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.