full-genome sequences were obtained, originating from four different islands and dating from the start of the outbreak (December 2013) to April 2015, when the outbreak was waning. High similarity (>99%) between sequences was found; nevertheless, all genome sequences were unique with a minimum of three SNPs differentiating one sequence from another. Thirty-three unique single nucleotide polymorphisms (SNPs) were identified, of which 29 were located in the coding regions of the genome. Eight SNPs were informative, and ten SNPs led to amino acid changes. Of the amino acid changes, nine were located in the non-structural proteins (1× nsP1, 5× nsP2, and 3× nsP3), and one was located in E2. In conclusion, we report the first whole-genome sequences of CHIKV isolates from the 2013 to 2015 outbreak that originated from the Dutch Caribbean islands. Sequencing of the remaining samples is still in progress.

A37 Transmission success of dengue virus type 1 lineages in a dynamic virus population: An evolutionary view

Carmen Koo,[1] Helen Xu,[1] Jayanthi Rajarethinam,[1] Yee Ling Lai,[1] Lee-Ching Ng,[1,2] and Hapuarachchige Chanditha Hapuarachchi[1]

[1]Environmental Health Institute, National Environment Agency, 11, Biopolis Way, #06-05-08, Singapore 138667, Singapore and [2]School of Biological Sciences, Nanyang Technological University, Singapore, Singapore

Arbovirus transmission involves an interplay between host, virus, and environmental factors. Because of the complexity of interactions, the transmission success of arboviruses could either be a function of viral fitness or be stochastic. In the present study, using 1,963 envelope (E) gene sequences and 239 whole genomes, we conducted a large-scale molecular epidemiological analysis of a dengue virus type 1 (DENV-1) population to understand the transmission success, evolution, and dispersal patterns of different lineages of DENV-1 circulating in Singapore from 2011 to 2016. The study population was highly dynamic and heterogeneous. However, only a handful of genetically distinct strains ($n = 6$) established sustained transmission, but at variable levels of dominance. Phylogeographic analysis revealed a weak spatial clustering and 35 well-supported diffusion pathways, implying widespread and complex dispersal of these strains in local settings. Yet, the dominant strains were neither evolving faster than less dominant ones nor under positive selection. These observations suggested that lineage dominance was likely to be stochastic and opportunistically driven by non-viral factors such as host immune pressure and vector abundance. Our findings, therefore, emphasize the implications of understanding the vector and human factors in parallel to virus dynamics on continuing efforts to control the arbovirus disease transmission in endemic regions.

A38 Genomic epidemiology quantifies gaps in Aedes-borne virus transmission in the Americas

Darlan da Silva Cândido,[1,2] Oliver G. Pybus,[2] and Nuno Rodrigues Faria[2]

[1]Radcliffe Department of Medicine, University of Oxford, Oxford, UK and [2]Department of Zoology, University of Oxford, Oxford OX1 3SY, UK

The rapid spread and severity of pathogens, such as Zika (ZIKV) and Chikungunya (CHIKV) viruses in the Americas, demonstrate the need for a better understanding of when and where outbreaks emerge. Sequence evolution of these viral pathogens occurs simultaneously with geographic spread, which allows phylodynamic processes to be recovered from genomic data. Here, we used time-calibrated phylogeographic analyses implemented in a Bayesian phylogenetic framework to characterize the date of introduction of ZIKV, CHIKV, dengue, and yellow fever viruses in different geographic regions of the Americas. To estimate 'surveillance gaps', we compared the estimated dates of introduction of these pathogens to the first confirmations of virus circulation in the region. Datasets included all publicly available geo-referenced and time-stamped genetic data from the Americas. A series of environmental and ecological covariates will be tested to infer what factors are associated with the delayed detection of arbovirus transmission in each geographic region. These results will provide important information on where to concentrate surveillance strengthening measures in order to prevent future mosquito-borne virus epidemics.

A39 Reconstruction of Ebola chains of transmission using sequence and epidemiological data

A. Robert,[1] J. Edmunds,[1] R. Eggo,[1] A.-M. Henao-Restrepo,[2] P.-S. Gsell,[2] C. H. Watson,[1] I. M. Longini,[3] A. Rambaut,[4,5] A. Camacho,[1,6] and S. Hué[1]

[1]Department of Infectious Disease Epidemiology, Centre for the Mathematical Modelling of Infectious Diseases, London School of Hygiene & Tropical Medicine, London, UK, [2]World Health Organization, Geneva, Switzerland, [3]Department of Biostatistics, University of Florida, Gainesville, FL, USA, [4]Institute of Evolutionary Biology, University of Edinburgh, Edinburgh, UK, [5]Ashworth Laboratories, Centre for Immunology, Infection and Evolution, University of Edinburgh, King's Buildings, Edinburgh EH9 3JT, UK and [6]Epicentre, Paris, France

Transmission trees can be established through detailed contact histories, statistical inference, phylogenetic inference, or a combination of methods. Each method has its limitations, and the extent to which they succeed in revealing a 'true' transmission history remains unclear. Moreover, the net value of pathogen sequencing in transmission tree reconstruction is yet to be assessed. We explored the accuracy and sensitivity to biases of a range of methods for transmission chain inference. We studied eight transmission chains determined by contact tracing, each one having more than a third of its cases sequenced (87 samples over 199 cases in total). We compared three inference methods on the selected transmission chains: (i) phylogenetic inference: the Ebola virus (EBOV) sequences derived from patients were mapped onto a dated EBOV phylogeny tree including 398 EBOV sequences sampled in Guinea between March 2014 and October 2015; (ii) statistical inference: we used the maximum likelihood framework developed by Wallinga and Teunis to infer the most likely transmitter-recipient relationships from the onset dates; (iii) combined method: we inferred probabilistic transmission events using both pathogen sequences and collection dates with the R package Outbreaker2. The cases coming from each transmission chain were mostly clustered together in the phylogenetic tree. The few misclassified cases were most likely allocated to the wrong chains of transmission because of the timing of their symptom onsets. Probabilistic transmission tree using only onset dates broadly matched the contact tracing data, but multiple potential infectors were identified for each case. The combined method showed that an *a priori* knowledge of the number of independent imports had an important impact on the outcome. Although cases were allocated to the correct transmission chains, discrepancies were found in identifying direct case linkage and transmission generations within a chain. Phylogenetic, epidemiological, and combined approaches for transmission chain reconstructions globally concurred in their output. Sequence data proved useful (if not necessary) to place the sampled cases in a wider context, identify transmission clusters, and misclassified cases when epidemiological chains are inferred from date of symptom onset only, and to identify links between supposedly independent chains of transmission.

A40 Estimation of Lassa virus emergence in Upper Guinea through a time-calibrated phylogeny

Elisabeth Fichet-Calvet,[1] N'Faly Magassouba,[2] and Stephan Günther[1]

[1]Bernhard-Nocht Institute of Tropical Medicine, Hamburg, Germany and [2]Laboratoire des Fièvres Hémorragiques Virales, Conakry, Guinea

Lassa fever is a hemorrhagic fever caused by an arenavirus, the Lassa virus (LASV), and can affect 150–200,000 persons per year in West Africa. The virus is hosted by several rodents, *Mastomys natalensis* and *M. erythroleucus*, *Hylomyscus pamfi*, and *Mus baoulei*. People can be contaminated at home or in the farms, by touching contaminated surfaces, eating contaminated food, or breathing aerosolized viral particles. Human-to-human transmission is occurring as well through infected bodily fluids. In Upper Guinea in particular, *M. natalensis* is the main host, with LASV prevalence of 14 per cent and IgG prevalence of 27 per cent. In humans, IgG prevalence is 40 per cent. This is, therefore, a hot spot for LASV transmission. In a previous phylogenetic study including 132 partial nucleoprotein (NP) sequences isolated from rodents, we showed that LASV could have emerged 90 years ago in the area. Here, we aim to revise the time of emergence upon analyzing the complete NP and polymerase genes of two strains coming from Upper Guinea: 'Bantou 366', a strain isolated from *M. natalensis* in

2003, and 'Faranah', a strain isolated from a human in 1996. They were aligned with 22 other LASV sequences belonging to all lineages and dated by their day of collection. In BEAST (v1.10) tree reconstruction, the following settings were used: GTR+gamma distributed rate variation (four discrete categories) across each codon position and constant population size demographic model. Four clock models were tested: strict, uncorrelated relaxed, random local, and fixed local. The best model was determined by comparing the resulting likelihoods using AICM model testing. Markov chain Monte Carlo (MCMC) sampling was performed for a total of 20 million states (sampling every 10,000 states) to obtain an effective sample size above 200 for all parameters. Results of MCMC sampling were examined in Tracer 1.6. The results showed that the Upper Guinea clade emerged 153 years ago when the phylogeny was reconstructed for partial NP (nt = 754, better model fit with strict clock), 208 years ago with complete NP (nt = 1,707, better model fit with random local clock), and 350 years ago with complete polymerase (nt = 6,681, better model fit with strict clock). The difference of emergence 1, 2, or 3 centuries ago, can be explained by the inclusion of some parts of the genome evolving slower than the partial NP. Therefore, the longer the sequence, the greater the divergence time. In order to have an accurate time of divergence, we suggest to use complete genes to perform a time-calibrated phylogeny.

A41 Deep sequencing of respiratory syncytial virus links viral diversity to disease severity

Inne Nauwelaers,[1,2] Tiina Talts,[2] Monica Galiano,[2] and Peter Openshaw[1]

[1]Section of Respiratory Infection, St Mary's Campus, Imperial College London, London, UK and [2]Respiratory Virus Unit, Public Health England, Colindale, UK

Respiratory syncytial virus (RSV) is a common virus that can cause bronchiolitis in infants and pneumonia in immunocompromised and elderly people. RSV belongs to the Pneumoviridae family and consists of a genome of 15 kb. Its genome contains ten genes that code for eleven proteins, with M2 coding for two different proteins in overlapping open reading frames. It is unclear why some infected children have severe disease and others have mild or asymptomatic disease. In this project, methods for complete genome sequencing of RSV via Sanger and Illumina MiSeq platforms were optimized. One hundred and twenty-four community samples (59 RSV A and 65 RSV B) from 2014 to 2018 were collected (in collaboration with the Royal College of General Practitioners) and sequenced. Samples were selected based on viral load (e.g. Ct values had to be < 30). The genotype of each sample was determined by constructing phylogenetic trees with reference sequences from all genotypes. Trees were reconstructed using the maximum likelihood method. Furthermore, Illumina sequencing was used to deep sequence seven community samples and four hospital samples that were spatiotemporally matched (obtained via Imperial College NHS Trust hospitals). Variants were studied to investigate if certain variants influence disease severity (e.g. cause mild (community samples) or severe infection (hospital samples)). Analysis so far showed that ON1 (with a seventy-two nucleotide duplication in attachment protein G) is the most common genotype in both community and hospitalized samples (90% and 75% of samples, respectively), with GA2 (without duplication) as the next most common genotype for RSV A subtypes (7% and 25%). Three per cent of community samples were of the GA5 genotype. Samples from the RSV B subgroup all belong to the BA genotypes with a 60-nucleotide duplication in G. Samples that were selected for Illumina sequencing had a Ct value between 19.0 and 29.1, while hospital samples had a Ct value of 18.3 to 29.1. Viral load, therefore, did not explain disease severity in these selected samples. The Shannon entropy from Illumina sequenced samples averaged at 22.78 in community samples (ranges from 15 to 28) and 38.78 in hospitalized samples (ranges from 31 to 57). This indicated that diversity of the virus pool might influence disease severity; however, more samples need to be analyzed. There are no specific variants that could explain disease severity. Diversity of the virus pool could explain the link between higher viral loads and disease severity, which is sometimes found but cannot always be confirmed. Higher viral loads can harbor more diverse viral particles compared to lower viral loads. Future work will focus on more in-depth variation and diversity analysis and on evolutionary analysis of both community and hospital samples. We will also investigate intra-host evolution of RSV in acute infections using consecutive samples and its possible implications on the host response.

A42 Next-generation sequencing to analyze multiple-strain infections, genotype distribution, and antiviral resistance in hematopoietic stem cell transplantation recipients with human cytomegalovirus infection

A. Dhingra,[1,2] N. Suárez,[3] P. Varanasi,[2,4] J. Zischke,[1,2] A. Heim,[1,2] E. Mischak-Weissinger,[2,4] T. Schulz,[1,2] A. Davison,[3] P. Kay-Fedorov,[1,2] and T. Ganzenmueller[1,2]

[1]Hannover Medical School, Institute of Virology, Hannover, Germany, [2]German Center for Infection Research (DZIF), Site Hannover-Braunschweig, Germany, [3]MRC-University of Glasgow, Centre for Virus Research, Glasgow, UK and [4]Department of Haematology, Haemostasis and Oncology, Hannover Medical School, Hannover, Germany

Next-generation sequencing (NGS) produces comprehensive insights across the entire genome of the human cytomegalovirus (HCMV), which is an important opportunistic pathogen following hematopoietic stem cell transplantation (HSCT). To assess the clinical impact of HCMV diversity, genotype distribution, and resistance mutations, we performed NGS directly on plasma specimens from HSCT recipients with HCMV reactivation. Twenty-nine HCMV-positive plasma samples (median viral load 1.7×10^3 IU/ml) collected from a prospective allogenic HSCT recipient cohort ($n = 16$) between 21 and 80 days after transplantation were sequenced on an Illumina MiSeq after preparation of target-enriched sequencing libraries. Consensus HCMV genome sequences were assembled for 24 samples. The presence of multiple-strain infections and antiviral resistance mutations in genes UL54 and UL97 was determined by variant analysis. Genotype distribution was determined by specific marker analysis of several hypervariable genes (RL5A, RL6, RL12, RL13, UL1, UL9, UL11, UL73, UL74, UL120, UL146, and UL139). Associations between genomic and clinical features (e.g. graft-versus-host disease (GvHD), donor/recipient HCMV serostatus, dynamics of HCMV antigenemia, survival) were explored. Multiple infections involving up to 3 HCMV strains were detected in seven out of sixteen patients, with one patient analyzed at > 2 time points, showing a switch of the dominant HCMV population. No known antiviral resistance mutations were detected, which may be expected due to sample collection early after HSCT from patients without antiviral prophylaxis. Multiple-strain infection was associated with an earlier peak of HCMV-antigenemia ($P = 0.054$), but not with duration of viremia, antigenemia peak values, donor/recipient HCMV serostatus, T-cell depletion, acute or chronic GvHD, disease relapse, or reduced survival. Genotype distribution analysis revealed a potential link of one genotype of the immunomodulatory gene UL11 with GvHD incidence after HCMV reactivation. NGS of HCMV diversity directly from plasma samples, even with low viral loads, enables the acquisition of data of potential clinical interest. To identify reliable associations between clinical features and HCMV diversity, further patient cohorts with suitable sample sizes are required.

A43 Translational research: NGS metagenomics into clinical diagnostics

D. Schmitz,[1,2] S. Nooij,[1,2] T. Janssens,[1] J. Cremer,[1] H. Vennema,[1] A. Kroneman,[1] and M. Koopmans[2]

[1]National Institute for Public Health and the Environment (RIVM), Bilthoven, the Netherlands and [2]Erasmus Medical Centre (EMC), Rotterdam, the Netherlands

As research next-generation sequencing (NGS) metagenomic pipelines transition to clinical diagnostics, the user-base changes from bioinformaticians to biologists, medical doctors, and lab-technicians. Besides the obvious need for benchmarking and assessment of diagnostic outcomes of the pipelines and tools, other focus points remain: reproducibility, data immutability, user-friendliness, portability/scalability, privacy, and a clear audit trail. We have a research metagenomics pipeline that takes raw fastq files and produces annotated contigs, but it is too complicated for non-bioinformaticians. Here, we present preliminary findings in adapting this pipeline for clinical diagnostics. We used information available on relevant fora (www.bioinfo-core.org) and experiences and publications from colleague bioinformaticians in other institutes (COMPARE, UBC, and LUMC). From this information, a robust and user-friendly storage and analysis workflow was designed for non-bioinformaticians in a clinical setting. Via Conda [https://conda.io] and Docker containers [http://www.docker.com], we made our