yeast genomes are being sequenced, allowing us to look more systematically at introgressed sequences. We have developed a hidden Markov model framework for identifying regions of *S. cerevisiae*genomes that are likely to have been introgressed from *S. paradoxus* or more distantly related species. We establish the sensitivity and specificity of the approach through simulations of hybridization under a variety of demographic models. We will present results that demonstrate our approach is effective under certain assumptions about the evolutionary history of budding yeast. We will also discuss findings from applying the framework to a set of 100 diverse *S. cerevisiae* whole genome sequences, highlighting specific introgressed genes that are of interest for further functional characterization. Finally, we will demonstrate how some aspects of the evolutionary history of budding yeast may be inferred from the distribution of predicted introgressed sequence lengths. This work provides insight into the evolutionary role of hybridization in wild yeast, as well as specific examples of sequences transferred through hybridization.

622

Supergene evolution favoured by the introgression of an inversion in *Heliconius*

Angeles de Cara[1], Paul Jay[2], Annabel Whibley[3], Mathieu Joron[2]

1. Lab d'Ecoanthropologie et Ethnobiologie, UMR 7206, Museum National d'Histoire Naturelle, CNRS, Paris, ILE DE FRANCE, France

2. Centre d'Ecologie Fonctionnelle et Evolutive, CNRS, Montpellier, France

3. John Innes Centre, Norwich, United Kingdom

Heliconius butterflies are a famous example of adaptive radiation driven by natural selection on wing-pattern mimicry. Among them is *Heliconius numata*, which shows a spectacular polymorphism with multiple coexisting forms mimicking distinct local butterfly species and their geographic variations. Wing-pattern polymorphism is controlled by a group of tightly-linked genetic elements, or supergene, maintained in linkage disequilibrium by polymorphic chromosomal rearrangements. Each arrangement is characterised by one or several inversions, associated with a given phenotypic form. Here we investigate the origin of inversion polymorphism and the role of introgression in the formation of the supergene. Breakpoint genotyping in related species shows that the main inversion is shared with a non-sister species, *H. pardalinus,* suggesting introgression may explain the emergence of a new allelic class in *H. numata*. Based on whole genome resequence data, f and D statistics (ABBA-BABA analyses) reveal an excess of shared derived mutations in the inversion between *H. numata* and *H. pardalinus*. Treemix analyses indicate a history of gene flow between the two taxa, and topological changes in the phylogeny across the genome show a pattern of haplotype sharing consistent with an ancient introgression of the inversion into *H. numata*. Multiple sequentially Markovian *coalescent* (MSMC) analyses corroborate this ancient origin, while the population branching statistic (PBS) and haplotype similarity confirm the common origin of the inversion. Finally, comparative demographic inferences and Approximate Bayesian Computation (ABC) simulations lead to the hypothesis that supergene formation and polymorphism could be associated with an increase in effective population size and gene flow in *H. numata*. We conclude that the introgression of an inversion kick-started the evolution of the supergene and enabled distinct adaptive morphologies to coexist. Contrary to well-known cases where mimicry shifts cause speciation, here inversion polymorphism and demographic events may favour the maintenance of intraspecific diversity and inhibit cladogenesis.

623

Developing a SNP toolkit for management of the koala, *Phascolarctos cinereus*: from pedigrees to population genomics

Siobhan Dennison[1], Greta J Frankham[1], Jason G Bragg[2], Linda E Neaves[1, 3], Sally Potter[2], Craig Moritz[2], Mark DB Eldridge[1], Rebecca N Johnson[1]

1. Australian Centre for Wildlife Genomics, Australian Museum, Sydney, NSW, Australia

2. Research School of Biology, The Australian National University, Canberra, NSW, Australia

3. Royal Botanic Garden Edinburgh, Edinburgh, United Kingdom

The koala (*Phascolarctos cinereus*) is an iconic Australian species with a complex management history. In the northern parts of their distribution, koala populations are in decline, primarily as a result of habitat loss and fragmentation, as well as disease (chlamydia and koala retrovirus (KoRV)). Conversely, in southern parts of their distribution, some koala populations are overabundant and require fertility control measures to manage numbers. The discrepancy in population trends across their distribution means that broad scale management of koalas is not an ideal strategy. Instead, management often occurs at local scales (in local government areas or on regional populations) with little reference to the broader Australian context. Previous population genetic studies on koalas have been carried out using a range of different marker sets. These are useful in isolation, but are often not comparable across studies. This study uses next generation methods to develop a SNP assay for use in koala population genetic studies at a range of scales, from individualisation within wild and captive populations to broad scale population genetics. An exon capture method has been used to identify both functional and neutral SNPs. These SNPs will allow us to investigate geneflow and neutral genetic diversity across the koala's distribution. Additionally, as part of the Koala Genome Consortium, we will also have the opportunity to map these SNPs to the Koala Genome, providing unprecedented power to investigate a range of questions involving traits under selection, linkage, and much more. It is our hope that these markers may be utilised in future population studies, and by other researchers and management bodies to facilitate more consistent and comparable data collection for conservation management.

624

Localizing Yiddish and non-Yiddish speaking Ashkenazic Jews to primeval villages in ancient Ashkenaz lands

Eran Elhaik[1], Ranajit Das[1], Mehdi Pirooznia[2], Paul Wexler[3]

1. The University of Sheffield, Sheffield, UK, United Kingdom

2. Psychiatry and Behavioral Sciences, Johns Hopkins University, Baltimore, Maryland, United States

3. Linguistics, Tel Aviv University, Tel-Aviv, Israel

The Yiddish language is over one thousand years old and incorporates German, Slavic, and Hebrew elements and one of the last European languages whose linguistic and geographical classifications remain unclear even after three centuries. The prevalent view claims Yiddish has a German origin, whereas the opposing view posits a Slavic origin with strong Iranian and weak Turkic substrata. The strong relationship between geography, genetics, and languages prompted us to investigate the geographical origin of 393 Yiddish and non-Yiddish speaking Ashkenazic Jews (AJs), Iranian, mountain Jews and over 600 non-Jewish genomes. The Geographic Population Structure (GPS) localized most AJs along major primeval trade routes in adjacent to four villages with names that may be derived from "Ashkenaz." These are the only placenames in the world derived from this ethnonym. AJs clustered adjacently to Iranian and mountain Jews in support of a common origin. Loss of maternal haplogroups was evident in non-Yiddish speaking AJs compared to the Yiddish speakers. Our results are compatible with linguistic evidence suggesting that Yiddish has multiple origins including German, Slavic, and Hebrew. This is the first study that analyzes genetic data of Yiddish speakers, and it is carried out at a most timely manner as individuals who speak solely Yiddish are increasingly difficult to find.

625

Hurdles out of Africa: how climate and terrain shaped the history of human migrations

Anders Eriksson[1, 2]

1. University of Cambridge, Cambridge, CAMBRIDGESHIRE, United Kingdom

2. Integrative Systems Biology Lab, Division of Biological and Environmental Sciences & Engineering, King Abdullah University of Science and Technology, Thuwal (KAUST), Thuwal, Saudi Arabia

Geography and the environment shape migrations, population bottlenecks and local movements of individuals, and thereby the patterns of genetic variation within the species. In this talk I will discuss how spatially explicit models, informed by past climate and ethnically diverse datasets of human genetic variation, can be used to infer how climate and vegetation affected the spread of anatomically modern humans out of Africa into Eurasia and the Americas and shaped genetic variation. During this process, humans encountered environments that differed dramatically from those where our species originated. This would have presented both challenges and opportunities, and set the stage for adaptation. However, the effects of specific adaptations on genetic variation can be confounded by the general demographic response to the new environments, such as local population bottlenecks. I will discuss how climate-informed spatial models can help to disentangle these factors by providing null models tailored to specific geographic contexts.

626

The global evolutionary history of *Arabidopsis thaliana*

Ashley Farlow[1], Magnus Nordborg[2]

1. University of Melbourne, Parkvile, VIC, Australia

2. Gregor Mendel Institute, Vienna, WIEN, Austria

The plant *Arabidopsis thaliana* serves not only as a model organism for fundamental molecular and cellular processes, but has also greatly advanced our understanding of the origins and consequences of intraspecific genome variation. We have produced a detailed variation map from 1135 high-quality resequenced genomes representing the global population in its native Eurasian and North African range, and in recently colonized North America.

We identified relict populations that continue to inhabit ancestral habitats, primarily in the Iberian Peninsula. They have mixed with a lineage that has spread to northern latitudes from an unknown glacial refugium and is now found in a much broader range of habitats.

627

Genomic analysis of coexistence and symbiosis in *Trifolium* populations

Colleen Friel[1], Katie Wozniak[1], Sharon Strauss[2], Maren Friesen[1]

1. Michigan State University, East Lansing, MICHIGAN, United States

2. University of California Davis, Davis, California, United States

At the Bodega Marine Reserve in California, eight species of the legume *Trifolium* coexist densely on a single hillside. These legumes form symbiotic interactions with different strains of the bacterium *Rhizobium leguminosarum bv. trifolii* (hereafter called rhizobia), which colonize plant roots and fix atmospheric nitrogen in exchange for photosynthetic carbon. As such, interactions with rhizobia are important factors in plant nutrient uptake strategies. Understanding the molecular mechanisms underpinning these densely coexisting populations requires an examination of both interactions with genetically diverse rhizobial populations and host transcriptomic responses to neighboring organisms. We isolated approximately 300 *R. leguminosarum* strains from *Trifolium* nodules collected at Bodega Marine Reserve and sequenced their genomes using Illumina sequencing. We examined the phylogenetic relationships of the plants and rhizobia and trace the evolutionary trajectory of the symbiosis and species coexistence. We also generated transcriptome data for the host plants and analyzed population genomic patterns to explore selection within the host genome. Interactions between species with larger phylogenetic distances resulted in a much wider variety of gene expression changes. In addition, rhizobial strains exhibited differences in selection pressures on core and accessory genomes. This work will provide important information regarding how diversity is maintained in this ecosystem and the evolutionary pressures affecting nitrogen fixation, an ecologically and agriculturally important function.

628

Spread of reduced activity of *STX* promoter in modern humans

Naoko T Fujito[1], Yoko Satta[1], Masaya Hane[2], Atsushi Matsui[3], Ken Kitajima[2], Chihiro Sato[2], Toshiyuki Hayakawa[4]

1. School of Advanced Sciences, SOKENDAI (the Graduate University for Advanced Studies), Hayama, Kanagawa, Japan

2. Bioscience and Biotechnology Center, Nagoya University , Nagoya, Aichi, Japan

3. Primate Research Institute, Kyoto University , Inuyama, Aichi, Japan

4. The Graduate School of Systems Life Sciences, Kyushu University, Hukuoka, Japan

STX is an enzyme responsible for the transfer of polysialic acids (PSA) to a neural cell adhesion molecule (NCAM). Three SNP sites (core SNPs) in the *STX* promoter region are known to alter the promoter activity and associated with various mental disorders, such as schizophrenia, bipolar disorder and autism, suggesting that STX plays an important role in the human specific brain activity. Based on the combination of core SNPs, the haplotypes of human *STX* promoter region can be classified into four major promoter types, i.e., “CGC”, “TGT”, “TCT” and “CGT”. Interestingly, the result of promoter assays indicated that the “CGC” type, which has high frequency (35%) in Asian populations, has significantly lower promoter activity than other types. A phylogenetic study using haplotype sequences determined by molecular cloning of 63 individuals from a wide range of ethnic groups revealed that all the promoter types emerged about 0.6 MYA and each type diversified 0.1~0.2 MYA, which is coincidentally prior to the African exodus. Further analysis using haplotype sequences from the 1000 genome project data (2504 individuals) reveals that “CGC” shows high homozygosity in the ~18 kb region surrounding the core SNPs. This is consistent with the small nucleotide diversity in the Asian “CGC” type, result of LRH test and site frequency spectrum analyses. Based on these results, we propose that positive selection has acted on “CGC” in the Asian populations throughout the Great Journey, and is still ongoing.

629

Whole genome sequencing reveals the complex phylogeography of canids

Shyam Gopalakrishnan[1], Mikkel Sinding[1], Jose Samaniego[1], Marcus Thomas Gilbert

1. Natural History Museum, Copenhagen University, Copenhagen, Denmark

The study of wolves and dogs offers a great opportunity to understand the origin of new species through domestication,and their interactions with the parental species. Much effort has been devoted to the study of dogs and their domestication. Conversely, fewer samples and resources has led to gaps in our understanding of the evolutionary history of wolves. One of the important questions that remains unanswered is the relationship between the wolf subspecies and their relationships to other canids such as african wolves and golden jackals (Rueness et al. 2011). Much of our knowledge about these relationships is from mtDNA, which offers limited information (Vila et al. 1999).

To address this question, we have sequenced the genomes of several canids, including more than 25 wolves sampled from across Europe, the middle east and Africa, 2 hunting dogs and 2 golden jackals. Combining these sequenced genomes with publicly available canid genomes, we used more than 50 whole genomes to construct a well resolved phylogeny of the canids of the world, allowing us to investigate the relationships between the basal canids. For the purposes of this study, we assembled a de-novo wolf reference genome.

Preliminary results show interesting patterns in the position of wolves that have previously been shown to be mitochondrially interesting, such as the Indian, Afghan, Saudi and Mongolian wolves. We also explore the relationships between African wolves, gray wolves and Golden jackals. Results and data from this study will be valuable in answering further questions such as the timing and location of dog domestication, and local admixture between dogs and wolves.

1. Vilà, C., Amorim, I. R., Leonard, J. A., Posada, D., Castroviejo, J., Petrucci-Fonseca, F., Crandall, K. A., Ellegren, H. and Wayne, R. K. (1999), Mitochondrial DNA phylogeography and population history of the grey wolf Canis lupus. Molecular Ecology, 8: 2089–2103.
2. Rueness EK, Asmyhr MG, Sillero-Zubiri C, Macdonald DW, Bekele A, et al. (2011) The Cryptic African Wolf: Canis aureus lupaster Is Not a Golden Jackal and Is Not Endemic to Egypt. PLoS ONE 6(1): e16385.

630

Demographic history of Northeast Africa revealed by genome-wide population-genetic data

Nina Hollfelder[1], Hiba M Babiker[2], Carina M Schlebusch[1], Mattias Jakobsson[1]

1. Uppsala University, Uppsala, UPPSALA LäN, Sweden

2. Max Planck Institute for the Science of Human History, Jena, Germany

The Nile River valley has a long history of human habitation, with finds of the earliest anatomically modern humans close by and housing some of the earliest civilizations, but it has not yet been extensively explored using genome-wide population-genetic data. Due to its geographic location, history and linguistic diversity, Sudan and South Sudan cover interesting areas to investigate human demographic history. Previous studies in the area have focused on uniparentally inherited markers, microsatellites or small SNP-marker panels. We investigate 18 Sudanese and South Sudanese populations by genotyping 221 individuals for approximately 5 million SNPs using the Illumina platform. We compare this novel dataset to available SNP data of surrounding geographic regions, as well as the 1000 Genomes and HGDP datasets. This dense marker set allows us to address questions relating to the demographic history of various populations from the area and reveals detailed population structure correlating with geography and language. By using a series of summary statistics, we infer further details about the populations' demographic history. A decrease in heterozygosity in migrant populations of the Sudan, such as the Hausa, indicates a small population size. We combine genetic, geographic and linguistic information to paint a detailed picture of the Northeast African population history.

631

Genome-wide analysis to identify locally adapted genes in humans

Risa Iwasaki[1], Yoko Satta[1]

1. SOKENDAI (The Graduate University for Advanced Studies), Hayama, KANAGAWA, Japan

Human populations may have experienced both cultural and environmental adaptations, and such independent adaptations may be the cause of local genetic differentiation. It is beneficial to investigate genes which have played important roles in human population differentiation for comprehensive understanding of the relationships between human cultural/environmental adaptation and genetic diversity of human genomes. The aim of this research is identification of genes which are associated with human population differentiation through genome-wide analysis. To approach this aim, we performed pairwise comparisons of SNP allele frequencies and pairwise Wright's F_{st} of chromosome 22 among 6 populations based on category of 1000 genome project database; African, European, American, South Asian, Han Chinese in Beijing, Japanese and the other East Asian populations. We focused on the Japanese population compared to the other populations and searched for SNPs with remarkably high or low frequency in Japanese. For results, we found 1) F_{st} values showed that Japanese is most closely related to East Asian populations (Han Chinese in Beijing and the other East Asian populations), and followed by South Asian and American populations. Relatively distant relationships with African and European populations were observed. 2) When comparing two populations, the more geographically differentiated they are, the more allele frequency differences were observed. 3) No remarkable changes of allele frequencies in Japanese relative to the other populations were detected, thus the other chromosomes will be further examined.

632

Mathematical Model of Missing African Genomic Variation: A case for robust sampling

Latifa Jackson[1], Raouf Ghomrasni[2], Fatimah Jackson[1]

1. Howard University, Washington, DISTRICT OF COLUMBIA, United States

2. African Institute for Mathematical Science, Cape Town, South Africa

Current African genetic databases are woefully inadequate to describe the genetic variation found among African Diaspora populations. This leads to significant difficulties in estimating admixture proportions in highly heterogeneous modern groups. What are the implications of this missing

diversity for developing robust reference databases? Using mathematical models to evaluate the impact of missing genetic diversity data, we estimate how much unaccounted for diversity data exists in Africa among different ethnic and regional populations. We estimate that we capture less than half of the genetic variation in exant human populations with the current publicly available population resources. These missing diversity demonstrate the lack of power in our current genetic sampling protocols and reveal the extent to which this absent African genetic diversity undermines comprehensive representations of the continent's people.

633

Evolution of the Coldblooded trotter breed: Where did the speed come from?

Kim Jäderkvist Fegraeus[1], Brandon D Velie[1], Maja Wiklund[1], Maria K Johansson[1], Jennifer R.S. Meadows[2], Gabriella Lindgren[1]

1. Swedish University of Agricultural Sciences, Uppsala, Sweden

2. Uppsala University, Uppsala, Sweden

The origin of the Coldblooded trotter (CT) provides a unique opportunity to identify genes influencing racing performance. The CT originates from the North Swedish draught horse (NS) and these two breeds retain high levels of genetic similarity (Fst = 0.08). However, prior to the introduction of paternity testing in 1969, crossbreeding with the Standardbred (SB) was used to improve CT performance. We hypothesize that the gains in CT performance over the last 50 years may in part be explained by the maintenance of favorable genetic variants originating from the SB. As such, the aim of the current study was to compare the genetic makeup of these three breeds and to identify genetic footprints of athletic performance. A sliding window Delta Fst analysis was performed across all breeds using data generated from the equine SNP50K array (CT, n=11; NS, n=19; SB, n=12). Five key regions were revealed where the CT and SB were genetically similar, but together differed from the NS. Seven genes reside in these segments, some of which affect muscle metabolism, such as regulation of cell growth in response to nutrient and mitochondrial DNA function in muscle. Genotyping of the top four differentiated markers in additional CT (n>130) showed that two were significantly associated ($p<0.05$) with performance traits, e.g. number of victories and time records. Additional genotyping will now be conducted in diverse racing breeds to ascertain the importance of these variants, and the current study will be augmented with additional Delta Fst analysis using whole genome sequence data.

634

Parallel speciation in an Australian wildflower

Maria C Melo[1, 2], Maddie E James[1], Federico Roda[1, 3], Diana Bernal-Franco[1, 4], Melanie J Wilkinson[1], Huanle Liu[1, 5], Daniel Ortiz-Barrientos[1]

1. School of Biological Sciences, The University of Queensland, Brisbane, Queensland, Australia

2. IST Austria, Klosterneuburg, Lower Austria, Austria

3. Department of Organismic and Evolutionary Biology, The Arnold Arboretum of Harvard University, Boston, Massachusetts, USA

4. Chemistry & Chemical Biology Department, Northeastern University, Boston, Massachusetts, USA

5. Institute for Molecular Bioscience, The University of Queensland, Brisbane, Queensland, Australia

In a number of animal species, reproductive isolation has repeatedly and independently evolved between populations adapted to contrasting but not to similar environments. This process is commonly known as parallel speciation, and examples in plants are enigmatically rare. Here, we show that natural selection has repeatedly and independently driven the evolution of reproductive isolation between the coastal ecotypes of the groundsel *Senecio lautus*, a herbaceous plant found in Australia. We found that in crosses between populations separated by half a million years of divergence, crosses within the same ecotype showed greater fertility than crosses between different ecotypes. This pattern of reproductive isolation was similar but weaker when crossing recently diverged populations of these ecotypes. Unexpectedly, molecular estimates of gene flow between parapatric populations were nil, suggesting that despite being largely interfertile, parapatric populations from different ecotypes must have experienced little gene flow in the recent past. These results suggest that in *Senecio*, ecological adaptation facilitates the evolution of both extrinsic and intrinsic reproductive isolation across a complex geographic landscape and leads to the repeated and independent origin of multiple plant species in parapatry.

637

Structural variants in yeast have strong effects on quantitative traits and reproductive isolation, and are transient in natural populations.

Daniel Jeffares[1], Clemency Jolly[1], Mimoza Hoti[1], Doug Speed[1], Charalampos Rallis[1], Christophe Dessimoz[1], Jürg Bähler[1], Fritz J Sedlazeck[2]

1. University College London, London, LONDON, United Kingdom

2. Johns Hopkins University, Baltimore, USA

Large structural variations (SVs) in the genomes are harder to identify than smaller genetic variants, and so their effects have been studies far less intensively. However, they are increasingly suspected to be major contributors to phenotypic diversity, reproductive isolation, adaptation therefore to evolution.

Using genome sequences from a worldwide sample of 160 natural isolates of the fission yeast *Schizosaccharomyces pombe*, we created a high-quality, curated catalog of structural variations, including duplications, deletions, inversions and translocations. We described the effects of these variants on gene expression, their contributions to 53 quantitative traits, and their influence on intrinsic reproductive isolation.

We uncovered several interesting facets of structural variant biology. We found that copy number variants (CNVs) frequently segregate within closely related clonal populations and are in weak linkage with single nucleotide polymorphisms (SNPs), indicating rapid turnover. These transient CNVs produce stoichiometric effects on gene expression. SVs in general contribute an average of 19% of trait variance (SNPs contribute 30% on average), with the majority of this effect being due to CNVs. Variation in some traits, including our recently characterized winemaking traits were entirely due to the effects of structural variants, with no measureable contribution from SNPs. Rearrangements (inversions & translocations), in contrast contribute strongly to reproductive isolation, but little to trait variation. Collectively, these findings have broad implications for evolution and for our understanding of quantitative traits including complex human diseases.

A preprint of this research is available on bioRxiv, see: danieljeffares.com/publications/

638

Consanguinity and runs of homozygosity in Jewish populations

Jonathan T. L. Kang[1], Amy Goldberg[1], Michael D. Edge[1], Noah A. Rosenberg[1]

1. Stanford University, Stanford, CALIFORNIA, United States

Recent studies have highlighted the potential of analyses of genomic sharing to produce insight into the demographic processes affecting human populations. We study runs of homozygosity (ROH) in 18 Jewish populations, examining these groups in relation to 123 non-Jewish populations sampled worldwide. By using a model-based clustering method to sort ROH into three length classes—short, intermediate, and long—we examine the impact of a variety of demographic processes on genomic patterns in Jewish populations. Interestingly, we find that the portion of the genome appearing in long ROH—the length class most directly related to recent consanguinity—closely accords with demographic data gathered during the 1950s on consanguineous unions in the various Jewish groups. The dissection of ROH into length classes and the comparison to consanguinity data provide insight into a number of additional phenomena, including differences between Jewish and non-Jewish populations in ROH patterns, the relative lengths of identity-by-descent tracts in different Jewish groups, and the nature of the "population isolate" status of the Ashkenazi Jews.

639

Breaking the infinite sites model: widespread mutational recurrence in exome sequence data from over 60,000 individuals

Konrad J Karczewski[2, 1], Monkol Lek[2, 1], Eric V Minikel[2, 1], Kaitlin E Samocha[2, 1], Jack Kosmicki[2, 1], Exome Aggregation Consortium, Mark J Daly[2, 1], Daniel G MacArthur[2, 1]

1. Broad Institute, Cambridge, Massachusetts, United States

2. Massachusetts General Hospital, Boston, Massachusetts, United States

Many population genetics models assume that if a variant is observed twice, the two observations are a result of identity-by-descent inheritance. However, as the number of sequenced individuals grows, the probability of observing two or more independent mutational events occurring at the same site in the genome increases. Here, we describe an analysis of widespread mutational recurrence observed in exome sequence data from 60,706 individuals from the Exome Aggregation Consortium (ExAC). This effect is most pronounced among highly mutable CpG transitions, and in this dataset, we observe over 60% of **all possible** synonymous CpG mutations and begin to saturate detection of these variants.

We find that approximately one-third of high-confidence validated *de novo* variants identified in external datasets of parent-offspring trios are also observed independently in the ExAC dataset, indicating that the same variant has arisen multiple times independently.

This process has a marked effect on the frequency spectrum in the ExAC data, resulting in a depletion of very low-frequency variants at sites with high mutation rates, even for synonymous sites. Specifically, we observe a strong correlation between site mutability inferred from sequence and singleton rates, as well as between site mutability and the probability of observing the variant in two separate populations.

We demonstrate that these patterns are only observed at a sample size greater than approximately 20,000 individuals, indicating that ExAC is the first such dataset to observe this phenomenon. Finally, we propose a correction factor to properly account for the impact of mutational recurrence on the frequency spectra of various functional classes, which enables us to provide robust estimates of their deleteriousness. We note that with a moderately larger sample size, we will be able to infer selection against individual CpG variants.

640

Inferring the unfolded site frequency spectrum and using it to quantify adaptive molecular evolution in Drosophila

Peter Keightley[1], Tom Booker[1], Jose Campos[1], Brian Charlesworth[1]

1. University of Edinburgh, Charlotte Auerbach Rd, UK, United Kingdom

The unfolded site frequency spectrum (uSFS) is a vector of counts of sites with different numbers of copies of the derived allele in a sample of gene copies from a population. The uSFS contains extra information compared to the folded SFS, a vector of counts of sites with different numbers of copies of the minor allele. Inferring the uSFS depends on using outgroups to estimate the frequency of sites with the ancestral versus the derived allele, potentially leading to statistical uncertainty because of multiple hits in the outgroup lineages. We present a new approach to inferring the uSFS, which we test by simulations. We show that there is usually a substantial increase in precision from using two outgroups rather than a single outgroup. We apply the approach to infer the uSFSs for synonymous and nonsynonymous sites of protein-coding genes in Drosophila using polymorphism data from whole-genome sequencing projects and the sequences of outgroup species. We then use the uSFSs along with the software DFE-alpha to infer the distribution of fitness effects of new mutations, i.e., the relative frequencies and effects of deleterious and advantageous mutations. We show that models with a significant fraction of advantageous mutations fit the polymorphism data substantially better than models that assume there are only deleterious or neutral mutations.

641

Genome-wide association study of copy number variation for detection of genes affecting fat distribution in pig

Sojeong Ka[1], Young-Sup Lee[1], Seoae Cho [2], Heebal Kim[1]

1. Department of Agricultural Biotechnology and Research Institute of Agriculture and Life Sciences, Seoul National University, Seoul, South Korea

2. CHO&KIM Genomics, Seoul National University Research Park, Seoul, Republic of Korea

The study of fat deposition is an important research area with implications for both human health and improvement of animal traits. The pig is one of the most important species to this field, as they are excellent models for research in energy metabolism and fatness. Recent biomedical studies have revealed that different fat depots may possess intrinsic characteristics against each other, which might be partly attributed to genetic factors.

In this study, we examined genetic variation and degree of obesity using data from the Korea Associated Resource cohorts, which included 8,842 individuals, in order to understand the genetic mechanisms underlying depot-specific fat accumulation. Genome-wide association studies of copy number variation (CNV) affecting subscapular (SUB) and suprailiac (SUP) skin fold thickness have identified many chromosomal regions and genes that affect human fat accumulation in male SUB, male SUP, female SUB, and female SUP sites. In order to relate human obesity to pig fatness, we compared human CNV genes to those located in previously known CNV regions identified using 18 diverse pig populations. Results of our

comparative analysis revealed that human and pig share multiple candidate CNV genes, and indicated that redundancy of those genes may play roles in pig depot-specific fatness.

Here, we investigate the distribution of the obesity-related CNVs in human genome and present CNV genes as potential candidate genes controlling fat deposition in pig. This comparative study on fat distribution provides a basis for parallel studies aimed at understanding similarities and differences in control of site-specific adiposity in pig and human.

642

Experimental evolution in *Drosophila* analysed with Gaussian process models

Agnes Jonas[1, 2], Carolin Kosiol[1]

1. Institute of Population Genetics, Vetmeduni Vienna, Vienna, Austria

2. Vienna Graduate School of Population Genetics, Vienna, Austria

In population genomics, the observation of changes in allele frequencies play an important role in understanding evolutionary forces. Recent advances in sequencing technologies have made it possible to observe evolutionary trajectories in more detail than ever. It has become feasible not only to sequence the last generation of a population at the end of long-term treatments but to monitor experimentally genome evolution at intermediate generations in *Drosophila*. However, distinguishing the alleles that are changing under selection from those just displaying genetic drift is challenging due to the large number of false positives.

Here we present a Gaussian Process (GP) approach to model the evolutionary time series data. First we infer the Single Nucleotide Polymorphism (SNP) frequencies and their observation noise variances from sequencing data under a Beta-Binomial model for the read counts from Next Generation Sequencing. Then, we fit time-dependent and time-independent GP models to logistic transformed frequencies, while incorporating the inferred noise variances in the models. Finally, we compute the Bayes Factors between time-dependent and time-independent GP models and rank the SNPs according to their Bayes Factors.

We compare the performances of our method and pairwise statistical test on a simulated dataset which mimics the sequencing data of Drosophila, with four replicates along eight generations. Results show that our method outperforms with a higher precision at the same recall rate and making use of the inferred noise variance in the GP models helps to decrease the number of false positives.

We present results from applying our approach on real data from *Drosophila simulans* experimental evolution for temperature adaptation. We also show preliminary results that the proposed method is able to find signatures of selection on ACE gene.

643

The impact of selection on cis-regulatory variation across the genome of an outcrossing plant

Benjamin Laenen[1], Kim A. Steige[1], Johan Reimegård[2], Douglas G. Scofield[3], Tanja Slotte[1]

1. Stockholm university, Solna, Sweden

2. Bioinformatics Long-term Support (WABI), Scilifelab, Solna, Sweden

3. Department of Ecology and Genetics, Evolutionary Biology, Evolutionsbiologiskt Centrum, Uppsala, Sweden

While cis-regulatory changes have long been suggested to be particularly important for adaptation, our understanding of what determines cis-regulatory variation remains limited in most species. Here, we have investigated the prevalence, selective importance, and genomic correlates of cis-regulatory variation in the outcrossing crucifer species Capsella grandiflora. We identify genes with cis-regulatory variation through analyses of allele-specific expression (ASE) in deep transcriptome sequencing data from flower buds and leaves, and use population genomic analyses of whole genome resequencing data from both a range-wide sample and a natural population to quantify the impact of positive and purifying selection on these genes. Our results show that in C. grandiflora, cis-regulatory variation is pervasive, affecting an average of 35% of genes within individual plants. Genes harboring cis-regulatory variation are (1) under weaker purifying selection, (2) significantly more likely to harbor nearby transposable element (TE) insertions, and (3) undergo lower rates of adaptive substitutions in comparison to other genes. These results are robust to correction for nonequilibrium demography and expression level variation among genes. In a logistic regression model, we identified presence of nearby TE insertions as a major factor increasing the odds of ASE, whereas gene body methylation was a major factor associated with reduced odds of ASE. Our findings suggest that gene body methylated genes are not only strongly conserved at the sequence level but also with respect to cis-regulatory variation. These results suggest that variation in the intensity of purifying selection across the genome, in part determined by gene body methylation, is a major determinant of the presence of intraspecific cis-regulatory variation in this outcrossing plant species.

644

Selective breeding mediated genetic mapping for tameness in mouse

Yuki Matsumoto[1, 2, 3], Jo Nishino[4], Tatsuhiko Goto[5, 3], Tsuyoshi Koide[1, 2]

1. Department of Genetics, SOKENDAI, Mishima, Shizuoka, Japan

2. Mouse Genomics Resource Laboratory, National Institute of Genetics, Mishima, Shizuoka, Japan

3. JSPS, Chiyoda-ku, Tokyo, Japan

4. Graduate School of Medicine, Nagoya University, Nagoya, Aichi, Japan

5. School of Life Sciences, The University of Nottingham, University Park, Nottingham, United Kingdom

Tameness is a major behavioral factor for the domestication. It can be divided into two potential components: motivation to approach humans (active tameness) and reluctance to avoid them (passive tameness), and we previously established behavioral assay for the quantification of the both types of tameness in mice (*Mus musculus*). To identify genes associated with active tameness, we performed selective breeding for contacting (defined as contacting human hand) using wild-derived heterogeneous stock (WHS). WHS is a mixed population derived from 8 wild mouse strains. At the 8th generation of the selective breeding, contacting in selected population increased although control group did not. Then by using GigaMUGA SNP genotyping array, we obtained 20,530 single nucleotide polymorphism (SNP) data of 32 mice in both selected and control population, and the 8 founder strains of WHS. Because the alleles associated with contacting should be increased in frequency by selective breeding, the selected loci can be identified by using the deviation from hypothetical allele frequency determined by computer simulation. We used the simulation based on non-selection model combined the pedigree information, genotype and the SNP position. Then we determined genome-wide thresholds for significant increase of allele frequency, and applied the threshold to observed data in the populations. As a result, we found a SNP on chromosome 11 exceeded the threshold. By using database, we will discuss about candidate genes within the detected SNP and the results of comparative genomics between mouse and dog, which is one of the major domesticated animal and shows tameness.

645

Can we see ecology in sequence data?

Michael J McDonald[1], Michael J McDonald[2]

1. School of Biological Sciences, Monash University, Melbourne, Australia

2. FAS Center for Systems Biology, Harvard University, Cambridge, Massachusetts, United States

Metagenomics has provided insight into the constituents of a variety of populations and ecosystems. In fact, often the only information we have on the members of an ecological community is their genome sequence. We are reasonably adept at detecting functions possessed by individuals, but can we use sequence data to determine the dynamics of evolution and ecology within these communities?

Experimental populations of microbes provide a great model system to address this question. Here I describe experiments in yeast and *E. coli* using time resolved sequencing to detect eco-evolutionary dynamics within evolving populations.

In yeast we study the spontaneous evolution of frequency dependence, and show how gene flow can allow the persistence of niche specific genotypes, even while newly arising mutations fix across the microcosm. I will also present new work from a long term *E. coli* evolution experiment, where our whole population sequencing at every 500 generations of a 60,000 generation experiment reveals the surprising prevalence of ecotype proliferation.

1. Sex speeds adaptation by altering the dynamics of molecular evolution. McDonald MJ, Rice DP, Desai MM. Nature. 2016 Mar 10;531(7593):233-6.
2. Crowded growth leads to the spontaneous evolution of semistable coexistence in laboratory yeast populations. Frenkel EM, McDonald MJ, Van Dyken JD, Kosheleva K, Lang GI, Desai MM. PNAS. 2015 Sep 8;112(36):11306-11.

647

Large-scale whole genome sequencing of the Estonian population reveals novel loss-of-function variants and new insights into the population history

Mait Metspalu[1], Georgi Hudjashov[1, 2], Mario Mitt[3], Luca Pagani[1, 4], Lauri Saag[1], Mart Kals[3], Kalle Pärn[3], Lili Milani[3], Daniel John Lawson[5], Tõnu Esko[3], Andres Metspalu[3], Reedik Mägi[3]

1. Evolutionary genomics group, Estonian Biocentre, Tartu, Estonia

2. Statistics and Bioinformatics Group, Institute of Fundamental Sciences, Massey University, , Palmerston, New Zealand

3. Estonian Genome Center, University of Tartu, Tartu, Estonia

4. Department of Biological Anthropology, University of Cambridge, Cambridge, United Kingdom

5. Integrative Epidemiology Unit, School of Social and Community Medicine, University of Bristol, Bristol, United Kingdom

Altogether 2244 whole genomes of geographically diverse individuals from Estonia were sequenced to a median depth of 30x using Illumina HiSeq with TruSeq PCR-free library preparation method. We found 19M SNVs and 6.6M indel variants with allele count larger than two and of which 8.4M were novel. Within this study we have analysed both loss-of-function variants revealed as well as the population structure of Estonia.

We found a total of 14,531 autosomal loss-of-function (LOF) SNVs and indels in 6,991 genes. Out of these genes, 3.3% contained homozygous or compound heterozygotes LOF variants with minor allele frequency less than 2%. By combining the data of complete gene knockouts of individuals with their disease history and variety of available endophenotypes (proteomics, NMR, biochemistry) will help us to study the function of these genes and will lead to better understanding the phenotypic consequences of the variation within these genes.

To study the fine-scale genetic structure of the Estonian population, we concentrate on a subset of the genomes (N=436), which comprehensively cover rural Estonia to minimize the mixing effect of historical urbanization. We further combine these genomes with a pan Eurasian panel of high coverage genomes from hundreds of populations. Using haplotype and allele frequency based methods we show that the genetic structure within Estonia is largely in line with the division of *inland* vs. *maritime* Estonia what has been proposed based on archaeological findings. Furthermore, we identify and quantify the relative contributions of the three major genetic domains of the European gene pool in Estonians and estimate split times from linguistically and geographically adjacent populations. We use Finestructure and inter-population doubleton distribution to reveal patterns of genetic sharing between Estonians and other European populations and infer population history in a series of population splits and admixture events in pre-historic and historic times.

648

kWIP: The *k*-mer Weighted Inner Product, a *de novo* estimator of genetic relatedness.

Kevin Murray[1], Christfried Webers[2], Cheng Soon Ong[2], Norman Warthmann[1]

1. ANU, Acton, ACT, Australia

2. NICTA, Canberra, ACT, Australia

Modern techniques in population genomics generate unprecedented quantities of data within which complex genetic histories reside. The scale and complexity of these data require the development of new approaches to the analysis of genetic data. We present the k-mer Weighted Inner Product, a de novo, alignment free measure of genetic similarity between samples in a population. kWIP, is an efficient tool implementing this metric that can determine the genetic relatedness between samples without alignment or assembly. We show kWIP can reconstruct the true relatedness between samples directly from sequencing reads generated with various modern sequencing platforms.

kWIP works by decomposing sequencing reads to short k-mers, hashing these k-mers using a constant-memory data structure, and performing pairwise distance calculation between these sample k-mer hashes. The power of kWIP comes from the weighting applied across different hash values, which decreases the effect of erroneous, rare or over-abundant k-mers while focusing on k-mers which give the most insight into the similarity of samples. We use simulation studies to quantify the increased accuracy of this weighting over existing

unweighted distance metrics. kWIP is free, open source software implemented in C++ and released under the GNU LGPL v3.

649

Functional Validation of Human Parkinson's Disease GWAS in the Fly

Lisa Oyston[3, 1, 2], Christian Weichenberger[4], Hagen Blakenburg[4], Andrew Hicks[4], Greg Neely[1, 2]

1. Charles Perkins Centre, School of Molecular Bioscience, University of Sydney, Sydney, NSW, Australia

2. Neuroscience Division, Garvan Institute of Medical Research, Sydney, NSW, Australia

3. St Vincent's Clinical School and School of Medical Sciences, University of New South Wales, Sydney, NSW, Australia

4. Centre for Biomedicine, EURAC Research, Bolzano, Italy

Sequencing of the human genome allowed for large-scale whole-genome population studies to identify the genetic underpinning of common diseases. While thousands of genome-wide association studies (GWAS) have now been published, a systematic functional validation of these efforts is lacking. Our goal in this study was to test if GWAS approaches do indeed identify genes/loci that modify disease severity. We focused on Parkinson's disease (PD), a complex neurodegenerative disease that severely impairs motor function with age. In *Drosophila melanogaster*, PD can be modeled by pan-neuronal (*elav-gal4*) ectopic expression of human alpha-synuclein (*UAS-αsyn*). Meta-analysis of GWAS data from ~25,000 PD patients and >100,000 age-matched controls identified SNPs of varying *p* values ($<10^{-4}$) which corresponded to 845 fly orthologs. Here we have systematically tested each candidate for PD-modifying effects by monitoring locomotor function throughout lifespan. Assessment of this initial dataset has allowed us to assess the validity of using GWAS to predict disease modifying genes. We have been able to show that probability of a disease-modifying phenotype inversely correlates with GWAS p-value and the distance between gene and SNP. Thus we conclude that at least for large studies, GWAS will pinpoint disease-modifying loci and has allowed us to identify multiple new conserved PD genes, which can inform on basic disease mechanisms or be considered as novel drug targets for this devastating illness.

650

Mitochondrial Genomes Reveal the Complex Demographic History of Human Populations in Eastern Pamirs

Min-Sheng Peng[1, 2], Wei-Fang Xu[3], Jiao-Jiao Song[1, 4], Xing Chen[1], Xierzhatijiang Sulaiman[3], Shi-Fang Wu[1], Gui-Mei Li[1], Najmidinov T Abdulloevich[5], Manilovna E Afanasovna[5], Khudoydodov B Ibrohimovich[5], Xi Chen[6, 7], Halmurat Upur[3], Ya-Ping Zhang[1, 2, 8]

1. Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, Yunnan, China

2. KIZ/CUHK Joint Laboratory of Bioresources and Molecular Research in Common Diseases, Kunming, Yunnan, China

3. Xinjiang Medical University, Ürümqi, Xinjiang, China

4. Institute of Health Sciences, Anhui University, Hefei, Anhui, China

5. E.N. Pavlovsky Institute of Zoology and Parasitology, Academy of Sciences of Republic of Tajikistan, Dushanbe, Tajikistan

6. Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences, Ürümqi, Xinjiang, China

7. Research Center for Ecology and Environment of Central Asia, Chinese Academy of Sciences, Ürümqi, Xinjiang, China

8. Yunnan University, Kunming, Yunnan, China

Eastern Pamirs is among the world's highest mountains in Central Asia. The population history for this region is still unclear. To address this issue, we analyzed mitochondrial DNA (mtDNA) sequence variations in Sarikoli (n=89), Wakhi (n=67), and Kyrgyz (n=69) populations from East Pamirs in Xinjiang, China. The surrounding lowland populations, including Tajiks (n=28) from Tajikistan, Kyrgyz (n=54) and Uyghur (n=27) people from western Xinjiang, China, were also investigated for comparisons. A total of 334 mitochondrial genomes were sequenced by using Ion Torrent PGM. The mtDNA haplogroup profiles for those populations were indicated. And some novel sub-haplogroups within haplogroups D4 and H were classified. In addition to East (e.g. A – D) and West (e.g. H – K) Eurasian mtDNA haplogroups, some South Asian specific lineages (e.g. M3 and M5) were detected in most populations. Within those haplogroups, the distribution of related sub-haplogroups showed differences among those populations, suggesting a complex admixture history in Eastern Pamirs. The Bayesian Skyline plots revealed different demographic history between highlanders and their surrounding lowlanders.

651

Robust identification of hard and soft sweeps in humans via machine learning

Daniel R Schrider[1], Andrew D Kern[1]

1. Rutgers University, Piscataway, NJ, USA

Detecting the targets of adaptive natural selection from whole genome sequencing data is a central problem for population genetics. Numerous approaches have been devised to detect the population genetic signature of a *de novo* beneficial mutation sweeping rapidly to fixation (a hard selective sweep). To date most of these methods to detect sweeps show poor performance under realistic demographic scenarios. Moreover, over the past decade there has been a renewed interest in determining the importance of selection from standing variation (soft sweeps) in adaptation of natural populations, yet few methods are sensitive to this mode of selection. Here we introduce a new tool, S/HIC, which uses supervised machine learning to precisely infer the location of both hard and soft selective sweeps. We show that S/HIC has unrivaled accuracy for detecting sweeps under demographic histories that are relevant to natural populations, and distinguishing sweeps from linked as well as neutrally evolving regions. Moreover we show that S/HIC is uniquely robust among its competitors to demographic misspecification: even if the true demographic model of a population differs catastrophically from that specified by the user, S/HIC still retains impressive discriminatory power. Next, we apply S/HIC to resequencing data from human European and African population samples from the 1000 Genomes Project. S/HIC reliably recovers selective sweeps that have been identified earlier using less specific and sensitive methods, and identifies several compelling novel candidates, including a tumor suppressor gene that is often mutated or deleted in breast tumors. Lastly we perform the first genome-wide examination of the prominence of hard versus soft sweeps in human populations, finding a much greater frequency of soft sweeps in Africa. This result confirms theoretical predictions that larger populations will more often respond to adaptive challenges by selecting on previously standing polymorphisms.

652

Beyond clines: lineages and haplotype blocks in hybrid zones

Alisa Sedghifar[1], Yaniv Brandvain[2], Peter Ralph[3]

1. Lewis-Sigler Institute for integrative genomics, Princeton University, Princeton, NJ, United States

2. Plant Biology, University of Minnesota, St Paul, MN, United States

3. Computational Biology and Bioinformatics, University of Southern California, Los Angeles, CA, United States

Hybrid zones formed between recently diverged populations offer an opportunity to study the mechanisms underlying reproductive isolation and the process of speciation. In particular, selection against hybrids can impact genome-wide patterns of introgression. Loci associated with incompatibilities are rapidly removed by selection when in the wrong species background and as a result, in populations away from the hybrid zone center, will be present in lineages that are relatively recent migrants. We expect this to be reflected in longer tracts of ancestry surrounding targets of selection.

To test this prediction, we use a combination of analytical theory and simulations to describe the movement and breaking up of lineages as a result of migration, hybridization and recombination. Under our Brownian model, we find that blocks of ancestry surrounding single-locus incompatibilities can be substantially longer than the genome-wide average, and that locally disadvantageous alleles tend to exist as smaller families.

These patterns may be used to characterize the age of hybrid zones and to identify targets of selection, thereby deepening our understanding of the genetic factors driving lineage-specific adaptation and reproductive isolation between species. Studies of selection in hybrid zones have traditionally focused on cline width at individual loci, and the availability of population genomic data now enables use of the additional information contained in tracts of unbroken ancestry to better understand the processes operating in these populations.

653

A molecular investigation of the relationships of karaka (*Corynocarpus laevigatus).*

Lara Shepherd[1], Robin Atherton[2], Peter de Lange[3], David Chagne[4], Nick Roskruge[2], Peter Lockhart[2]

1. Museum of New Zealand Te Papa Tongarewa, Wellington, New Zealand

2. Massey University, Palmerston North

3. Department of Conservation , Auckland

4. Plant and Food Research, Palmerston North

Corynocarpus laevigatus (karaka, kōpi) is a small tree whose kernels provided an important food source for New Zealand Māori before European settlement. Much of the contemporary distribution of karaka, including its presence on the distant Kermadec and Chatham Island groups, is considered to have resulted from translocations as part of its cultivation.

We examined the patterns of pre-European translocation of karaka using sequences from two nuclear loci and SNPs developed from whole chloroplast DNA sequences screened with high resolution melting (HRM). Our results indicated low levels of genetic diversity in karaka at both chloroplast and nuclear loci. However, our New Zealand-wide sampling revealed a reduction in genetic diversity in the translocated populations compared to the natural range in northern New Zealand. The distinctiveness of specimens from the Three Kings Islands excluded this region as the source population for translocated karaka. For the remaining populations the lack of genetic diversity, combined with low levels of genetic structuring, prevented more precise identification of the source of translocated karaka.

654

Detecting introgressed archaic haplotypes in Oceanic population genome sequences

Laurits LS Skov[1], Anders Bergstrom[1], Yali Xue[1], Chris Tyler-Smith[1], Richard Durbin[1]

1. Sanger Institute, Hinxton, CAMBRIDGE, United Kingdom

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

Most published methods for identification of archaic haplotypes rely on ancient DNA samples from the archaic population, to compare the modern samples with. Here we present a method for identifying regions of modern whole genome sequences that have been introgressed into a subset of modern humans from an ancestor with a long history of separation from the modern human lineage, and apply it to Oceanian genomes. The method takes advantage of the fact that introgressed regions show different patterns of LD and a high density of SNPs private to the population that received the introgression. For the Oceanian samples, we look for regions that have a clear excess of variants not seen in any 1000 Genomes Project sample, in order to identify regions introgressed from archaic populations other than Neanderthals, after the separation of Oceanians from other modern Asian ancestral populations.

We identify tens of Mb of potentially private introgressed sequence for each of the individuals in the study. The regions will be compared to regions found by other methods traditionally used for identifying introgressed regions, to known archaic sequences from ancient DNA, to each other to understand their diversity, and to other modern human sequences to estimate their original separation time.

656

ddRAD-based target capture across *Jasus* lobster species to assess spatial-temporal adaptive variation

Carla A Souza[1], Nick Murphy[1], Bridget Green[2], James Bell[3], Jan Strugnell[1]

1. La Trobe University, Melbourne, VIC, Australia

2. Institue for the Marine and Antartic Studies, University of Tasmania, Hobart, TAS, Australia

3. Victoria University of Wellington, Wellington, New Zealand

ddRAD-Seq[1] and target capture[2] sequencing methods have widely been used to explore a range of population and phylogenetic questions[3] including historic samples[4]. ddRAD-seq methods result in large amounts of missing data compared to target capture due to allelic dropout (especially when comparing across species) and requires very high quality DNA[5]. Even though most target enrichment methods strongly rely on pre-existing transcriptome or genome resources, such data are still scarce or poorly annotated for the *Jasus* genus and closely related species. The *Jasus* clade encompasses economically important lobster fisheries around the Southern Hemisphere. These species have some of

the longest planktonic larval stages (1-2 years), which can be transported over long distances[6]. Here, we tested a target enrichment approach based on a set of loci discovered in ddRAD-seq libraries of *Jasus edwardsii* and *Sagmariasus verreauxi* to investigate the role of adaptation and self-recruitment processes underlying diversification in a range of *Jasus* species. A total of 2,241 probes were designed to target specific regions in contemporary and historic specimens (museum-collection) from *J. caveorum, J. edwardsii, J. frontalis, J. lalandii, J. paulensis and J. tristani.* The method efficiency was evaluated with respect to number of targets successfully recovered across species, which varied from 671 for *J. frontalis* to 1,308 for *J. tristani*. Only 279 were shared across all valid individuals from the six species. Patterns of mismatches (C to T and G to A substitution bias) resulting from DNA damage were compared between historic and contemporary samples enabling further data filtering. Our results show the feasibility of integrating two genomic methods to enrich genetic data for which no reference genome nor transcriptome are available. Moreover, the methods we employed were successfully applied to museum collection samples, which are unique resources to compare progressive changes in genetic diversity in lobster populations over time.

1. 1. Peterson, B. K. et al. Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. PLoS One 7, e37135 (2012).
2. 2. Rohland, N. & Reich, D. Cost-effective, high-throughput DNA sequencing libraries for multiplexed target capture. Genome Res. 22, 939–946 (2012).
3. 3. McCormack, J. E. et al. Applications of next-generation sequencing to phylogeography and phylogenetics. Mol. Phylogenet. Evol. 66, 526–38 (2013).
4. 4. Bi, K. et al. Unlocking the vault: Next-generation museum population genomics. Mol. Ecol. 22, 6018–6032 (2013).
5. 5. Leache, a. D. et al. Phylogenomics of Phrynosomatid Lizards: Conflicting Signals from Sequence Capture Versus Restriction Site Associated DNA Sequencing. Genome Biol. Evol. 7, 706–719 (2015).
6. 6. Punt, A. E. et al. Estimating the size-transition matrix for Tasmanian rock lobster, Jasus edwardsii. Mar. Freshw. Res. 48, 981 (1997).

657

Ecological divergence and reproductive isolation in a colour polymorphic Anolis lizard.

Jessica Stapley[1], Adam C Algar[2], Brian Bock[3], Eldredge Bermingham[4], Jon Slate[5], Stuart R Dennis[6]

1. ETH, Zurich, Zurich, ZURICH, Switzerland

2. University of Nottingham, Nottingham, United Kingdom

3. University of Antioquia, Medellín, Colombia

4. Patricia and Phillip Frost Museum of Science, Florida, USA

5. University of Sheffield, Sheffield, United Kingdom

6. Aquatic Research, EAWAG, Dübendorf, Switzerland

The *Anolis* lizards are a well-known example of an adaptive radiation, but the evolution of one of their most characteristic features, the dewlap, remains somewhat of a mystery. The dewlap is an extendable throat-fan, which is used for sexual displays and visual communication. Across the genus there is spectacular variation in colour, pattern and size of the dewlap, but our understanding of how this variation contributes to population divergence and speciation is limited. To address this gap, I have leveraged a discrete colour-pattern polymorphism in a Panamanian Anolis lizard, with mate choice experiments, lab crosses, RAD-tag sequencing and environmental data to investigate the evolution and maintenance of this colour polymorphism. I have determined that dewlap colour-pattern is a mendelian trait. The distribution of colour-pattern morphs is related to an environmental gradient, consistent with what we predict based on signalling theory: darker dewlaps in drier environments, and individuals with a different dewlap colour-pattern vary in life history characteristics, including growth and reproductive rate. I also found evidence of incomplete reproductive isolation between morphs and populations, in the form of colour assortative mating and reduced fitness of F1 hybrids. Using genome-wide SNP data I have estimated genetic differentiation between these populations and found that this was related to the environmental gradient and phenotypic divergence. Between populations geneflow was high, but several loci were found to be strongly associated with the environmental gradient and will be interesting candidate loci for future work. This study represents a comprehensive body of work detailing the relationships between environmental, phenotypic and genetic variation in this colour polymorphic species. The results suggest that ecologically driven divergence in dewlap colour-pattern is contributing to reproductive isolation and population divergence.

658

Prioritizing Candidate Genetic Variants Driving Adaptations in Human Populations

Michal Szpak[1], Yali Xue[1], Qasim Ayub[1], Yuan Chen[1], Chris Tyler-Smith[1]

1. Wellcome Trust Sanger Institute, Hinxton, United Kingdom

The focus of positive selection studies in humans needs to move from candidate locus discovery to pinpointing underlying causal variants and further investigation of their biological significance. We performed a meta-analysis of published selection screens and extended these with an analysis of the 1000 Genomes Project Phase 3 SNP dataset using our newly developed method of Fine-Mapping of Adaptive Variation (*FineMAV*). The *FineMAV* score combines population differentiation, derived allele frequency and a measure of molecular functionality to produce a refined list of candidate variants for functional follow-up. We calibrated and tested *FineMAV* using eight 'gold standard' examples of experimentally-validated causal variants underlying adaptations, and were able to pick out the known functional allele in all instances. We used this approach to identify the best candidate variants driving positive selection in Africans, Europeans, East and South Asians, and report many novel examples including rs6048066 in *TGM3* associated with curly hair, and rs7547313 in *SPTA1* associated with erythrocyte shape and possibly malaria resistance in Africans. We also picked up rs201075024 in*PRSS53* associated with hair shape in South Asians. *FineMAV* now offers a better way to identify specific variants for functional follow-up and paves the way for identification of causative alleles driving phenotypic differences among human populations.

659

Using whole genome sequencing of pooled samples to detect local adaptation of teosintes along two altitudinal gradients

Maud I Tenaillon[1], Margaux-Alison Fustier[1], Brandenburg Jean-Tristan[2], Simon Boitard[3], Jason Lapeyronnie[3], Luis E Eguiarte[4], Yves Vigouroux[5], Domenica Manicacci[6]

1. CNRS, Gif-sur-Yvette, FRANCE, France
2. INRA, Gif-sur-Yvette, France
3. INRA, Jouy-en-Josas, France
4. UNAM, Mexico
5. IRD, Montpellier, France
6. Université Paris-Sud, Gif-sur-Yvette, France

We aimed at mining the determinants of local adaptation in the two closest maize wild relatives, the teosintes *Zea mays* ssp *parviglumis* and ssp. *mexicana*. We sequenced 20 individuals from two lowland and two highland populations sampled along two altitudinal gradients, considered as biological replicates. Sequencing of two middle elevation *parviglumis* and *mexicana* populations in one gradient helped controlling for subspecies differentiation. We sequenced 20 individuals from the 6 populations and detected 8,479,581 SNPs. Population differentiation was greater between subspecies than within subspecies. We combined differentiation- and diversity-based methods to detect outlier SNPs that were further tested for correlation of allele frequencies with environmental variables. Outliers defined 43 candidate regions. We modified a haplotype-based method to incorporate genotype uncertainties in haplotype calling. We found haplotype signal consistency in 58% of our candidate regions along with several instances of the same haplotype being selected in lowland or highland populations. We found frequent co-localization between our candidate regions and loci involved in the variation of traits linked to plant-soil interactions indicating that soil is an important factor driving local adaptation in teosintes.

660

Evolutionary history of European Bison (*Bison bonasus*)

Ayla L van Loenen, Bastien Llamas, Julien Soubrier, Holly Heiniger, Emilia Hofman-Kamińska, Rafal Kowalczyk, Alan Cooper

Very little is known about the evolutionary history of the European bison (*Bison bonasus*). The extensive morphological diversity, particularly in body size, skull and horn shape, has been used by palaeontologists to classify more than 50 species and sub-species during the late Pleistocene (120-11ky BP). Many of these forms appear roughly contemporaneously, and no clear pattern of succession can be discerned. Ancient DNA (aDNA) provides a unique opportunity to directly observe genetic evolution by investigating the changes in genetic structure of species and populations in real time. A previous study of the mitochondrial control region of 448 bone samples from Beringia (Russia/Alaska/Canada) revealed a dynamic series of events through time, including range shifts, migrations, and widespread extinctions (Shapiro et al., 2004). Bison are one of the few species to have survived the mass megafaunal extinction during the Pleistocene/Holocene transition (12-9ky BP). Understanding how will provide insight about how large mammals respond to environmental variation and adapt to periods of rapid climate change. Here we describe the evolutionary patterns observed in high-resolution mitochondrial sequencing data—i.e. complete mitochondrial genomes—from a number of ancient European bison samples, specifically patterns of succession of various bison ecomorphs across a broad geographical and temporal range.

661

The factors that constrain or promote the evolution of alternative genome architectures in an RNA virus

Anouk Willemsen[1], Mark P. Zwart[1, 2], Santiago F. Elena[1, 3]

1. Instituto de Biología Molecular y Celular de Plantas (CSIC-UPV), Valencia, VALENCIA, Spain
2. Institute of Theoretical Physics, University of Cologne, Cologne, Germany
3. The Santa Fe Institute, Santa Fe, USA

There are clear differences in genome architecture over organisms, however its evolution is poorly understood. Here, we study three specific processes of the evolution of genome architecture in viruses: (*i*) the reshuffling of existing elements, (*ii*) the decrease of genome complexity through loss of redundant or unnecessary genetic material, and (*iii*) the increase of genome complexity through the acquisition of new genes. We address these topics *in vivo* using a plant RNA virus as a model organism. Important changes in the viral genome were generated followed by experimental evolution to observe how these changes were accommodated. The evolved and ancestral lineages were compared by next-generation sequencing and measurements of virulence, viral accumulation and within-host competitive fitness. First, we identified multiple barriers to the evolution of alternative gene orders. Second, we observed differences in the deletion dynamics of genetically and functionally redundant sequences and we developed a model to predict the stability of gene insertions. Third, we found an exogenous sequence that was evolutionary stable in our model virus genome that does not appear to affect viral fitness and can act as a backup in case of failure of the viral protein responsible for blocking RNAi-mediated plant defenses. Lastly, we observed that a host species jump can be a game changer for evolutionary dynamics, allowing unstable viruses to be competitive in alternative hosts. The results of this study serve as a road map for future research on genome architecture evolution across different viruses as well as different organisms.

662

Genetic origins, population structure and admixture of Xinjiang's Uyghurs

Shuhua Xu[1, 2]

1. Max Planck Independent Research Group on Population Genomics, CAS-MPG Partner Institute for Computational Biology, Shanghai, China
2. School of Life Science and Technology, ShanghaiTech University, Shanghai, China

With a population size more than 10 million, the Uyghur people residing in Xinjiang are believed to be descendants of the most ancient of Turkish tribes with mixed Caucasian and East Asian ancestries. However, their genetic origins, population structure and admixture history remain poorly understood and debatable. Here we systematically assessed genetic diversity and individual ancestry composition of the Xinjiang's Uyghurs (XJU) by genotyping 951 Uyghur individuals, roughly proportional to population size of 13 geographical regions, with high-density single nucleotide polymorphism arrays. We observed a southwest-northeast differentiation within the XJU, which is different from the expected north-south differentiation as separated by Tianshan Mountain. In the context of comparative analyses of 2,477 individuals representing 206 worldwide populations, four major ancestries were identified in XJU without very much variation among individuals, i.e., East Asia, Siberia, West Eurasia, and South Asia. However, XJU showed an overall unique genetic make-up and divergent history from surrounding neighbors including the other modern Turkic speaking populations. The results suggest a long history of population admixture and isolation. Facilitated by new methods including one developed in this study, our analyses shed exciting new light on genetic origins and admixture history of Uyghurs.

663

Genomic divergence between two sympatric sibling species in the mangrove genus *Rhizophora* detected by RAD sequencing

Yubin Yan[1], Mei Sun[1]

1. The University of Hong Kong, Hong Kong, HONG KONG

Uncovering the genomic landscape of divergence in taxa with "porous" species boundaries is of great interest in evolutionary and ecological biology. With the advances of next generation sequencing (NGS), many thousands of single nucleotide polymorphisms (SNPs) can be genotyped for hundreds of individuals in divergent species within reasonable time and budget. This makes the studies of genome-wide divergence feasible in non-model organisms. We investigated the genomic divergence and introgression between *Rhizophora mucronata* and *R. stylosa*, two sympatric mangrove species with detectable ecological divergence and minor morphological differences, using restriction site associated DNA (RAD) sequencing. The reduced genomes of 94 individuals from six populations in three sympatric sites of *R. mucronata* and *R. stylosa* were sequenced using Illumina HiSeq 4000. The short reads were mapped to *R. apiculata* scaffold genome and >42 000 SNPs were discovered. The overall genomic differentiation between *R. mucronata* and *R. stylosa* was substantial in SE Asia, with high F_{ST} at many genomic regions. In contrast, the relative low genomic divergence between sympatric populations in Australia may indicate a lack of reproductive isolation and introgression between the two close-related taxa, or various degree of admixture in ancestry among the individuals at the site. We also detected extensive heterogeneity along the genome in divergence between *R. mucronata* and*R. stylosa*, as some extremely high F_{ST} as well as a few very lows were detected in some genomic regions, which could result from divergent or convergent selection at different loci in the genome.

664

A Functional MiR-124 Binding-Site Polymorphism in IQGAP1 Affects Human Cognitive Performance

Lixin Yang[1], Bing Su[1]

1. Kunming institute of zoology , Chinese Academy of Sciences, Wu Hua, YUN NAN, China

As a product of the unique evolution of the human brain, human cognitive performance is largely a collection of heritable traits. Rather surprisingly, to date there have been no reported cases to highlight genes that underwent adaptive evolution in humans and which carry polymorphisms that have a marked effect on cognitive performance. IQ motif containing GTPase activating protein 1 (IQGAP1), a scaffold protein, affects learning and memory in a dose-dependent manner. Its expression is regulated by miR-124 through the binding sites in the 3'UTR, where a SNP (rs1042538) exists in the core-binding motif. Here we showed that this SNP can influence the miR-target interaction both in vitro and in vivo. Individuals carrying the derived T alleles have higher IQGAP1 expression in the brain as compared to the ancestral A allele carriers. We bserved a significant and male-specific association between rs1042538 and tactile performances in two independent cohorts. Males with the derived allele displayed higher tactual performances as compared to those with the ancestral allele. Furthermore, we found a highly diverged allele-frequency distribution of rs1042538 among world human populations, likely caused by natural selection and/or recent population expansion. These results suggest that current human populations still carry sequence variations that affect cognitive performances and that these genetic variants may likely have been subject to comparatively recent natural selection.

665

Y-chromosome diversity suggests southern origin and Paleolithic backwave migration of Austro- Asiatic speakers from eastern Asia to the Indian subcontinent

Xiaoming Zhang[1], Bing Su[1]

1. Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, Yunnan, China

Analyses of an Asian-specific Y-chromosome lineage **(O2a1-M95)**—the dominant paternal lineage in **Austro-Asiatic (AA)** speaking populations, who are found on both sides of the Bay of Bengal—led to two competing hypothesis of this group's geographic origin and migratory routes. One hypothesis posits the origin of the AA speakers in India and an eastward dispersal to Southeast Asia, while the other places an origin in Southeast Asia with westward dispersal to India. Here, we collected samples of AA-speaking populations from mainland Southeast Asia (MSEA) and southern China, and genotyped **16 Y-STRs of 343 male**s who belong to the O2a1-M95 lineage. Combining our samples with previous data, we analyzed both the Y-chromosome and mtDNA diversities. We generated a comprehensive picture of the O2a1-M95 lineage in Asia. We demonstrated that the O2a1-M95 lineage originated in the southern East Asia among the Daic-speaking populations ~20–40 thousand years ago and then dispersed southward to Southeast Asia after the Last Glacial Maximum before moving westward to the Indian subcontinent. This migration resulted in the current distribution of this Y-chromosome lineage in the AA-speaking populations. Further analysis of mtDNA diversity showed a different pattern, supporting a previously proposed **sex-biased admixture** of the AA-speaking populations in India.

666

Adaptive introgression in agricultural *Campylobacter coli*

Daniel Falush[1], Samuel Sheppard[1], Guillaume Meric[1], Leonardos Mageiros[1], Xavier Didelot[2]

1. University of Swansea, Swansea, United Kingdom

2. Imperial College London, London, UK

Campylobacter jejuni and *Campylobacter coli* are two zoonotic pathogens with a wide host range that differ by about 15% at the nucleotide level. Campylobacter coli is largely restricted to ducks but two lineages have invaded agriculture, namely the ST828 complex and the ST1150 complex which colonize chickens and pigs. Remarkably, both lineages have undergone an enormous adaptive introgression of up to 35% of their genome from Campylobacter jejuni. We use more than 500 ST828 genomes and a large collection of isolates from C. jejuni and non-agricultural C. coli in order to describe the ongoing pattern of introgression.

667

Detection of *Drosophila melanogaster* line admixture through targeted barcoding sequencing

Hangxing Jia[1], Jiawei Huang[1], Wenwen Shi[1], Shengjun Tan[1], Huijing Ma[1], Chenyu Ma[1], Yong E. Zhang[1]

1. Institute of Zoology, Chinese Academy of Sciences, Beijing, China

Drosophila melanogaster is a widely used model organism in many biological studies. Usually *D. melanogaster* inbreeding lines maintained in lab are with a high density, therefore one line might be invaded by other lines. Up till now, there is no effective way to check the line admixture in a large scale. In this study, we serve a new method, Line-seq, to solve this problem by taking advantage of both targeted sequencing and barcoding sequencing techniques. Using Line-seq, we successfully detected line admixture in 23 inbreeding lines and compared the performance of Line-seq and traditional Sanger sequencing in eight artificially admixed samples. Line-seq can detect exogenous DNA admixtures as low as 6%, while Sanger sequencing can do so when the proportion of of exogenous DNA rises up to around 20%. Our results highlight Line-seq is an sensitive, accurate, as well as cost-effective method to detect *D. melanogaster* line admixture.

668

Blood ties: Metabolic convergence among gammaproteobacterial endosymbionts from blood-feeding arthropods and the Mexican leech *Haementeria officinalis*

Alejandro Manzano-Marín[1], Alejandro Oceguera-Figueroa[2], Amparo Latorre[1], Luis F. Jiménez-García[3], Andrés Moya[1]

1. ICBiBE. University of Valencia, Paterna, VALENCIA, Spain

2. Laboratorio de Helmintología, Departamento de Zoología, Instituto de Biología, Universidad Nacional Autónoma de México, Distrito Federal, México

3. Departamento de Biología Celular, Facultad de Ciencias, Universidad Nacional Autónoma de México, Distrito Federal, México

Endosymbiosis between eukaryotic hosts and microorganisms is a common phenomenon in insects, whose typically unbalanced diets are usually complemented by their obligate endosymbionts. While much interest and focus has been directed towards phloem-feeders like aphids and mealybugs, blood-feeders such as the Lone star tick (*Amblyomma americanum*), Glossina flies, and the human body louse (*Pediculus humanus corporis*) also depend on obligate endosymbionts to complement their B-vitamin-deficient diets, and thus are required for growth and survival. Strict blood-feeding glossiphoniid leeches, contrary to the predatory species, have also been found to harbour distinct endosymbionts belonging to the Gamma and Alphaproteobacteria housed in specialised morphologically-diverse organs. The Mexican leech, *Haementeria officinalis* is associated to the obligate endosymbiont *Candidatus* Providencia siddallii (Gammaproteobacteria) . This symbiont resides intracellularly in spherical bacteriomes attached to the oesophagus, and possesses a highly-reduced genome with high A+T content and a reduced set of metabolic capabilities, all of which are common characteristics of ancient obligate endosymbionts of arthropods. Its genome has retained many pathways related to the biosynthesis of B-vitamins, pointing towards a role in supplementing the blood-restricted diet of its host. Through genomic comparison against the endosymbionts of the different blood-feeding arthropods, we were able to detect a high degree of metabolic convergence among these very distantly related endosymbiotic bacteria. These findings strongly support the widespread and conserved metabolic dependence of the strict blood-feeders in their bacterial endosymbionts and the similar constraints these have undergone in their evolution as obligate nutritional partners.

1. Smith TA, Driscoll T, Gillespie JJ, Raghavan R. A Coxiella-like endosymbiont is a potential vitamin source for the Lone Star tick. Genome Biol Evol. 2015;7: 831–8. doi:10.1093/gbe/evv016
2. Akman L, Yamashita A, Watanabe H, Oshima K, Shiba T, Hattori M, et al. Genome sequence of the endocellular obligate symbiont of tsetse flies, Wigglesworthia glossinidia. Nat Genet. 2002;32: 402–7. doi:10.1038/ng986
3. Kirkness EF, Haas BJ, Sun W, Braig HR, Perotti MA, Clark JM, et al. Genome sequences of the human body louse and its primary endosymbiont provide insights into the permanent parasitic lifestyle. Proc Natl Acad Sci U S A. 2010;107: 12168–12173. doi:10.1073/pnas.1003379107
4. Manzano-Marín A, Oceguera-Figueroa A, Latorre A, Jiménez-García LF, Moya A. Solving a Bloody Mess: B-Vitamin Independent Metabolic Convergence among Gammaproteobacterial Obligate Endosymbionts from Blood-Feeding Arthropods and the Leech Haementeria officinalis. Genome Biol Evol. 2015;7: 2871–2884. doi:10.1093/gbe/evv188

669

Using molecular clocks to investigate beneficial (and deleterious) microbe-host interactions in the agroecosystem

Omar Rota-Stabelli[1], Tobias Weil[1], Lino Ometto[1], Andrea Campisano[1], Keivan Karimi[1], Ilaria Pertot[1], Gianfranco Anfora[1]

1. Department of Sustainable Agro-ecosystems and Bioresources, Fondazione Edmund Mach, San Michele all' Adige, Italy

The molecular clock is a powerful technique used to estimate divergence time among organisms using molecules. Although widely used in animal and plant studies, the molecular clock is rarely applied to microbes and microbiomes: while in few cases co-radiation with host can be exploited, calibration of molecules is generally impaired by a lack of fossils and a poor knowledge of generation times outside model organisms. Here we outline, however, how molecular clocks can provide interesting insight into the biology of complex microbe-host interaction within various types of agro-ecosystems. Our case studies include: 1) the concomitant radiation of a phytoplasma with its apple host and its insect vector: a complex partnership further characterized by endosymbionts with putatively protective role against the pathogenic phytoplasma; 2) the origin of a likely beneficial new grapevine endosymbiont whose divergence matches human domestication; 3) the co-radiation of garden strawberry with its main anthracnose fungal endophytic agent. Although methodologically challenging, these examples illustrate that molecular clock is a promising and powerful tool to study the evolution of microbes and microbiomes in the agroecosystems.

670

Exploring the role of segmental duplications in the phenotypic differences between humans and other great apes

Marina Brasó-Vives[1], Diego A Hartasánchez[1], Arcadi Navarro[1, 2, 3, 4]

1. *Institute of Evolutionary Biology (Universitat Pompeu Fabra – CSIC), Barcelona, BARCELONA, Spain*
2. *National Institute for Bioinformatics (INB), Barcelona*
3. *Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona*
4. *Centre for Genomic Regulation (CRG), Barcelona*

Duplicated sequences are one of the main sources of variation in eukaryotic genomes and are known to give rise to new genes and functions. Large (1-200 Kb) and highly identical (≥90%) duplications named segmental duplications (SDs) had a particularly important role in the evolution of African great apes (including humans). SDs compose around 5% of the human genome and are shared with the other African great apes more than what one would expect given their nucleotide divergence. In other words, the rate at which SDs appeared was specially high in the African great ape ancestor, after the split from orangutans. SDs that arose during that period are strong candidates to account for part of the phenotypic differences between these species that point-mutations cannot explain. Here, we use inferences of copy-number along great ape genomes to classify human SDs according to the period of time in which they appeared. We identify, first, human specific SDs, second, human SDs that appeared during the burst of duplications and, finally, older human SDs that are shared with all great apes, including orangutans. We explore the characteristics of these three groups of duplications trying to understand both the causes of the increase in duplication rate during the time of the African great ape ancestor and its phenotypic consequences. We also differentiate between tandem, non-tandem intrachromosomal and interchromosomal SDs. We find differences in length, gene content and Alu content between these groups. These differences point towards different duplication mechanisms of the SDs in these three types of duplications. Moreover, we use sequence similarity inside and outside shared exons in duplications to identify candidate signals of selection in human SDs.

671

GRIDSS: detecting structural variation using positional de Bruijn graphs

Daniel L Cameron[1, 2], Terence P Speed[1, 3], Tony Papenfuss[1, 2]

1. *Walter and Eliza Hall Institute of Medical Research, Parkville, VICTORIA, Australia*
2. *Medical Biology, The University of Melbourne, Parkville, VICTORIA, Australia*
3. *Mathematics and Statistics, The University of Melbourne, Parkville, VICTORIA, Australia*

The identification of genomic rearrangements with high sensitivity and specificity remains a major challenge. Whilst new sequencing technologies enable the detection of a wider range of events, there is still scope for significant improvement of short-read based approaches. Here, we present the Genome Rearrangement IDentification Software Suite (GRIDSS). GRIDSS is composed of an assembler that performs alignment-constrained whole genome breakend assembly using a novel positional de Bruijn graph algorithm and a probabilistic structural variant caller that combines assembly, split read, and read pair evidence in a unified variant scoring model.

Our novel assembly approach identifies breakend contigs, that is, contigs assembled from a single side of a breakpoint. By incorporating positional information into the assembly graph GRIDSS performs whole genome assembly of all contigs, with no separation into windows or prior identification of candidate regions required. Although this approach results in an assembly graph that is approximately 100 times larger than the equivalent whole genome de novo de Bruijn graph, our method assembles a 50x WGS data set in less than 4 CPU hours using 16GB of memory.

GRIDSS achieves high sensitivity and specificity on cell line and patient tumour datasets. Results on well-characterised Genome in a Bottle data demonstrate improved sensitivity whilst retaining a false discovery rate less than half that of other recent methods. GRIDSS detects micro-homologies and non-templated sequence insertions at the breakpoint, and can perform combined variant discovery on multiple related samples and population data. GRIDSS is freely available at https://github.com/PapenfussLab/gridss.

672

Segregation of chromosomes during the meiosis of pentaploid in yeast

Ying Chen[1]

1. *Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Chaoyang, China*

While it is generally accepted that triploid organisms are sterile, the exact reason of it remains elusive. With six copies of each homolog being segregated into four meiotic products, meiosis in triploid causes aneuploid gametes with reduced viability. However, it is unclear that if there are additional processes during meiosis also contribute to the sterility of triploid organisms. To examine the segregation behavior in details and find phenomena that have not been observed in previous studies of triploid, we constructed a pentaploid strain of *Saccharomyces cerevisiae*, so that each spore has at least one copy of each chromosome even if trivalent paring occurs and all three chromosomes segregate into one pole. Using single-nucleotide polymorphisms among strains, we generated genome-wide maps of crossover of all four products derived from eleven meioses. Among these 176 homolog pairing and segregation events, we observed 35 trivalent pairings and 10 bivalent/univalent pairings. The karyotype patterns indicated that following bivalent/univalent pairings, two combined homologs segregate to opposite poles while the third homolog randomly segregates to either pole, and following trivalent pairings, all three homologs randomly segregate to either pole rather than to the only one pole. In addition, high-resolution recombination maps revealed numerous chromosome breakage events, resulting in partial instead of the complete chromosomes in the meiotic products. Remarkably, the terminal region of all these fragments are long terminal repeat (LTR) sequence of genome, suggesting that chromosome breakage happens more easily in these regions. Taken together, we studied meiotic chromosome pairing and segregation patterns in pentaploidy, demonstrated the prevalence of multivalent paring during meiosis, and suggested the potential role of LTR in chromosome rearrangement that occurred frequently during the genomic evolution of yeast.

673

Structural variation in genome could make novel complex trait in mouse behavior

Maryam Keshavarz[1], Diethard Tautz[1]

1. *max planck institute for evolutionary Biology, Plön, SCHLESWIG - HOLSTEIN, Germany*

Throughout evolutionary history, organisms have evolved a variety of sophisticated novel traits for survival and reproduction. One of these novel traits that recently have been known is house mouse mate choice preference that strongly influenced by their father origin; however the molecular mechanism of this complex trait remains unknown so far. Here for the first time we showed how structural variation in genome could lead to a novel trait during mouse evolution. Genome and transcriptome sequencing data analysis of ten different mouse populations revealed Peg13 (a paternal gene) on chromosome 15 and an imprinted cluster in chromosome 7 which has paternal biased expression have been highly differentiated between

mouse populations. Data from transcriptome assembly showed different number of paternal new gens in every population in chromosome 7 which only express in brain and all of them were classified as long non coding RNA by lncseeker program and then confirmed by ribosomal profiling data analysis. Interestingly we also detected two repeated clusters which compare to other mammals is two-five times more expanded in mouse. Our small RNA sequencing data revealed that these repeated clusters are host for tow paternal snoRNA families which highly duplicated in mouse and interestingly rate of snoRNA duplication for both families is different from one population to another. Further analysis on snoRNA sequences showed although members of each snoRNA family in each population are differentiated from each other, still they are much closer to each other compare to members of same family in other populations and showed a nice concerted evolution manner. Since these regions in brain have known role in speech and cognition in human and also regulate ultrasonic vocalization in mouse, we assume that these specific paternal structural variations in mouse could be responsible for their paternal mate choice preference and subsequently their fast evolution

674

Systems Genetics identifies two structural variants of P450 genes that confer resistance to insecticides in *Drosophila melanogaster*.

Paul Battlay[1], Llewellyn Green[1], Alexandre Fournier-Level[1], Charles Robin[1]

1. University of Melbourne, Parkville, Vic, Australia

The *Drosophila melanogaster* Genetic Reference Panel, consisting of 205 inbred, sequenced lines, is a powerful tool for elucidating the genetic basis of phenotypic variation in *Drosophila melanogaster* through genome-wide association studies. The recent addition to of 185 sequenced transcriptomes has expanded the utility of this resource to allow the incorporation of genotype, phenotype, and gene expression into a single model. By applying this 'systems genetics' approach to insecticide resistance phenotypes we identified *Cyp6g1* and *Cyp12d1* as top candidate resistance loci for the insecticides azinphos-methyl and chlorantraniliprole respectively. We found, in each of these genes, that expression level was correlated with structural variation, and validated the involvement of increased expression in resistance transgenically. These findings highlight the transcriptomic and phenotypic relevance of structural variation, which is becoming more apparent as we expand beyond SNP-based association studies.

675

LINE-1-like retrotransposons contribute to RNA-based gene duplication in dicots

Shengjun Tan[1], Yong Zhang[1]

1. Institute of Zoology, Chinese Academy of Sciences, Chaoyang District, BEIJING, China

RNA-based duplicated genes or functional retrocopies (retrogenes) are known to drive phenotypic evolution. Retrogenes emerge via retroposition, which is mainly mediated by long interspersed nuclear element 1 (LINE-1 or L1) retrotransposons in mammals. By contrast, long terminal repeat (LTR) retrotransposons appear to be the major player in plants, although an L1-like mechanism has also been hypothesized to be involved in retroposition. We tested this hypothesis by searching for young retrocopies, as these still retain the sequence features associated with the underlying retroposition mechanism. Specifically, we identified polymorphic retrocopies by analyzing public Arabidopsis (*Arabidopsis thaliana*) resequencing data. Furthermore, we searched for recently originated retrocopies encoded by the reference genome of Arabidopsis and *Manihot esculenta*. Across these two datasets, we found cases with L1-like hallmarks, namely, the expected target site sequence, a polyA tail and target site duplications. Such data suggest that an L1-like mechanism could operate in plants, especially dicots.

676

The Recruitment of Proteins into a Scorpion Venom

Galit Blecher[1], Oksana Daron[1], Noam Zilberberg[1]

1. Ben Gurion University in the Negev, Beer Sheva, Israel

Scorpion venoms are composed of a vast variety of bioactive peptides including neurotoxins that target receptors and ion channels, enzymes, and peptides with cytolytic activity. The order *Scorpiones* constitute one of the most ancient groups of animals on earth (>400 million years), with ~1,700 known species that can be divided into four groups. The venoms of scorpions from the most studied group, the *Buthida*, are a rich source for these bioactive peptides. Most of the venom peptides are small (23-78 amino acids-long), are well-packed by several disulfide bridges, and affect ion channel function in excitable and non-excitable tissues.

Here we report the identification of venom gland-specific transcripts from the Israeli scorpion *Buthacus leptochelys*. By using next generation RNA-seq, we have generated transcriptomes for two different abdomen segments: one with (telson) and one without the venom glands. Out of ~70,000 contigs in the telson transcriptome, 270 coded for venom-selective secreted proteins, out of which 110 were homologous to known toxins. An additional 360 contigs coded for paralogs of the venom-selective proteins but were not selectively expressed in the venom glands. Based on the two data sets, we were able to distinguish between venom gland-expressed and venom gland-selective proteins. Furthermore, most of the venom gland-selective proteins were found to have paralogs, which are not specifically expressed within the venom gland, as well as orthologs. This information will help us to learn about the process of protein recruitment into the venom gland and subsequently the evolution of several toxin families.

677

Comparative proteomic and transcriptomic analysis of venom producing posterior salivary glands of the blue ringed octopus (*Hapalochlaena maculosa*) and the southern sand octopus (*Octopus kaurna*).

Brooke Whitelaw[1], Jan Strugnell[1], Pierre Faou[2], Rute Fonseca[3], Nathan Hall[2, 4], Mark Norman[5], Julian Finn[5], Ira Cooke[6]

1. Department of Ecology, Environment and Evolution, La Trobe University, Melbourne, Victoria, Australia

2. Department of Biochemistry, La Trobe University, Melbourne, Victoria, Australia

3. The Bioinformatics Centre, Department of Biology, University of Copenhagen, Copenhagen, Denmark

4. Agribio, La Trobe University, Melbourne, Victoria, Australia

5. Sciences, Museum Victoria, Melbourne, Victoria, Australia

6. James Cook University, Townsville, QUEENSLAND, Australia

The venom secretions used by octopods to immobilise and kill prey have been studied for over a century. Until very recently however, the molecular composition of these venoms was almost entirely unknown. Recent studies have used high throughput molecular techniques such as transcriptome sequencing to propose a wide variety of proteins that may play a role in octopod venom. In this work we used a combination of transcriptomic sequencing and mass spectrometry based proteomics to directly identify proteins from the posterior salivary glands of two octopod species, *H. maculosa* and *O. kaurna*. While both species are thought to have similar dietary preferences, the venom of *H. maculosa* is unique in that it is known to contain a potent non-proteinaceous neurotoxin (Tetrodotoxin) produced by bacteria.

Our analysis showed that the salivary gland proteomes of the two species were dominated by a similar suite of molecules with serine proteases being particularly diverse and abundant. Many of the most abundant proteins were shared between both species (orthologs) while 12 of the top 20 proteins across both species were not homologous to any proteins on the NCBI nr database by BLAST search. Among less abundant proteins we also identified representatives from nine other families of proteins that have been recruited to toxic function in other taxa. We examined the relative abundance and diversity of these between the two species as well as the molecular evolution of expanded protein families such as serine proteases.

678

Intragenome diversity of gene families encoding toxin-like proteins in venomous animals

Ricardo C Rodríguez de la Vega[1], Tatiana Giraud[1]

1. Ecologie Systematique et Evolution, Univ. Paris-Sud, CNRS, AgroParisTech, Université Paris-Saclay, Orsay, France

The remarkable diversity of action and specificity exerted by animal venoms is tightly linked to the prevalence of the different component toxins. The evolution of venoms is thus the story of how toxins arise and the processes that generate and maintain their diversity. For animal venoms these processes include recruitment for expression in the venom gland, neofunctionalization, paralogous expansions and functional divergence[1]. The systematic study of these processes requires the reliable identification the venom components involved in antagonistic interactions. While -omic approaches have the potential of uncovering the entire set of toxins in a given organism, the existence of non-venom toxin homologs and the misleading effect of partial census of the intragenome molecular diversity of toxins make necessary to collect complementary evidence to distinguish true toxins from their non-venom homologs. Here we analyzed the whole genomes of two scorpions, one spider and one snake aiming at the identification of the full repertoire of toxin-like protein coding genes. We classified the entire set protein coding genes into paralogous groups and monotypic genes, identified the toxin-like protein coding genes based on known toxin families, and quantified their expression in both venom-glands and pooled tissues. Our results confirm that toxin-like protein coding genes in a range of venomous animals are part of multigene families originated by recruitment events from non-toxin paralogs and followed by expansions of the toxin-like protein coding genes. However, we also show that failing to account for sequence similarity with non-toxin proteins has a considerable misleading effect that can be greatly reduced by comparative transcriptomics.

Acknowledgments: Research funded by a Marie Skowldowska-Curie / Pres-SUD grant (Cofund No. 246556). This contribution greatly benefited from discussions carried out at a NESCent catalysis meeting and the SICB's symposium on Integrative and Comparative Biology of Venom organized by M. Daly and L. Gibbs.

1. Annual Review of Genomics and Human Genetics 10:438-511, 2009, http://dx.doi.org/10.1146/annurev.genom.9.081307.164356

679

A combined transcriptomic and proteomic approach reveals putative toxins in the slime secretions of the southern bottletail squid, *Sepiadarium austrinum* (Cephalopoda)

Nikeisha J Caruana[1], Ira R Cooke[2], Pierre Faou[1], Julian Finn[3], Nathan E Hall[1], Mark D Norman[3], Sandy Pineda[4], Jan M Strugnell[1]

1. La Trobe University, Melbourne, VIC, Australia

2. James Cook University, Townsville, Queensland, Australia

3. Museum Victoria, Melbourne, Vic, Australia

4. Institute for Molecular Bioscience, The University of Queensland, Brisbane, Qld, Australia

Sepiadarium austrinum, the southern bottletail squid, is a small squid that inhabits sediments along Australia's south-east coast. When provoked, it rapidly secretes large volumes of slime, and field observations suggest that this is toxic to crabs. This study provides the first proteomic analysis of a slime secretion from a cephalopod and the first investigation of a member of the family Sepiadariidae using proteomic methods. The proteomic composition of this slime was analyzed using a combination of tandem mass spectrometry and transcriptomics and found that it was remarkably complex with 1735 identified protein groups (FDR: 0.01). Of these, 15 were identified as putative toxins including three short (80-130AA) cysteine rich secreted proteins with no homology to proteins on the NCBI or UniProt databases. Our *S. austrinum* protein database of 40,475 proteins was created from the transcriptome by combining predictions from TransDecoder software with an additional 84 novel proteins identified by proteogenomics. This last step proved crucial for the identification of toxin-like proteins with the most abundant protein in slime (also toxin-like) being identified through this method. Our study highlights the importance of proteomics in toxin discovery by using direct proteomic measurement within a toxic secretion (slime) rather than its parent gland.

680

Expression patterns of cnidarian toxins reveal dynamic gene family evolution and regulation

Joachim M Surm[1, 2], Peter J Prentis[3, 4], Ana Pavasovic[1, 2]

1. School of Biomedical Sciences, Faculty of Health, Queensland University of Technology, Brisbane, QLD, Australia

2. Institute of Health and Biomedical Innovation, Queensland University of Technology, Brisbane, QLD, Australia

3. School of Earth, Environmental and Biological Sciences, Science and Engineering Faculty, Queensland University of Technology, Brisbane, QLD, Australia

4. Institute for Future Environments, Queensland University of Technology, Brisbane, QLD, Australia

Cnidarians are the oldest venomous lineage, characterised by a diverse range of proteinaceous toxins found in nematocysts (stinging cells). While the distribution of some gene families that encode toxin proteins have been studied in detail, we know very little about the expression patterns of toxin genes across different tissues within an organism. In this study we examined expression patterns of toxin and toxin-like gene families across three tissue types in the sea anemone *Actinia tenebrosa.* Tissues were selected as they contain the highest density of nematocysts and included acrorhagi (ring of modified tentacles used in intraspecific aggressive encounters), tentacle (epithelial projections apical aspect and used in predation and defence) and mesenteric filaments (multifunctional morphological structures principally used in digestion). A fully replicated RNA-Seq experiment was performed across the three tissues isolated from nine individuals (three replicate pools of three individuals for each tissue type). Sequencing generated at least 80 million paired-end reads per replicate, while the bioinformatic analysis pipeline consisted of *de novo* assembly followed by remapping of individual samples to quantify gene expression patterns across the tissue types. Overall we found 24,449 differentially expressed transcripts across the three tissue types of which toxin genes localised to nematocysts were significantly over-represented. Over 200 toxin and toxin-like genes, both widespread and cnidarian specific, were differentially expressed. These results show that the expression profiles of toxin and toxin-like genes are unique to specific tissue types and possibly relate to their different roles in attack, defence and prey capture.

682

Lokiarchaea is not the missing link between Archaea and Eukaryotes.

Violette Da Cunha[1, 2], Morgan Gaia[1], Arshan Nasir[3], Patrick Forterre[1, 2]

1. Microbiology Departement, Institut Pasteur, Paris, FRANCE, France

2. Institute for Integrative Biology of the Cell (I2BC), CEA, CNRS, Univ. Paris-Sud, Université Paris-Saclay, , Gif-sur-Yvette, France

3. Departments of Biosciences, COMSATS Institute of Information Technology, Islamabad, Pakistan

The topology of the tree of Life, and particularly the evolutionary relationship between Archaea and Eukarya, is a major biological issue, with presently two conflicting hypotheses. In the first one, Archaea and Eukarya are sister groups[1,2], whereas in the other, Eukarya emerged within Archaea[3]. The latter hypothesis has been boosted by the reconstruction, from metagenomic data, of three partial genomes from members of a possible new archaeal phylum, Lokiarchaeota. In a phylogenetic tree constructed from the concatenation of 36 universal proteins, Eukarya was shown branching within Lokiarchaeota[4]. The lokiarchaeota were rapidly presented as the “missing link” between “Prokaryotes” (Archaea) and Eukarya, and a definitive argument supporting the archaeal ancestor scenario for eukaryogenesis[3]. In reanalyzing the data, we observed that most lokiarchaeal universal proteins branch within Archaea. In one of the minority proteins that branch as sister group to Eukarya, we observed several specific insertions shared between the Loki protein and specific eukaryotic proteins. This suggests that the position of the lokiarchaea in the original analysis is probably biased by reconstructions problems. We proposed a new position for the *bona fide* Lokiarchaea, based on a robust phylogenetic analysis of the two large RNA polymerase subunits concatenated, and supported by several lines of evidence. Our results clearly show that the Lokiarchaeon is not more the missing link to Eukarya, and supports the classical Woese tree of Life. This shift from one scenario to its opposite clearly shows that the process of eukaryogenesis is still open for debate.

1. Woese, C. R. Interpreting the universal phylogenetic tree. Proc. Natl. Acad. Sci. U. S. A. 97, 8392–8396 (2000).
2. Forterre, P. The universal tree of life: an update. Front. Microbiol. 6, (2015).
3. Guy, L. & Ettema, T. J. G. The archaeal ‘TACK’ superphylum and the origin of eukaryotes. Trends Microbiol. 19, 580–587 (2011).
4. Spang, A. et al. Complex archaea that bridge the gap between prokaryotes and eukaryotes. Nature 521, 173–179 (2015).

683

Metagenomics exploration of novel archaea sheds new light on the early evolution of eukaryotes

Eva F. Caceres[1], Katarzyna Zaremba-Niedzwiedzka[1], Jimmy H. Saw[1], Disa Bäckström[1], Lina Juzokaite[1], Emmelien Vancaester[1], Kiley W. Seitz[2], Karthik Anantharaman[3], Kasper U. Kjeldsen[4], Matthew B. Stott[5], Takuro Nunoura[6], Jillian F. Banfield[3], Andreas Schramm[4], Brett J. Baker[2], Anja Spang[1], Thijs J.G. Ettema[1]

1. Department of Cell- and Molecular Biology, Science for Life Laboratory, Uppsala University, Uppsala, Sweden

2. Department of Marine Science, University of Texas-Austin, Marine Science Institute, Port Aransas, USA

3. Department of Earth and Planetary Sciences, and Department of Environmental Science, Policy, and Management, University of California, Berkeley, CA, USA

4. Section for Microbiology & Center for Geomicrobiology, Department of Bioscience, Aarhus University, Aarhus, Denmark

5. GNS Science, Extremophile Research Group, Private Bag 2000, Taupō, New Zealand

6. Research and Development Center for Marine Biosciences, Japan Agency for Marine-Earth Science & Technology, Yokosuka , Japan

The origin of eukaryotes and their cellular complexity remains one of the main enigmas in biology. It is generally accepted that eukaryotes emerged from an endosymbiosis event between an archaeal host and a bacterial symbiont. Although the identity of the host cell remains an ongoing matter of debate, a recent study has shown that it seems to be most closely related to Lokiarcheaota, a new phylum of uncultivated archaea [1]. Lokiarchaeota not only forms a monophyletic group with Eukaryotes in phylogenetic analyses but also harbors a large amount of eukaryotic signature proteins, suggesting that the genetic repertoire of the putative archaeal ancestor was more complex than previously thought [1].

To be able to shed further light onto the potential archaeal ancestry of eukaryotes, we are using cultivation-independent approaches to reconstruct genomes of archaeal lineages related to Lokiarchaeota from all around the world. Preliminary results of phylogenetic and comparative genomic analyses of these novel organisms support the emergence of Eukaryotes from within the archaeal domain, i.e. from Lokiarchaeota and related lineages and reveal the presence of previously detected as well as novel eukaryotic signature proteins. Intriguingly, these proteins are involved in processes such as ubiquitin modifier systems and ESCRT- and trafficking machineries in Eukaryotes or represent major components of eukaryotic cytoskeletons. Altogether, these findings reveal fascinating novel insights into the genetic potential of the archaeal domain and the origin of eukaryotes.

1. Spang, A. et al. Complex archaea that bridge the gap between prokaryotes and eukaryotes. Nature 521, 173-179, doi:10.1038/nature14447 (2015).

684

Rings Reconcile Genotypic and Phenotypic Evolution within the Proteobacteria

James A Lake[1], Joseph Larsen[2], Brooke Sarna[1], Rafael R de la Haba[3], Yiyi Pu[4], HyunMin Koo[5], Jun Zhao[6], Janet S Sinsheimer[7]

1. MCDB, UCLA, Los Angeles, California, United States

2. UCLA, Castaic, CALIFORNIA, United States

3. University of Sevilla, Sevilla, Spain

4. Zhejiang University, Hangzhou, Zhejiang, China

5. University of Alabama, Birmingham, Birmingham, Alabama, United States

6. Peking University, Haidian, Beijing, China

7. Biomathematics, UCLA, Los Angeles, California, United States

Phylogenetic trees are the traditional way to represent evolutionary relationships. However, they fail to adequately represent species evolution even with the sequencing of whole genomes. To address this issue we utilized a novel modeling technique, the phylogenetic rings and applied our methods to the relationships among the proteobacteria, which is the most speciose proteolytic phylum known in science, containing free living, pathogenic, photosynthetic, sulfur metabolizing, and symbiotic species. We develop new rooted ring analyses and studied proteobacterial evolution using protein family data and outgroup rooting procedures. We discover and map the origins of significant gene flows in the rooted proteobacterial rings, and recognized that the evolution of "Alpha-","Beta-", and "Gamma"-proteobacteria is represented by unique set of rings. Our analyses utilize gene presences and absences to determine complex gene flows. Thus, unlike trees which only show branching arrangements, our graphs depict the complex flow of genes producing phenotypes. Through our analyses we are able to identify the gene flows that led to photosynthesis in Alpha-, Beta-, and Gamma-proteobacteria from the common ancestor of the Actinobacteria and Firmicutes. From our study of the rooted rings of proteobacteria we find consistency with the observed genotypic and phenotypic relationships observed among the various proteobacterial classes. Ring phylogenies can explain the evolution of both genotypes and phenotypes of biological processes in robust and complex groups such as proteobacteria.

685

Transcriptional variation in the protist *Trichomonas vaginalis* caused by insertion of DNA transposable elements

Martina Bradic[1], Sally D. Warring [1], Jonathan Bermeo[1], Jane Carlton[1]

1. Center for Genomics and Systems Biology, Department of Biology, New York University, New York , NY, United States

Transposable elements (TEs) comprise a large portion of eukaryotic genomes. Their capacity to jump within and between genomes contributes to genome plasticity and genetic diversity. However, their role in transcriptional regulation has been less studied, primarily because of the notion that TEs represent "junk". Understanding TE insertion characteristics, and how their introduction and expansion is controlled, represent central questions in evolutionary biology. *Trichomonas vaginalis*, a haploid protist whose genome is composed of ~60% TEs provides an excellent system to address these questions. The majority of TEs in this microbe are DNA transposons, including a family of ~1,000 *Mariner* elements and ~3,000 *Maverick* elements, which use a cut-and paste mechanism to jump. First we undertook a pilot study to determine *in silico* insertion preference for these two TE families in a reference genome assembly of *T. vaginalis*, and found that the interruption of gene expression is inversely correlated with distance of TEs from genes. Surface protein gene families, such as the BspA family, were most frequently disrupted by TE insertions, which might be related to the parasite's plasticity to establish or modulate infection. Next we looked at insertions across the whole genome using "Transposon Display-Seq" and RNA-Seq of 17 geographically and phenotypically distinct *T. vaginalis* isolates, and identified a total of 1,411 *Maverick* and 1,385 *Mariner* insertions. Approximately 1.7% of *Maverick* and 0.5% of *Mariner* insertions were fixed in all samples, some of which showed abolished gene expression in all isolates. Our study represents the first genome-wide analysis of the impact of TE insertions on the *T. vaginalis* transcriptional landscape and shows that TEs play an important role in gene regulation of this parasite.

686

Should we reconsider the chromosomal gene movement of retrocopies?

Joanna Ciomborowska[1], Michal Kabza[1], Izabela Makalowska[1]

1. Faculty of Biology, Department of Bioinformatics, Adam Mickiewicz University in Poznan, Poznan, POZNAN, Poland

One of most common hallmarks of retrocopies (usually intronless RNA-based duplicates created from multi-exon parental genes through reverse transcription) is chromosomal gene movement, so situation in which retrocopy is located in a completely different genomic surrounding than a source gene. Number of previous studies showed a strong tendency for retrocopies to „escape" from chromosome X after retroposition, probably in order to avoid meiotic sex chromosome inactivation.

Here we examined, using bioinformatics and statistics methods, movement pattern for retrocopies in 15 animal genomes collected in RetrogeneDB (http://retrogenedb.amu.edu.pl). However, we used slightly different approach to these analyses because we decided to follow changes in localisation for each chromosome separately, not only for basic groups (all autosomes vs. chromosome X etc.). What is more, we analysed the pattern for all, as well as for expressed retrocopies only (according to our RNA-seq data analysis).

Our findings showed that chromosome X is on the top of the list of donor but for expressed retrocopies only for five species, while for majority of organisms retrocopies originated from one of the autosomes (i.e. chromosome 13 for human, 11 for marmoset, 6 for gorilla or 19 for chimp). Considering chromosomes that accept all retrocopies, for twelve species chromosome X presents maximum excess value, so there is visible trend for moving from automosome to sex chromosome. Only for human (chr 19), marmoset (chr 18), gorilla (chr 19) autosomes tend to gain more retrosequences. For expressed retrocopies, observed proportions are equal. We also tried to find if there are any factors describing those chromosomes served as donors or acceptors, especially for human.

Summing up, thanks to our approach, we can follow movement of retrocopies in details and observe differences related to expression pattern. Our results present an interesting perspective and will help to understand retroposition and it's evolutionary consequences better.

687

Resolving kangaroo phylogeny and overcoming retroposon ascertainment bias

William G Dodt[2, 1], Susanne Gallus[3], Axel Janke[3], Matthew J Phillips[2, 1], Maria Nilsson[3]

1. Queensland University of Technology, Brisbane

2. Queensland University of Technology, Wavell Heights, QUEENSLAND, Australia

3. Senckenberg Biodiversity and Research Centre, Frankfurt

The genus *Macropus* contains many of the most recognizable kangaroo species, which include the largest living marsupials. Despite being a well-studied group, the phylogenetic relationships within this genus remain poorly resolved. With the development of next generation sequencing, it has become possible to investigate phylogenetic relationships using genome level characters. I will discuss the use of retrotransposons as phylogenetic markers, with a focus on kangaroo evolution. A particular class of retrotransposon – an endogenous retrovirus - has been prolific during the evolution of kangaroos. We have utilized presence/absence information of retrotransposons to shed light on the phylogenetic relationships among members of the genus *Macropus,* and close relatives, and address the statistical support for retroposon analyses when only a single reference genome is available.

688

High similarity between distantly related species of a plant SINE family is consistent with a scenario of vertical transmission without horizontal transfer.

Jeffrey Fawcett[1], Hideki Innan[1]

1. SOKENDAI (The Graduate University for Advanced Studies), Hayama, KANAGAWA, Japan

Many transposable element (TE) families show surprisingly high levels of similarity between distantly related species. How can we explain such observations? One possibility is frequent horizontal transfers of TEs, which is often at least partly based on our intuition that TE sequences between such distantly related species “should not be that similar”. However, do we really know how similar the TE sequences should be? In this study, based on careful comparative genomic analyses, we reconstructed the evolutionary history of a particular TE family in plants called the Au SINE. Our results suggest that the Au SINE originated >150 million years ago (mya) in the common ancestor of all angiosperms, and retained ~80% nucleotide similarity between many plant species that diverged >100 mya, probably because maintaining their sequences was important for their survival. Thus, even if TE sequences between certain species may seem “too similar”, a simple model of vertical transmission without any horizontal transfers may sometimes provide a sufficient explanation.

689

Identification of polymorphic L1 insertions in mice

Patricia Gerdes[1], Sandra R. Richardson[1], Geoffrey J. Faulkner[1, 2]

1. Mater Research Institute - University of Queensland, Woolloongabba, QLD, Australia

2. School of Biomedical Sciences, University of Queensland, St. Lucia, QLD, Australia

Long interspersed element 1 (L1 or LINE-1) retrotransposons are mobile genetic elements that comprise 17% of the human and 18% of the mouse genome. L1 mobilizes, or retrotransposes, via the reverse transcription and integration of an RNA intermediate. In mice, ~3000 copies per individual are potentially active and belong to three subfamilies (TF, GF and A), whereas the human genome harbours only ~80-100 potentially active copies. L1 mobilization has to occur in the germline or pluripotent cells of the early embryo prior to germline specification to be transmissible and have an ongoing impact on genome evolution. Importantly, L1 insertions contribute to genomic diversity thereby creating variation among individuals of a species but insertions within and proximal to genes can disrupt gene function and cause genetic disease.

Here, we used mouse retrotransposon capture sequencing (mRC-seq) to identify endogenous retrotransposition events in pedigrees of C57BL/6J mice. We identified 35 polymorphic insertions including L1 TF and GF subfamilies as well as B1 and B2 short interspersed elements (SINEs) that appeared throughout our pedigrees but were absent from the reference genome. 18/35 polymorphic insertions were varying in presence among our animals. 7 out of these 18 insertions were validated by PCR and fully characterized identifying target site duplications (TSDs) and other structural hallmarks of target-site primed reverse transcription (TPRT). In addition, 3’ transductions on two polymorphic L1 insertions allows identification of the progenitor L1 element. Two L1 insertions differentially present within our mouse pedigrees inserted into introns of genes, allowing the opportunity to study the functional impact of retroelement polymorphisms in mammals. Together, these studies will elucidate the ways in which retrotransposon activity can impact the genomic landscape of a species.

690

Phylogenetic and genomic analyses resolve the origin of important plant genes derived from transposable elements

Zoé Joly-Lopez[1], Douglas R Hoen[2], Mathieu Blanchette[3], Thomas E Bureau[2]

1. Center For Genomics And Systems Biology, New York University, New York, NY, United States

2. Biology, McGill University, Montreal, Qc, Canada

3. School of Computer Sciences, McGill University, Montreal, Qc, Canada

Once perceived as merely selfish, transposable elements (TEs) are now recognized as potent agents of adaptation. One way TEs contribute to evolution is through TE exaptation (also referred to as co-option or molecular domestication), a process whereby TEs, which persist by replicating in the genome, transform into novel host genes, and persist by conferring phenotypic benefits. In eukaryotes, TE exaptation has made possible major evolutionary innovations, including the vertebrate adaptive immune system and the mammalian placenta; yet little is known about this process. To better understand TE exaptation, we designed an approach to resolve the phylogenetic context and timing of exaptation events and subsequent patterns of exapted TE (ETE) diversification. Starting with known ETEs, we search in diverse genomes for basal ETEs and closely-related TEs, carefully curate the numerous candidate sequences, and infer detailed phylogenies. To distinguish TEs from ETEs, we also weigh several key genomic characteristics including repetitiveness, terminal repeats, pseudogenic features, and conserved domains. Applying this approach to the well-characterized plant ETEs, MUSTANG (MUG) and FAR-RED ELONGATED HYPOCOTYL3 (FHY3), we show that each group is paraphyletic and we argue that this pattern demonstrates that each originated in not one but multiple exaptation events. These exaptations and subsequent ETE diversification occurred throughout angiosperm evolution including the crown group expansion, the angiosperm radiation, and the primitive evolution of angiosperms. In addition, we detect evidence of several putative novel ETE families. Our findings support the hypothesis

that TE exaptation generates novel genes more frequently than is currently thought, often coinciding with key periods of evolution. Our approach is not limited to plant genomes and would be informative to apply, for example, in mammals.

691

Applying complementary NGS approaches for *in silico* resolution of the present and past mobilome of *Clostridium difficile* R078 isolates

Bethany R Jose[1, 2]

1. University of Canterbury, Christchurch, New Zealand

2. Department of Molecular and Cell Biology, University of Leicester, Leicester, Leicestershire, United Kingdom

The human pathogen *Clostridium difficile* is responsible for many severe cases of hospital-acquired diarrhoea, and the many mobile genetic elements (MGEs) in the *C. difficile* genome contribute to its virulence through the transfer of toxin and antibiotic resistance genes. Resultant repeat-rich regions makes it difficult to resolve these elements and generate high quality draft genomes with short read NGS technologies.

Using Illumina reads from a previous sequencing project, an assembly strategy was developed to create an improved draft genome of the *C. difficile* R078 estuarine isolate CD105HS26. The newly SMRT-sequenced genomes of the clinical reference strain M120 and estuarine isolate CD105HS27 (both R078) were used as references to resolve difficult sequence regions. Annotation of the draft assembly identified MGE content, and showed improved resolution of repeat regions and pathogenicity-related genes. CD105HS26 was found to contain both transposons present in M120 and a unique transposon-like element from CD105HS27, which could be partly resolved using reference sequences during assembly.

The CRISPR/cas system provides adaptive immunity against bacteriophage infections by storing viral sequences as spacers. Spacers from the R078 CRISPR arrays were searched for identical matches to *C. difficile* phages, genomes and plasmids to determine phage resistance. Non-identical matches were used to assess potential phage evolution, predict the existence of novel phages and establish a potential host range of existing uncharacterised phages. This analysis indicates the R078 isolates have high resistance to bacteriophage infection and shows their CRISPR arrays acquire new spacers slowly, suggesting CRISPR spacer analysis could be used for strain typing.

692

Bioinformatic Analysis of Expression and Regulation of Human Antisense transcripts by Transposable elements in Human Full-Length cDNA Sequences

Jaeeun Jung[1, 2], **Kunhyang Park**[3], **Chang hyeon Park**[1], **Min ah Jung**[1, 2], **Sugi Lee**[1, 2], **NamSoon Kim**[1, 2], **Dae-Soo Kim**[1, 2]

1. Rare Disease Research Center, Korea Research Institute of Bioscience & Biotechnology, Youseong-gu, Daejeon, South Korea

2. University of science and technology, Youseong-gu, Daejeon, South Korea

3. Core Facility Management Research Center, Korea Research Institute of Bioscience & Biotechnology, Youseong-gu, Daejeon, South Korea

Genetic complexity of an organism studies aimed at defining the structure of gene and determining the processes responsible for how gene expression at the transcripts and protein levels are regulated have led to regulatory mechanism. And also, insertion of transposable element into the promoter region could affect the transcription of cellular genes. The main aim of our study was to find antisense transcripts in human genome that derived from the transposable elements. Using a bioinformatics approach, we searched the antisense transcripts derived from the transposable elements using the full-length cDNA sequence within human genome. In this study, we established a set of very stringent criteria to identify the correct orientation of each transcript. From our *in silico* analysis of human genome indicated that 423 antisense transcripts gene pair was identified to have been affected by transposable elements during the cellular gene expression. The large number of sense-antisense transcripts suggests that gene regulation by antisense transcripts derived from the transposable elements in human genome. Antisense transcripts regulate of transposable elements may also provide understanding of the complex regulation networks and dynamic evolutionary features during human evolution.

693

Macaca specific exon creation event generates a novel *ZKSCAN5* transcript

Young-Hyun Kim[1, 2], **Se-Hee Choe**[1, 2], **Ja-Rang Lee**[1], **Sang-Je Park**[1], **Sang-Rae Lee**[1, 2], **Jae-Won Huh**[1, 2], **Kyu-Tae Chang**[1, 2]

1. National Primate Research Center, Korea Research Institute of Bioscience and Biotechnology (KRIBB), Cheongju, Chungbuk, Republic of Korea

2. University of Science & Technology (UST), National Primate Research Center, Korea Research Institute of Bioscience and Biotechnology (KRIBB), Cheongju, Chungbuk, Republic of Korea

ZKSCAN5 (also known as ZFP95) is a zinc-finger protein belonging to the Krűppel family. ZKSCAN5 contains a SCAN box and a KRAB A domain and is proposed to play a distinct role during spermatogenesis. In humans, alternatively spliced *ZKSCAN5* transcripts with different 5′-untranslated regions (*UTRs*) have been identified. However, investigation of our *Macaca* UniGene Database revealed novel alternative *ZKSCAN5* transcripts that arose due to an exon creation event. Therefore, in this study, we identified the full-length sequences of *ZKSCAN5* and its alternative transcripts in *Macaca* spp. Additionally, we investigated different nonhuman primate sequences to elucidate the molecular mechanism underlying the exon creation event. We analyzed the evolutionary features of the *ZKSCAN5* transcripts by reverse transcription polymerase chain reaction (RT-PCR) and genomic PCR, and by sequencing various nonhuman primate DNA and RNA samples. The exon-created transcript was only detected in the *Macaca* lineage (crab-eating monkey and rhesus monkey). Full-length sequence analysis by rapid amplification of cDNA ends (RACE) identified ten full-length transcripts and four functional isoforms of ZKSCAN5. Protein sequence analyses revealed the presence of two groups of isoforms that arose because of differences in start-codon usage. Together, our results demonstrate that there has been specific selection for a discrete set of *ZKSCAN5* variants in the *Macaca* lineage. Furthermore, study of this locus (and perhaps others) in *Macaca* spp. might facilitate our understanding of the evolutionary pressures that have shaped the mechanism of exon creation in primates.

694

Functional analysis of a natural mutational hotspot in the proximal promoter of a stress-response gene in *Drosophila melanogaster*

Miriam Merenciano[1], **Anna Ullastres**[1], **Josefa Gonzalez**[1]

1. Institute of Evolutionary Biology (CSIC-Pompeu Fabra University), Barcelona, Spain

Mutational hotspots are common in disease genes and in complex genomic regions not amenable to sequencing by short reads technologies. However, it is still not clear what makes a particular genomic region more prone to mutations and whether the different mutations located in a hotspot are functionally equivalent. In this work, we have discovered and characterized in detail an insertional hotspot in the promoter region of a stress-response gene in the fruitfly *Drosophila melanogaster*. The nine transposable elements insertions described are clustered in a small 368 bp region and all belong to the same family of transposable elements: the *roo* family. Although the sequences of these insertions are highly similar, their molecular and functional consequences are different: only *FBti0019985* insertion is associated with increased resistance to cold-stress. Interestingly, the previously described insertional hotspot in the *D. melanogaster* genome was also located in the promoter of a stress response gene suggesting that selection may favour the maintenance of genetic variability in these genes.

695

The role of heterozygous transposable elements in reproductive isolation among *Chironomus riparius* populations

Ann-Marie Oppold[2, 1], Marcel Rose[1], Hanno Schmidt[2], Sören L Hellmann[3], Fabian Ripp[3], Erwin Schmidt[3], Robert Kofler[4], Thomas Hankeln[3], Markus Pfenninger[2, 1]

1. Molecular Ecology Group, Goethe-University, Frankfurt am Main, Germany

2. Senckenberg Biodiversity and Climate Research Centre, Frankfurt am Main, Germany

3. Institute of Molecular Genetics, Johannes Gutenberg-University, Mainz, Germany

4. Department of Biomedical Sciences, Vetmeduni Vienna, Vienna, Austria

Transposable elements (TEs) have been shown to shape genome evolution and are suspected to drive speciation processes. However, besides the well-studied cases of classic *Drosophila*-specific TEs, knowledge on the mechanisms of how TEs may actually confer reproductive isolation (RI) is lacking. In our study on the non-biting midge *Chironomus riparius*, we propose a novel mechanism how transposons can contribute to species divergence. Focusing on a specific type of TE, the mostly tandem-repetitive, minisatellite-like *Cla-element*, whose increased activity is possibly associated with speciation events in the genus *Chironomus*, we investigated whether differential TE activity may be responsible for RI among conspecific populations.

With reciprocal crossing experiments, we found initial stages of RI among geographically and ecologically most distant *C. riparius* populations and visualized aberrations in giant polytene chromosomes of hybrid individuals. Using a novel *C. riparius* draft genome assembly, we unveiled diverging TE distribution patterns between populations. A highly significant correlation of the pairwise population F_{ST} as inferred by genome wide SNPs with the F_{ST} estimated from TEs suggested genomic drift as the major force driving TE population differentiation. However, there is significant indication for negative selection against heterozygous *Cla-element* insertions that especially occur in inter-population hybrids. Together with a possible imperfect pairing of homologous chromosomes in regions of heterozygous *Cla-element* bands, we suggest a new hypothesis on how the *Cla-element* might be involved in conferring RI in *C. riparius*, and, more generally, how TEs can contribute to species divergence. Our study thus shows that the role of TEs in the organisation of genomic architecture can influence fitness and directly contribute to evolutionary processes.

696

Do transposable elements facilitate adaptation?

Jessica Stapley[1]

1. ETH, Zurich, Zurich, ZURICH, Switzerland

Adaptation is fuelled by genetic variation and the ultimate source of genetic variation is mutation. One powerful mutagenic force in the genome is Transposable elements (TEs), which have the capacity to move and replicate within the genome. Recent work has demonstrated that TEs have contributed to the structure, function and evolution of genomes, and to heritable phenotypic change. Several studies have also demonstrated the adaptive role of specific TEs in natural populations. Despite these advances we still know very little about how often TEs have been involved in adaptation and if TE remodelling of the genomic landscape helps or hinders adaptation. To begin to address these questions I am investigating the distribution of TEs in relation to genes, adaptive loci and other genomic features across a diverse range of organisms. Using whole genome sequence data available online and published data of adaptive loci and/or genomic regions under selection, I will test if genomic regions involved in adaptation to different selective pressures are enriched for TEs, or specific families of TEs. I will also test if TE-rich regions are consistently enriched for genes belonging to particular GO terms across multiple species. The results will help us understand if TE activity influences adaptation and gene distribution across a broad range of taxa.

697

Transposons horizontally transferred between parasitic plants and hosts are still actively expressed in some recipients

Ting Sun[1], Yuxing Xu[1], Yan Qin[1], Guiling Sun[1]

1. Kunming Institute of Botany, CAS, Kunming, YUNNAN, China

Growing evidence is pointing to an important role of horizontal gene transfer (HGT) in the evolution of higher plants. However, reports of HGTs of transposable elements (TEs) in plants are still scarce, and only one case is known of a class II transposon horizontally transferred in grasses. To investigate possible TE transfers in dicots, we performed transcriptome screening in the obligate root parasite *Phelipanche aegyptiaca* (Orobanchaceae), data-mining in the draft genome assemblies of four other Orobanchaceae, gene cloning, gene annotation in species with genomic information, and a molecular phylogenetic analysis. We discovered that the broomrape genera *Phelipanche* and *Orobanche* acquired two related nuclear genes (christened *BO* transposase genes), a new group of the *hAT* superfamily of class II transposons, from Asian Sisymbrieae or a closely related tribe of Brassicaceae, by HGT. The lack of a classic border structure and low expression levels suggest that *BO* transposase genes rarely transpose in Brassicaceae, whereas their high transcriptional levels in *P. aegyptiaca* imply that they are still able to transpose or act as functional genes.

698

Multiple domestications of Asian rice with limited inter-group gene flow.

Peter Civáň[1], Terence A. Brown[1]

1. Manchester Institute of Biotechnology, University of Manchester, Manchester, United Kingdom

Domesticated rice (*Oryza sativa* L.) accompanied the dawn of Asian civilisation and has become one of world's most important staple crops. Although *O. sativa* is genetically differentiated into several groups with limited crossability, it is generally accepted that the genealogical history of at least some parts of rice genome trace back to a single domestication process. This is indicated by allelic uniformity of several *O. sativa* genes considered to be crucial for the domestication phenotype. However, whether these alleles originated in the wild population or during cultivation – a critical question for the interpretation of the domestication process – remains unclear.

We conducted a multi-layer analysis of a published whole-genome dataset of wild and cultivated rice, including reconstruction of complete chloroplast genomes, comparison of genome-wide selective sweep patterns, quantification of shared derived variants, as well as examination of the diversity associated with the domestication genes. Analyses of the chloroplast haplotypes and selective sweep patterns confirm the general distinctiveness of the *indica*, *japonica* and *aus* groups, and quantification of shared derived variants fails to support a domestication model where crucial alleles spread across groups by introgressive hybridization. We also found that domestication alleles such as *rc* and *laba1* (previously assumed to have emerged under cultivation) do occur in wild populations where they display higher associated diversity, indicating their pre-domestication origins.

All acquired evidence is consistent with geographically separate and genetically independent domestications leading to *indica*, *japonica* and *aus* rice. Given the presence of the *sh4*, *prog1*, *rc* and *laba1* alleles in wild populations, their allelic uniformity in *O. sativa* is parsimoniously explained by parallel selection of closely related haplotypes from standing variation of *O. rufipogon*. Domestication of rice was therefore a multiregional process that did not depend on transcontinental cultural interaction and probably followed different dynamics in various parts of Asia.

699

The Origin and Evolution of Fibromelanosis Locus in Domesticated Chickens: Comparison between Indonesian Cemani and Chinese Silky Genomes

Anik Budhi Dharmayanthi, Yohey Terai, Yoko Satta

The fibromelanosis (Fm) phenotype observed in both Indonesian Cemani and Chinese Silky chickens is represented by black pigmentation in skin and several organs. The mutation in the Fm region, which is located from 10.2 Mbp to 11.7 Mbp on chromosome 20, is known to be responsible for the black phenotype. The causal mutation is segmental duplication containing Endothelin3 (*EDN3*). The purpose of this study is to identify artificial selection in the Fm region and to elucidate the evolutionary history of Fm phenotypes in both Cemani and Silky chickens. We analyzed nine fragments of ~3 kb each in the segmental duplication and surrounding region of Cemani, Silky and other domesticated chicken populations. We also analyzed the whole region covering the nine fragments using Whole Genome Sequences from each of one individual of these chickens. We found a reduction of heterozygosity in the region close to the duplication region containing *EDN3* in Cemani and Silky populations. This indicates that selection has acted and reduced the variability, and this reduction is caused by selective sweeps with the target selection site. Comparison of heterozygosity and nucleotide diversity showed the distinct boundary of reduction of variability between Cemani and Silky, suggesting that Cemani and Silky have a different timing of selection. Moreover, the calculation of divergence time based on the region downstream of duplicated *EDN3* showed that this duplication in Fm region has occurred ~0.0135 mya just before the chicken domestication process in China ~0.010 mya.

700

What can morphometrics tell us about domestication at the age of genomics ?

Allowen EVIN[1], Thomas CUCCHI[2]

1. Institut des Sciences de l'Evolution, CNRS-Univ. Montpellier, Montpellier, France

2. Archéozoologie, Archéobotanique : Sociétés, Pratiques et Environnements, CNRS-MNHN, Paris, France

Following the development of high-throughput sequencing, the field of evolutionary genomics has undergone an unprecedented boom, but the limitations may now lie in the association of genomic variation with detailed phenotypic information. With the development of geometric morphometrics, a revolution has occurred in the field of morphometrics and phenotypic analyses. Geometric morphometrics offer powerful tools to dissect phenotypic variation and is often used to address questions in evolutionary biology. By focussing on research on the domestication of animals, we will reveal the advantages of combining genomic approaches with top-of-the art morphometric analyses.

Starting with the examples of pig and dog domestication, we will show why and how morphometric analyses have enabled to resolve questions that cannot be answered yet with genomic approaches. Especially when studying archaeological remains, the combined use of both approaches appears to be the best way forward.

701

The probability of monophyly of a sample of gene lineages given a species tree: an application to maize domestication.

Rohan Mehta[1], David Bryant[2], Noah Rosenberg[1]

1. Stanford University, Stanford, CA, United States

2. Mathematics and Statistics, University of Otago, Dunedin, New Zealand

Monophyletic groups---groups that consist of all the descendants of a most recent common ancestor---arise naturally as a consequence of descent processes that result in meaningful distinctions between organisms. Domestication induces strong selection and population bottlenecks, both of which can lead to an increase in the observed frequency of monophyletic groups in gene genealogies involving loci that may have been important in the domestication process. Here we present a theoretical formula for the probability that a particular sample of gene lineages is monophyletic given a species tree under a neutral coalescent model. Our formula extends previous work on two-species trees to arbitrarily many species. We study the effects of species tree topology and branch lengths on the monophyly probability, revealing new behavior, including the maintenance of nontrivial monophyly probabilities for gene lineage samples that span multiple species and

even for lineages that do not derive from a monophyletic species group. We also perform an example comparison of observed monophyly frequencies to theoretical monophyly probabilities for a dataset from maize and teosinte. The observed frequencies indicate that domesticated maize lineages show increased levels of monophyly compared both to theoretical expectations and to nondomesticated teosinte lineages. We present a software package, *Monophyler*, that facilitates computation of theoretical monophyly probabilities. Our results suggest that theoretical monophyly probabilities can be useful in the study of domestication.

702

Changing Patterns of Genomic Variability Following Domestication of Sheep

Marina M Naval Sanchez[1], Rudiger R Brauning[2], Shannon S Clarke[2], Alan A McCulloch[2], Noelle N Cockett[3], Ross R Tellam[1], Tony T Vuocolo[1], Wahid W Zamani[4], Francois F Pompanon[5], Pierre P Taberlet[5], Sean S McWilliam[1], Hans H Daetwyler[6], James J Kijas[1]

1. CSIRO, St Lucia, QLD, Australia

2. AgResearch Ltd, Invermay, Mosgiel , New Zealand

3. Utah State University, Logan, UT, USA

4. Department of Environmental Sciences, Tarbiat Modares University, Noor, Iran

5. Laboratoire d' Ecologie Alpine, Universite de Grenoble Alpes, Grenoble, France

6. Department of Economic Development Jobs, Jobs, Transports and Resources, Bundoora, Victoria, Australia

Patterns of genome variation in domestic animals have been shaped by the evolutionary and developmental process of domestication. Using whole-sheep genomes from a broad collection of domestic breeds and their wild ancestor the Asian mouflon (*O.orientialis*) we aim to assess the impact of domestication at a genome-wide scale. Following variant calling to identify approximately 45 million high confidence SNP, we find nucleotide diversity was generally higher in mouflon compared with their domesticated ancestors. To approach a deeper understanding of changing patterns of genomic diversity post domestication, we partitioned the sheep genome into 44 functional classifications. This utilised the current ovine annotation, sheep specific epigenomic datasets and human regulatory data (ENCODE and Epigenetics Roadmap) to predict the location of coding regions, promoters, enhancers and other elements of the gene regulatory machinery. Comparison of nucleotide diversity between functional classifications revealed coding regions contain the highest constraint, followed by promoters and finally enhancers. In addition, constraint within predicted promoters and enhancers is correlated to the number of tissues where they are active. Genomic regions that displayed marked reduction in diversity within domestic sheep, compared with the wild ancestors, were evaluated against each functional classification. We find strong evidence for enrichment for enhancer regions, suggesting a major consequence of domestication and selection has occurred in regulatory regions of the sheep genome. Currently, we are further characterizing these selective sweeps and their predicted consequences on gene expression.

703

Cracking the nut: genome and transcriptome sequencing of *Macadamia integrifolia* (Proteaceae)

Catherine J Nock[1], A K M Abdul Baten[1], Bronwyn J Barkla[1], Agnelo Furtado[2], Robert J Henry[2], Graham J King[1]

1. Southern Cross University, Lismore, NSW, Australia

2. QAAFI, University of Queensland, Brisbane, QLD, Australia

Macadamia is a native Australian nut crop belonging to the diverse Gondwanan family Proteaceae. Genetic resources are extremely limited, restricting macadamia breeding programs and comparative genomic studies. In this study, over 95 gigabases of sequence data from the genome and transcriptome of *Macadamia integrifolia* cultivar HAES 741 were assembled. In total, 35,337 protein-coding genes were predicted and of these, 90% were supported by RNA-seq based expression evidence. *Macadamia* represents an ancient rainforest-restricted Proteaceae lineage and comparative gene family analysis provides evidence for an expansion of gene families involved in pathogen recognition, plant defense and monoterpene synthesis. Although rare among plant species, many domesticated food plants are cyanogenic. The relatively high proportion of cyanogenic species in the Proteaceae indicates that cyanogenesis is an important defense strategy in this family. Several of the candidate genes for cyanogenic glycoside biosynthesis that were identified were highly expressed in the *M. integrifolia* leaf, shoot and flower tissues examined. These first genomic resources for the Proteaceae and for macadamia provide a platform for comparative genomics, and new opportunities to identify the genes and markers associated with traits of importance for conservation, domestication and plant breeding.

704

A discriminative model-based approach to inferring the geographic origin of domestic species

Anna AR Rudzinski[1]

1. UCL, London, UK, United Kingdom

I present a spatially-explicit discriminative modeling approach to infer the location of origin of domesticated plant and animal species from genetic and morphometric variation data. The model is based on the expected monotonic reduction in diversity with geographic distance from region of origin. Such a pattern is expected because as a population expands in space, variation is sampled on the wavefront of expansion, leading to a loss of diversity. My approach performs a search geographic space to identify the region where this correlation is maximized. I account for sparse and uneven sampling, and the possibility of high homozygosity through selfing in plant and animal species, by implementing a spatial kernel. I include a permutation test in order to assign significance for inference of location of origin.

The method has been applied to various species such as broomcorn millet (*Panicum miliaceum*) microsatellite data to infer that the crop spread from northeast China. We also apply our approach to morphological variation data of the Polynesian rat (Rattus exulans) to investigate the origin of expansion.

705

Characterization of grain amaranth (*Amaranthus L*) domestication with Genotyping by Sequencing and whole genome sequencing

Markus G Stetter[1], Georg Haberer[2], Karl J Schmid[1]

1. University of Hohenheim - Institute of Plant Breeding and Seed Science, Stuttgart, BADEN-WüRTTEMBERG, Germany

2. MIPS/IBIS Helmholtz Zentrum, München, Bavaria, Germany

Grain amaranth is a pseudo-cereal and an ancient crop of Central and South America. Of the three species of grain amaranth, *Amaranthus caudatus* is mainly cultivated in the Andean region. To investigate the domestication history of *A. caudatus* and its relationship to the two wild relatives *A. quitensis*and *A. hybridus*, we used Genotyping by Sequencing (GBS) to genotype 119 amaranth accessions from the Andean region. We determined the genome sizes of the three species and compared phenotypic variation in two domestication-related traits, seed size and seed color. A population genetic analysis revealed very little genetic differentiation between the two wild species, suggesting they are the same species, but showed a strong differentiation between wild and domesticated amaranths despite evidence for a significant level of recent gene flow. Genome sizes and seed sizes were not significantly different between wild and domesticated amaranths, although a genetically distinct cluster of Bolivian accessions had significantly larger seeds. The analysis of seed size and seed color indicates that South American grain amaranth is an incompletely domesticated species, either because it was not strongly selected or because high levels of gene flow from its sympatric wild relatives counteract the fixation of key domestication traits. We sequenced *de novo* the *A. caudatus* genome with 150 x coverage resulting in a N50 of 150 kb. Additionally, we re-sequenced 120 individuals from all three grain species (*A.* caudatus, *A. cruentus* and *A. hypochondriacus)* and their potential wild ancestors (*A. hybrudus* and *A. hybridus* sp *quitensis)* with 10 x coverage to further trace the genetic background of the incomplete domestication. The genomic data together with transcriptomic information is being used for demographic analysis of amaranth domestication. Understanding the reason for incomplete domestication in amaranth can give additional insights into the process of crop domestication.

706

The genomic signature of a 4,600 year old Scandinavian dog adds a time-depth to modern basal dog breeds

Emma Svensson[1], Torsten Günther[1], Helena Malmström[1], Arielle Munters[1], Jan Storå[2], Mattias Jakobsson[1]

1. Uppsala University, Uppsala, UPPLAND, Sweden

2. The Osteo-archaeology Research Laboratory, Stockholm University, Stockholm, Sweden

The dog (*Canis familiaris*) was the first animal to be domesticated, several thousands of years prior to the domestication of other animals. However the timing and location of domestication still remains difficult to pinpoint. Today there are thousands of different types of dogs in the world, but most of the modern breeds just trace their origin back to the last 200 years. However genome wide SNP data has shown that there are some modern breeds that appear basal compared to all others. These are breeds can mostly be found in the geographical fringes of dog distribution (for example Scandinavia), and they have been genetically isolated from other dogs for a long time.

Here we have generated a low coverage genome from an ancient dog dated to ca 2830-2485 cal BC. We sampled a tooth from a complete skeleton found in a burial in South Eastern Sweden. The dog is morphologically similar to modern dogs of the Spitz-type which include a number of basal breeds.

The genome of the ancient Swedish dog is very similar to Spitz-type dogs and other breeds that have been identified as 'basal' or 'ancient' breeds based on their genetic signature. This pattern suggests that the foundation for some dog breeds had started more than 4,600 years ago and that such basal breeds have received relatively little admixture from other dogs since then. Furthermore, we investigate whether the ancient Swedish dog received introgession from wolves, in particular from an extinct Siberian wolf linage which has admixed with modern breeds from high latitudes.

The genome from our 4,600 year old dog provides a first step towards understanding the origin of dog breeds beyond the historical records of the past 200 years which can also help to shed light on dog domestication.

707

Out of southern East Asia: the natural history of domestic dogs across the world

Guo-Dong Wang[1], Weiwei Zhai[2], He-Chuan Yang[1], Lu Wang[3], Andrei Poyarkov[4], David Irwin[5], Hannes Lohi[6], Chung-I Wu[7], Peter Savolainen[8], Ya-Ping Zhang[1]

1. Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, YUNNAN, China

2. Human Genetics, Genome Institute of Sin-gapore, Singapore

3. Yunnan University, Kunming, Yunnan, China

4. Severtsov Institute of Ecology and Evolution, Russian Academy of Science, Moscow

5. University of Toronto, Toronto, Canada

6. University of Helsinki, Helsinki

7. University of Chicago, Chicago

8. KTH-Royal Institute of Technology, Science for Life Laboratory, Solna

The origin and evolution of domestic dogs has been an interesting and controversial question for the scientific community. With whole genome sequences gathered for a total of 58 canids (12 grey wolves, 27 primitive dogs from Asia and Africa and a collection of 19 diverse breeds from across the world), we revealed an ancient origin of domestic dogs in Southern East Asia 33,000 years ago. After living in East Asia for thousands of years, a subset of dog ancestors started migrating to the Middle East, Africa and Europe, about 15,000 years ago, possibly through the Indian coastal areas. Interestingly, one of the Out of Asia lineages also migrated back to the east, creating a series of admixed populations with the endemic Asian lineages in Northern China before migrating to the New World.

This study opens many potential avenues for future research. Collection of additional samples from other parts of the World (especially the Indian coastal region and Northern Eurasia) should allow us to draw a more complete picture of the worldwide migration patterns, and their association with human populations. The study of the Chinese indigenous dogs has provided an unprecedented opportunity for illuminating the history of selection during dog domestication. For example, as dogs established stronger bonds with humans, possibly empowered by the origin of modern agriculture in the Middle East and China, strong selection for genes involved in metabolism and morphology/development emerged. Our study, for the first time, begins to reveal the extraordinary journey that our best friend has traveled on this planet, and a large and complex landscape upon which a cascade of positive selective sweeps occurred during the domestication of the dog.

1. Guo-Dong Wang, Weiwei Zhai, He-Chuan Yang, Lu Wang, Li Zhong, Yan-Hu Liu, Ruo-Xi Fan, Ting-Ting Yin, Chun-Ling Zhu, Andrei D Poyarkov, David M Irwin, Marjo K Hytönen, Hannes Lohi, Chung-I Wu, Peter Savolainen, Ya-Ping Zhang. 2016. Out of southern East Asia: the natural history of domestic dogs across the world. Cell Res. doi: 10.1038/cr.2015.147
2. Guo-Dong Wang, Min-Sheng Peng, He-Chuan Yang, Peter Savolainen, Ya-Ping Zhang. 2016. Questioning the evidence for a Central Asian domestication origin of dogs. PNAS. doi: 10.1073/pnas.1600225113
3. Jiao-Jiao Songa,Wen-Zhi Wang, Newton O. Oteckob, Min-Sheng Peng, Ya-Ping Zhang, 2016. Reconciling the conflicts between mitochondrial DNA haplogroup trees of Canis lupus. Forensic Science International: Genetics. doi:10.1016/j.fsigen.2016.03.008

708

Genome variability in the ecologically significant marine diatom *Leptocylindrus* (Bacillariophyta) in a southern hemisphere upwelling system

Penelope A Ajani[2, 1], Gurjeet S Kohli[3], Aaron E Darling[4], Michael Liu[4], Martin Ostrowski[5], Shauna A Murray[2, 1]

1. Climate Change Cluster, University of Technology Sydney, Ultimo, NSW, Australia

2. Sydney Institute of Marine Science, Mosman, NSW, Australia

3. Singapore Centre for Environmental Life Sciences Engineering, Nanyang Technological University, Singapore

4. i3 institute, University of Technology Sydney, Ultimo, NSW, Australia

5. Department of Chemistry and Biomolecular Science and Macquarie Marine Research Centre, Macquarie University, Sydney , NSW, Australia

Diatoms are a highly productive and diverse class of unicellular marine eukaryotes, carrying out 25% of the world's photosynthesis and generating ~ 40% of organic matter produced by the ocean each year. The diatom *Leptocylindrus* Cleve is a major component of phytoplankton blooms in coastal ecosystems and upwelling regions worldwide and although reported from Australia since the 1930s, there is little known about this genus in the southern hemisphere. Using light and transmission electron microscopy and molecular phylogenetics based on the nuclear-encoded ITS1/5.8S/ITS2 rDNA region, our study has characterised three species, *L. aporus*, *L. convexus* and *L. danicus*, from 55 clonal isolates. Using Illumina high throughput sequencing technology, de novo genome assemblies are currently being examined to characterise and compare the genome architecture of these three species. Furthermore, by mapping the genomes of multiple strains of *L. danicus*, our study aims to investigate the intraspecific diversity of diatom populations along this coastline. It is envisaged that both inter-species and intra-species genetic information on these ecologically significant diatoms will provide critical insights into modes of evolution, stress response and the adaptive capacity of marine organisms to changing ocean conditions.

709

Global phylogeography of *Coccidioides spp.*; the etiologic agent of Valley Fever

David M Engelthaler[1], Chandler C. Roe[1], Crystal Hepp[2], Marcus Teixeira[1], Elizabeth M. Driebe[1], James M. Schupp[1], Lalitha Gade[3], Victor Waddell[4], Kenneth Komatsu[4], Eduardo Arathoon[5], Heidi Logemann[6], George R Thompson, III[7], Tom Chiller[3], Bridget Barker[1], Paul Keim[1, 8], Anastasia P. Litvintseva[3]

1. TGen North, Translational Genomics Research Institute, Flagstaff, AZ, USA

2. Informatics and Computing Program, Northern Arizona University, Flagstaff, AZ, United States

3. Mycotic Diseases Branch, National Center for Emerging and Zoonotic Infectious Diseases, Centers For Disease Control and Prevention, Atlanta, GA, USA

4. Division of Public Health Services, Arizona Department of Health Services, Phoenix, AZ, USA

5. Asociación de Salud Integral, Guatemala City, Guatemala

6. Universidad de San Carlos, Guatemala City, Guatemala

7. Division of Infectious Diseases, Department of Medicine, University of California Davis, Davis, CA, USA

8. Center for Microbial Genetics and Genomics, Northern Arizona University, Flagstaff, AZ, USA

Each year, the etiologic agents, *Coccidioides immitis* and *C. posadasii*, cause Valley Fever in tens of thousands of individuals. While it widely accepted that this fungal disease is endemic to arid locations, such as the southwestern United States, the recent discovery of endemic clusters in Washington state suggests an expansion of the geographic range. Here, we present a whole genome analysis of 86 genomes, where 68 are unique to this study. The incorporation of Bayesian phylogenetics resulted in the identification of phylogeographic structure of both species, and calibrations on the root node reveal that *C. posadasii* is the more ancient of the two species. Taken together, we propose that *C. posadasii* originated near the Arizona-Mexico border, and we suggest a subsequent dispersal mechanism and route of spread.

710

Identification of microsporidia host-exposed proteins reveals a repertoire of large paralogous gene families and rapidly evolving proteins

Aaron Reinke[1], Eric Bennett[1], Emily Troemel[1]

1. UCSD, La Jolla, CALIFORNIA, United States

Although it is clear that intracellular pathogens use a variety of secreted and surface proteins to interact with and manipulate their hosts, a systematic approach for identifying such proteins has been lacking. Because of this, the identity of these host-exposed proteins is often unknown in many pathogens. Additionally, little is known about how conserved repertoires of host-exposed proteins are between related species. Microsporidia are a large phylum of eukaryotic obligate intracellular parasites that can specifically infect a variety of different animal species. To identify host-exposed proteins from microsporidia, we used spatially restricted enzymatic tagging followed by mass spectrometry on *C. elegans* infected with two related species of *Nematocida* microsporidia. Using this approach, we identified 82 microsporidia proteins that are exposed inside of host intestinal cells, including several in the nucleus. These proteins are enriched in targeting signals, lack conservation with other microsporidia species, are rapidly evolving, and lack domains with known function. Almost half of the identified proteins belong to large, *Nematocida*-specific gene families that are undergoing species-specific radiations. We also find that large, species-specific families with targeting signals are common throughout microsporidia species. Our data suggest that the use of a large number of rapidly evolving species-specific proteins represent a common strategy for these intracellular pathogens to interact with their hosts.

711

Comparative and population genomics of *Phellinus noxius* causing brown root rot disease in trees

Isheng J Tsai[1], Chia-Lin Chung[2], Hsin-Han Lee[2], Jiaye T Lee[1], Tzuhao Kuo[1], Huei-mie Ke[1], Yen-Ping Hsueh[3], Meiyeh J Lu[1], Chien-Yu Chen[4], Pao-Jen Ann[5], Jyh-Nong Tsai[5], Mitsuteru Akiba[6], Tsutomu Hattori[6], Yuko Ota[6], Norio Sahashi[6], Taisei Kikuchi[7], Ruey-Fen Liou[2]

1. Biodiversity Research Center, Academia Sinica, Taipei City, Taiwan

2. Department of Plant Pathology and Microbiology, National Taiwan University, Taipei City, Taiwan
3. Institute of Molecular Biology, Academia Sinica, Taipei City, Taiwan
4. Department of Bio-Industrial Mechatronics Engineering, National Taiwan University, Taipei City, Taiwan
5. Plant Pathology Division, Taiwan Agricultural Research Institute, Taichung City, Taiwan
6. Department of Forest Microbiology, Forestry and Forest Products Research Institute, Tsukuba, Japan
7. Division of Parasitology, Faculty of Medicine, University of Miyazaki, Miyazaki, Japan

Phellinus species are basidiomycete fungi that are widely distributed tropical and sub-tropical regions of the world. In particular, *Phellinus noxius* causes brown root rot disease in 200+ known tree species. Originally played a role of replacing old trees in natural forests, they have recently have emerged to be problems to greater disease foci in forests, monoculture plantations and amenity trees in urban cities. We produced a 12 piece 31.6Mb genome assembly from a Japanese *Phellinus noxius* KPN-91 isolate using long read sequencing, as well as draft assemblies from three other members of Hymenochaetaceae. Repeat induced mutations are apparent in the genomes of these species, with greater than 40kb transposable element clusters indicative of centromeres. *P. noxius* genome presents itself with little variability in gene density, simple repeat density or conserved genes along its scaffolds, with the exception of expansion of nucleotide-binding oligomerization domain (NOD)-like receptors (NLRs). Interestingly, degraded mesosynteny are observed between *P. noxius* and *C. coprinopsis,* suggesting karyotype conservation are still conserved in basidiomycetes. Sequencing of 61 isolates of *P. noxius* collected in Japan and Taiwan over 2012-2016 period reveals a single population with exceptionally high diversity at synonymous sites of 0.15. Transcriptome comparisons between *P. noxius* fungal mats of infected trees against fruiting bodies reveal upregulated genes involved in trehalose biosynthetic pathway which has been implicated in infectivity by fungal pathogens. The Hymenochaetales are phylogenetically placed between the better-studied Polyporales/Agaricales and Tremellales/Ustillaginales orders, making the small assembly of *P. noxius* an attractive genome to study the evolutionary transition between these orders of Basidiomycota.

712

Molecular ecology of palytoxin producing protist *Ostreopsis siamensis* (Alveolata) along the East Australian Current

Arjun Verma[1], David J Hughs[1], Gurjeet S Kohli[1], Tim D Harwood[2], David J Suggett[1], Peter J Ralph[1], Shauna A Murray[1]

1. Plant functional biology and Climate Change Cluster, University of Technology, Sydney, Sydney, NSW, Australia
2. Cawthron Institute, Nelson, New Zealand

Ostreopsis siamensis is a morphologically cryptic marine protist (Alveolata: Dinophyceae) that produces toxic and complex non-peptide compounds, palytoxins (PLTX). The intraspecific genetic diversity of marine protists across environmental gradients of tens to hundreds of kilometres is still relatively little known, and the extent to which these differences are linked to functional traits has largely not been explored. Despite a widespread distribution along the East Australian coastline, and ecologically detrimental blooms of *O. siamensis*, little is known about the interspecific genetic and phenotypic diversity driving PLTX biosynthesis amongst their populations. Such information is important in furthering our understanding of likely changes in the populations of this species, as the East Australian Current is considered a 'climate change hotspot', with increases of up to 2.0 °C over the past 100 years, and a more southern range extension. In our study, we analysed ribosomal (rDNA) data, morphological features, photophysiological traits (FRRf) and toxin profiles (LC-MS/MS) from 55 *O. siamensis* clonal isolates from 8 different sites along the east Australian coastline, to determine the cooperative and specific traits amongst and between different geographic populations that promote the ecological success of this species. RNA sequencing (RNA-seq) analyses were performed to identify genes encoding for key metabolic pathways such as toxin biosynthesis. Our findings suggest that *O. siamensis*may allow mutualistic intraspecific facilitation in multiple ways, thereby promoting the overall success of the species and facilitation of its expansion.

713

Why do ciliates have such low mutation rates, and why don't we?

David J Winter[1], Hongan Long[2], Allan YC Chang[3], Ricardo A Azevedo[3], Reed A Cartwright[1], Rebecca B Zufall[3]

1. Arizona State University, Tempe, AZ, United States
2. Department of Biology, Indiana University, , Bloomington, IN, United States
3. Biology and Biochemistry, University of Houston, Houston, TX, United States

Ciliates, microbial eukaryotes that contain separate germline and somatic nuclei, provide a near perfect system in which to study mutation. During vegetative growth, the ciliate germline genome is duplicated but not expressed. Thus, mutations can accrue over many rounds of cell division without being exposed to natural selection, allowing the full range of spontaneous mutations to be studied. However, detecting these mutations through short read sequencing is particularly difficult. A lack of complete reference genomes, the binuclear nature of ciliates and the large number of genomic rearrangements that occur during ciliate development all contribute to a high error-rate when using traditional mutation detection methods.

Using a novel approach to mutation detection, we have estimated the mutation rate for the ciliate *Tetrahymena thermophila* and found that it is one of the lowest if not the lowest ever recorded. The only other ciliate for which a direct estimate of the mutation rate is available, *Paramecium tetraurelia*, is among the very few species with a comparable rate.Here we describe the unique challenges associated with taking whole genome approaches to ciliates, and how they can be overcome. We also propose and test mechanistic and evolutionary hypotheses that might explain the low mutation rates observed in this group.

714

The genome of a nematode isolated from the deep, hot terrestrial subsurface reveals horizontal transfer and amplification of Hsp70 genes as an adaptive strategy

Sarah Allen[1], **John R Bracht**[1]

1. American University, Washington DC

Genetic signatures associated with adaptation to extreme conditions are of interest for understanding the limits of life on earth and in the search for extraterrestrial life, but animals are relatively rare in extreme habitats. We have performed sequencing and protein prediction on a unique extremophile metazoan, *Halicephalobus mephisto,* a nematode isolated from the Beatrix gold mine in South Africa, just over a kilometer below the earth's surface. The deep subterranean environment subjects *H. mephisto* to high pressure, heat (up to 41 degrees Celsius), utter darkness, and low oxygen. Here we present data supporting an unusual evolutionary signature in its genome consistent with adaptation to an extremely warm environment. Specifically, a sixteen-nematode comparative analysis revealed an expanded repertoire of 70 kilodalton heat-shock protein (Hsp70) in *H. mephisto,* which contains over 100 detected Hsp70 gene paralogs as compared to just 16 in *C. elegans*. *H. mephisto* exceeds all sequenced nematodes (if not all sequenced organisms) in Hsp70 gene content, which is significant because Hsp70 proteins are specialized chaperones for re-folding heat-damaged proteins. In addition to a large number of Hsp70 proteins, we found evidence for 15 of these Hsp70 proteins being derived from bacterial sources, some of which are also extremophiles. These data are consistent with horizontal gene transfer as a mechanism of animal adaptation to the deep, hot terrestrial subsurface.

715

Discrete dynamics of stem cell niches - a cellular automaton model

Alonso Sandoval[1]

1. National Autonomous University of Mexico, Coyoacán, MEXICO CITY, Mexico

Stem cell microenvironment is involved in regulating stem cell fate regarding self-renewal, quiescence, and differentiation. Mathematical models may be useful in understanding the dynamics of the regulation and geometrical organization of such microenvironments.

Stem cells in the tissues of both animals and plants are often located and controlled by special microenvironments known as niche. Studies in different stem cells niches in model systems (*Drosophila* ovary and testes, bone marrow, hair follicles, subventricular zone of the brain, and villi of the intestinal tract in mammals, apical meristems and radicals in vascular plants) have revealed adhesive interactions, changes in the cell cycle, cell signaling and common spatial organizations (confinement) operating to control the behavior of stem cells. Thus it seems that these niches are an ancient evolutionary module with characteristics conserved in different places, tissues and organisms regarding their role and especially their organization.

In this work we conceived the activity and fate of stem cells as a function of local interaction with their environment, and by generalizing and organizing this molecular interactions in short-range and long-range interactions, we build a 'cellular automaton' model that represents the organizational geometry of the stem cell niches in several organisms along the multicellular linage. The model takes into account all the different types of interactions that exist in the niche, and through alterations demonstrates the robustness of the system and thus the main molecular and genetic factors involved in it.