polymorphism (SNP) alignments, containing both modern and ancient sequences, it is common to find positions with unknown nucleotides (gaps) that could generate problems in the phylogenetic reconstruction. Thus, the use of complete deletion alignments is fairly common. This practice, however, could cause the loss of potentially important information, so we aim to identify the most suitable deletion threshold for the proportion of unknown sites allowed for a given alignment before proceeding to analyze the data in BEAST. Here, I present the temporal signal of 204 whole-genome sequences of *Yersinia pestis*, a zoonotic gram-negative bacteria and causal agent of the bubonic, pneumonic, and systemic plagues. I demonstrate measurable temporal signal for the alignment with thresholds of 0–10 per cent for the proportion of unknown sites per SNP. The results showed that a complete deletion alignment presented the lowest correlation and greatest residual mean squared values. The best threshold depends on the method used to find the best root, but appears to be between 7–9 per cent.

A59 A strategy to studying zoonotic infectious diseases

P. Anh,[1] P. Tam,[1] N. Tue,[1] M. Rabaa,[1,2] G. Thwaites,[1,3] and S. Baker[1,3]

[1]Oxford University, Ho Chi Minh City, Vietnam, [2]University of Edinburgh, Edinburgh, Scotland, UK and [3]Oxford University, Oxford, UK

An increasing number of zoonotic viruses have been detected in animals, especially in poultry species. Understanding the diversity of zoonotic infections and the local behavior helps to characterize the pathogen diversity in human and animals and predict the risk of pathogen spill-over from animals to human. Vietnam is considered, along with other countries in Southeast Asia, as a hotspot for zoonotic viruses. In Vietnam, domestic animals are typically farmed in close proximity to humans, which may increase the risk of transmission of zoonotic pathogens. Our previous studies found the presence of some zoonotic viruses (e.g. rotavirus group A, hepatitis E virus) in domestic pigs. However, the risk of pathogenic transmission from domestic animals to humans has not been determined. Detailed genomic sequence data may help to track the origin and evolution of zoonotic pathogens. To understand the origins and emergence of zoonotic infections in people, who have regular contact with animals, we will investigate the viral diversity in farmers and domestic animals in their farm, using high-throughput sequencing technique. Viral RNA was extracted from pooled fecal samples of 30 farmers and 50 pigs, and used as input for SureSelect target enrichment and Illumina MiSeq sequencing.

A60 Revealing the evolution of virulence in RNA viruses

Marina Escalera-Zamudio, Bernardo Gutiérrez, Julien Thézé, and Oliver G. Pybus

Department of Zoology, Oxford University, Parks Rd Oxford, OX1 3PS, UK

A combination of high rates of mutation and replication, coupled with strong natural selection, ensures that RNA viruses experience rapid genotypic and phenotypic evolution. Such a 'fast-forward' evolution enables viruses to rapidly adapt to new host species, evade host immune responses, and to develop resistance to anti-viral drugs. Similarly, rapid evolution allows viruses to attain new levels of virulence, defined as the ability to cause severe disease in hosts. We hypothesize that distinct viral groups share genetic determinants that modulate virulence that have been acquired through convergent evolution. Thus, common patterns reflecting changing virulence-related specific viral groups could be detected. The main goals for this project are (1) to understand how genetic and phenotypic diversity can be generated among different viral groups by analyzing the variation patterns and determining the selective forces behind them (impact in viral fitness) and (2) to understand how fixed mutations can modulate virulence within different viral groups by performing comparison of strains with differing virulence within a longitudinal timescale. The subject of the study is key emerging and re-emerging virus families of medical importance. Such groups include: Coronaviridae (severe acute respiratory syndrome and Middle East respiratory syndrome-associated coronaviruses), Picornaviridae (Hepatitis A virus), Flaviviridae (Yellow fever, West Nile, Hepatitis C, Dengue, and Zika viruses), Togaviridae (Rubella and Chikungunya virus), Bornaviridae (Borna-disease virus), Filoviridae (Ebola and Marburg viruses), Paramyxoviridae (Measles, Nipah, and Hendra viruses), Rhabdoviridae (Lyssaviruses), Arenaviridae (Lassa virus), Bunyaviridae (Hanta- and Crimean-Congo hemorrhagic fever viruses), and Orthomyxoviridae (Influenza A viruses). Viral genomes collected at different time points, different hosts (human and their most closely related animal reservoirs) and different locations will be compiled. Extensive molecular evolutionary analyses will be carried out to infer gene expansion/contraction within groups, rates of evolution, and changes in selection pressure, including the detection of positive selected genes and sites (adaptive evolution). Positively selected sites will be mapped onto the viral protein structures to reveal their impact on function, and hence the location of potential virulence determinants. Virulence changes among particular viral strains and types will be defined and measured according to definitions based on an increase in: (1) transmissibility, (2) host tropism, (3) immune evasion, (4) morbidity and mortality, (5) drug resistance, and by the incorporation of epidemiological data to determine whether high or low virulence strains within different hosts and localities are spreading most efficiently in nature.

A61 Large RNA genomes: Is RNA polymerase fidelity enough?

F. Ferron and B. Canard

Centre National de la Recherche Scientifique, and Aix-Marseille University, Université d'AIX-Marseille, UMR7257-AFMB, 163 avenue de Luminy, 13288 Marseille cedex 09, France

Large-genome Nidoviruses and Nidovirus-like viruses reside at the current boundary of largest RNA genome sizes. They code for an unusually large number of gene products matching that of small DNA viruses (e.g. DNA bacteriophages). The order of appearance and distribution of enzyme genes along various virus families (e.g. helicase and ExoN) may be seen as an evolutionary marker in these large RNA genomes lying at the genome size boundary. A positive correlation exists between (+)RNA virus genome sizes and the presence of the RNA helicase and the ExoN domains. Although the mechanistic basis of the presence of the helicase is still unclear, the role of the ExoN activity has been linked to the existence of an RNA synthesis proofreading system. In large Nidovirales, ExoN is bound to a processive replicative RNA-dependent RNA polymerase (RdRp) and corrects mismatched bases during viral RNA synthesis. Over the last decade, a view of the overall process has been refined in Coronaviruses, and in particular in our lab (Ferron et al., PNAS, 2018). We have identified genetic markers of large RNA genomes that we wish to use to data-mine currently existing metagenomic datasets. We have also initiated a collaboration to sequence and explore new viromes that will be searched according to these criteria. Likewise, we have a collection of purified viral RdRps that are currently being used to generate RNA synthesis products that will be compared to existing NGS datasets of cognate viruses. We will be able to have an idea about how much genetic diversity is possibly achievable by viral RdRp ('tunable fidelity') versus the detectable diversity (i.e. after selection in the infected cell) that is actually produced.

A62 A major likelihood-based approach gives problematic estimates of diversification dynamics and rates

Chenlu Di,[1] Andreas L. S. Meyer,[1,2] and John J. Wiens[1]

[1]Department of Ecology and Evolution, University of Arizona, Tucson, AZ 85721, USA and [2]Graduate Program in Zoology, Universidade Federal do Paraná, Curitiba, PR 81531-990, Brazil

The diversity of life is shaped by rates of speciation and extinction, and so estimating these rates correctly is crucial for understanding diversity patterns among clades, regions, and habitats. In 2011, Morlon and collaborators developed a promising likelihood-based approach to estimate speciation and extinction and to infer the model describing how these rates change over time based on AICc. This approach is now implemented in an R package (RPANDA). Here, we test the accuracy of this approach under simulated conditions, to evaluate its ability to correctly estimate rates of speciation, extinction, and diversification (speciation—extinction) and to choose the correct underlying model of diversification (e.g. constant or changing rates of speciation and extinction over time). We found that this likelihood-based approach frequently picked the incorrect model. For example, with changing speciation rates over time, the correct model was chosen in only ~10 per cent of replicates. There were significant relationships between true and estimated speciation rates using this approach, but relationships were weak when speciation rates were constant within clades.

Relationships were consistently weak between true and estimated rates of extinction and of diversification. Overall, we suggest that results from this approach should be interpreted with considerable caution.

A63 Quantifying the dynamics of evolutionary rates through time

J. V. Membrebe,[1] G. Baele,[1] M. A. Suchard,[2,3] and P. Lemey[1]

[1]Department of Microbiology and Immunology, Rega Institute, KU Leuven, Leuven, Belgium, [2]Departments of Biomathematics and Human Genetics, David Geffen School of Medicine at UCLA, University of California, Los Angeles, CA 90095, USA and [3]Department of Biostatistics, UCLA Fielding School of Public Health, University of California, Los Angeles, CA 90095, USA

The availability of evolutionary rate estimates in recent years led to the observation that they may depend on the time-scale on which they are measured. Specifically, RNA virus evolutionary rates are frequently estimated to be low towards the past and high towards the present. This time-dependent rate phenomenon (TDRP) has important implications for evolutionary studies as it could severely bias divergence time estimates. While recent studies are providing insights into the relationship between viral evolutionary rate and time, formal probabilistic models to draw inference under TDRP scenarios remain lacking. Here, we adopt epoch-modelling to develop a Bayesian model of discrete rate changes through time in an unknown evolutionary history and combine this with a log-linear parameterization of rates as a function of times in the past. We provide an implementation for nucleotide substitution rates as well as for nonsynonymous rates change in a codon substitution model. Using a foamy virus dataset for which internal node calibrations can be applied based on host-virus co-divergence, we estimate a significant decline in evolutionary rates as a function of time into the past for nucleotide substitutions as well as for non-synonymous substitutions in a codon model. We also estimate a deep evolutionary history for primate Lentiviruses by combining an HIV-1 group M node calibration and a biogeographic calibration for viruses in drill monkeys in the TDRP model. Our analyses lead to the conclusion that studies of evolutionary timescales require a reconsideration of substitution rates, in either codon and nucleotide substitution model, as a dynamic feature of molecular evolution.

A64 Viral sequence classification using deep learning algorithms

David Nieuwenhuijse, Bas Oude Munnink, My Phan, and Marion Koopmans

[1]Department of Viroscience, Erasmus Medical Center, Rotterdam, the Netherlands

Sewage samples have a high potential benefit for surveillance of circulating pathogens because they are easy to obtain and reflect population-wide circulation of pathogens. These type of samples typically contain a great diversity of viruses. Therefore, one of the main challenges of metagenomic sequencing of sewage for surveillance is sequence annotation and interpretation. Especially for high-threat viruses, false positive signals can trigger unnecessary alerts, but true positives should not be missed. Annotation thus requires high sensitivity and specificity. To better interpret annotated reads for high-threat viruses, we attempt to determine how classifiable they are in a background of reads of closely related low-threat viruses. As an example, we attempted to distinguish poliovirus reads, a virus of high public health importance, from other enterovirus reads. A sequence-based deep learning algorithm was used to classify reads as either polio or non-polio enterovirus. Short reads were generated from 500 polio and 2,000 non-polio enterovirus genomes as a training set. By training the algorithm on this dataset we try to determine, on a single read level, which short reads can reliably be labeled as poliovirus and which cannot. After training the deep learning algorithm on the generated reads we were able to calculate the probability with which a read can be assigned to a poliovirus genome or a non-poliovirus genome. We show that the algorithm succeeds in classifying the reads with high accuracy. The probability of assigning the read to the correct class was related to the location in the genome to which the read mapped, which conformed with our expectations since some regions of the genome are more conserved than others. Classifying short reads of high-threat viral pathogens seems to be a promising application of sequence-based deep learning algorithms. Also, recent developments in software and hardware have facilitated the development and training of deep learning algorithms. Further plans of this work are to characterize the hard-to-classify regions of the poliovirus genome, build larger training databases, and expand on the current approach to other viruses.

A65 Characterization of endolysin gene of bacteriophages infecting *Listeria* spp. isolated from dairy industry wastewater

Blanco Fernández,[1,2] M. E. Barrios,[1,2] R. V. Cammarata,[1,2] C. Torres,[1,2] and V. A. Mbayed[1,2]

[1]Cátedra de Virología, Facultad de Farmacia y Bioquímica, Universidad de Buenos Aires, Buenos Aires, Argentina and [2]Consejo Nacional de Investigaciones Científicas y Tecnológicas, CONICIT, San José, Costa Rica

Bacteriophages and their endolysins, enzymes that degrade the cell walls of bacteria, are emerging as alternative tools to detect and inhibit growth of pathogen bacteria. *Listeria monocytogenes* is a foodborne pathogen that causes listeriosis, a serious invasive disease that affects both humans and a wide range of animals. *Listeria* spp. are ubiquitous in the dairy farm environment and could be present in dairy-processing plants and wastewater. All Listeria-specific bacteriophages found to date are members of the Caudovirales, of the Siphoviridae or Myoviridae families. Myophages infecting Listeria have been recently classified by the ICTV in the Spounavirinae subfamily, as well as in the P100 virus genus. The aim of this work was to isolate *Listeria* spp. bacteriophages and their endolysin codifying genes from wastewater of a dairy industry. Wastewater with and without treatment was sampled during the course of a year, and isolation of bacteriophages was performed after an enrichment step using as hosts *L. innocua, L. ivanovii,* and *L. monocytogenes* serotypes 1/2a, 1/2b, and 4b. Bacteriophages infecting *L. innocua* and *L. ivanovii* were isolated ($n = 24$) from 3 out of 12 samples. Bacteriophages were purified, and the host range was determined using spot test and EOP against five collection strains and several field isolates of *Listeria* spp. Two bacteriophages of narrow and broad host range, vB_Lino_VEfB7, and vB_Liva_VAfA18, were selected for further characterization. High titer stocks of bacteriophages were purified by centrifugation with ammonium acetate, and morphological information on the purified bacteriophages was obtained by negative staining and transmission electronic microscopy. Their morphology, size, and contractile tails indicated that these bacteriophages belonged to the Myoviridae family. Bacteriophage genomes were extracted using phenol-chloroform, followed by ethanol precipitation, and tested by digestion with RNAsa A and DNAse I. RFLP was performed, digesting genomes with restriction enzymes HindIII and NcoI. Consistent with the morphological findings, bacteriophages contained dsDNA genomes but showed different RFLP patterns. A PCR designed to amplify conserved domains of endolysins—PGRP and CwlA—was applied to characterize this gene. Another PCR was designed to amplify the complete endolysin gene, and the complete sequence of this gene was obtained and analyzed. Substitution model selection and a maximum likelihood phylogenetic tree of the endolysin gene was carried out using IQ-Tree software. The sequences of the endolysin gene indicated that the codified enzyme is an N-acetyl-muramoyl-L-alanine amidase, related to A511 and P100 species of the recently described P100virus genus. Further evolutionary analyses are needed to evaluate their belonging to this species or their taxonomy within this genus.

A66 Tracing the evolutionary history of an emerging Salmonella 4,[5],12:i:- clone in the United States

E. Elnekave,[1] S. Hong,[1] A. Taylor,[2] D. Boxrud,[2] A. Rovira,[3] and J. Alvarez[4]

[1]Department of Veterinary Population Medicine, College of Veterinary Medicine, University of Minnesota, Saint Paul, MN, USA, [2]Minnesota Department of Health, St Paul, MN, USA, [3]Veterinary Diagnostic Laboratory, College of Veterinary Medicine, University of Minnesota, St Paul, MN, USA and [4]Centro de Vigilancia Sanitaria Veterinaria (VISAVET), Universidad Complutense de Madrid, Madrid, Spain

Salmonellosis is one of the leading causes of foodborne disease worldwide, with an estimated one million cases a year in the United States. Salmonella 4,[5],12: i:-, a monophasic variant of *Salmonella typhimurium*, is an emerging serovar that has been