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

disparate pipeline processes self-contained and reproducible. Furthermore, we moved all pipeline settings into a separate JSON file. After every analysis, the pipeline settings and virtual-environment recipes will be archived (immutably) under a persistent unique identifier. This allows long-term precise reproducibility. Likewise, after every run the raw data and final products will be automatically archived, complying with data retention laws/guidelines. All the disparate processes in the pipeline are parallelized and automated via Snakemake1 (i.e. end-users need no coding skills). In addition, interactive web-reports such as MultiQC [http://multiqc.info] and Krona2 are generated automatically. By combining Snakemake, Conda, and containers, our pipeline is highly portable and easily scaled up for outbreak situations, or scaled down to reduce costs. Since patient privacy is a concern, our pipeline automatically removes human genetic data. Moreover, all source code will be stored on an internal Gitlab server, and, combined with the archived data, ensures a clear audit trail. Nevertheless, challenges remain: (1) reproducible reference databases, e.g. being able to revert to an older version to reproduce old analyses. (2) A user-friendly GUI. (3) Connecting the pipeline and NGS data to in-house LIMS. (4) Efficient long-term storage, e.g. lossless compression algorithms. Nevertheless, this work represents a step forward in making user-friendly clinical diagnostic workflows.

A44 Genome analysis of bovine enterovirus variants isolated from cattle in Thailand

N. Kosoltanapiwat,[1] N. Income,[1] D. Cadar,[2] J. Schmidt-Chanasit,[2] and J. Tongshoob[1]

[1]Department of Microbiology and Immunology, Faculty of Tropical Medicine, Mahidol University, Thailand and [2]Bernhard Nocht Institute for Tropical Medicine, Germany

Bovine enteroviruses (BEV) are non-enveloped RNA viruses of the genus Enterovirus, family Picornaviridae, which are commonly found in cattle. They have been classified into two species, enterovirus E (EV-E) and enterovirus F (EV-F). The viruses were previously considered non-pathogenic, but recent evidences suggest their association with pathogenesis in cattle. BEV-like enteroviruses have also been increasingly isolated from a wide range of animals, such as sheep, goats, horses, geese, possum, and deer, from many countries. The isolation and characterization of novel enteroviruses expands the range of the genus. Our data show that both EV-E and EV-F are circulating in cattle in Thailand. The viruses have been detected in 35–67 per cent of dairy and meat cattle feces in Kanchanaburi Province. Recently, we retrieved EV-E isolates from cattle feces by virus isolation in Madin-Darby Bovine Kidney cells. Four virus isolates were subjected to whole-genome sequencing using Illumina next-generation sequencing. A phylogenetic analysis of VP1 capsid protein, which is used for virus genotyping, suggested that there are at least two EV-E genotypes circulating in cattle in the area of study. Two virus strains, closely related to EV-E1 with amino acid sequence identities >88 per cent were identified as EV-E1. The other two strains, closely related to EV-E2 with amino acid sequence identities < 85 per cent, were likely to constitute a new EV-E genotype separate from the existing EV-E2.

A45 Genetic diversity of anelloviruses in the blood virome

Marijn Thijssen,[1] Leen Beller,[2] Kwe Claude Yinda,[2] Ward Deboutte,[2] Piet Maes,[1] Jelle Matthijnssen,[2] Marc Van Ranst,[1] and Mahmoudreza Pourkarim[1,3]

[1]Laboratory of Clinical Virology, Department of Microbiology and Immunology, KU Leuven-University of Leuven, Rega Institute, Leuven, Belgium and [2]Laboratory of Viral Metagenomics, Department of Microbiology and Immunology, KU Leuven-University of Leuven, Rega Institute, Leuven, Belgium

The microbiome has an important impact on human health. The microbiome is a complex ecosystem that contains of a wide variety of microorganisms shaped by the immune system, host genetic factors, and the environment. Studies of the human virome have identified a diverse group of viruses in different compartments of the body, including viruses of the Anelloviridae family. These viruses are widespread among the general population. In various clinical conditions an association has been found between the Anelloviridae abundance and the patient's immune status. However, no pathological consequences have been identified for this viral family. In this study, we analyzed the Anelloviridae diversity in plasma samples of liver transplant recipients. The virome contents of plasma samples from liver transplant recipients were sequenced by next-generation sequencing techniques on an Illumina platform (NextSeq). Complete Anelloviridae ORF1 contigs were extracted from metagenomic data and aligned with 66 RefSeq anellovirus sequences for phylogenetic analysis. The study included 144 plasma samples of 24 liver transplant recipients who had been infected by the hepatitis B virus and developed end-stage liver disease. The identified Anelloviridae viruses belong to the Alphatorquevirus, Betatorquevirus, and Gammatorquevirus genera. In total, we were able to retrieve 142 unique anellovirus contigs that were less than 95 per cent identical on the nucleotide level. A phylogenetic tree was constructed from these contigs with 65 RefSeq sequences retrieved from GenBank. The majority of the identified Anelloviridae sequences were assigned to the Alphatorquevirus genus, which represents the largest group of anelloviruses. We were able to identify a high diversity of Anelloviridae viruses in serum samples of liver transplant recipients. Phylogenetic analysis showed that the majority of anelloviruses belonged to the Alphatorque genus. Future research should focus at elucidating the role of these commensal viruses in both immunocompromised and healthy individuals.

A46 Hand, foot, and mouth disease in Vietnam

Le Nguyen Truc Nhu,[1] Hoang Minh Tu Van,[1] Le Nguyen Thanh Nhan,[2] Nguyen To Anh,[1] Tran Tan Thanh,[1] Vu Thi Ty Hang,[1] Nguyen Thi Han Ny,[1] Nguyen Thi Thu Hong,[1] Nguyen Thanh Hung,[2] Truong Huu Khanh,[2] Du Tuan Quy,[2] Ha Manh Tuan,[3] Ho Lu Viet,[3] Do Chau Viet,[3] Nguyen Tran Nam,[3] Nguyen Thi My Thanh,[4] Saraswathy Sabanathan,[1] Phan Tu Qui,[1,4] Nguyen Van Vinh Chau,[4] Guy E. Thwaites,[1,5] Bridget Wills,[1,5] C. Louise Thwaites,[1,5] H. Rogier van Doorn,[1,5] and Le Van Tan[1]

[1]Oxford Clinical Research Unit, Ho Chi Minh City, Vietnam, [2]Children's Hospital 1, Ho Chi Minh City, Vietnam, [3]Children's Hospital 2, Ho Chi Minh City, Vietnam, [4]Hospital of Tropical Diseases, Ho Chi Minh City, Vietnam and [5]Nuffield Department of Medicine, Centre for Tropical Medicine, University of Oxford, Oxford, UK

Hand, Foot, and Mouth Disease (HFMD) is a major public health issue in the Asia-Pacific region. Our research program aims to address unanswered questions about clinical, epidemiology, pathogen evolution, cost of illness, and host-genetic makers associated with severe HFMD in Vietnam. A multi-hospital-based observational study has been conducted at three referral hospitals in Ho Chi Minh City, Vietnam since 2013. Demographic, clinical data, and cost of illness were collected alongside clinical specimens. Multiplex PCR and next-generation sequencing were employed to identify enterovirus serotypes and to study pathogen evolution, respectively. A genome-wide association-based approach was used to explore genetic markers of disease severity. From 2013 to 2017, 2,191 HFMD patients were enrolled. More than twenty enterovirus serotypes were detected in 84.3 per cent of patients. EV-A71 was the major cause, accounting for 22 per cent of total number of cases, followed by CV-A6 (21%), CV-A16 (13%), and CV-A10 (8%). Interestingly, these four common enteroviruses replaced each other during the study period. EV-A71 and CV-A6 were the two most predominant viruses detected in 2013 and 2014. However, CV-A6 was replaced by CV-A16 and CV-A10 in 2015 and 2016, respectively. A total of 396 whole-genome sequences (EV-A71 ($n = 200$), CV-A6 ($n = 98$), CV-A10 ($n = 66$), and CV-A16 ($n = 32$)) were obtained. Phylogenetic analysis showed that EV-A71 subgenogroup B5 has replaced C4 in 2012, and, since then, B5 has continued to circulate predominantly, while C4 has been sporadically detected. All Vietnamese CV-A6 isolates belonged to genogroup A, which has caused large outbreaks of HFMD worldwide. Costs of illness varied between disease severities, ranging from $USD 244 [95% confidence interval (95% CI): 230–258] per patient for grade 2A (mild) to $USD 1984 (95% CI: 1,752–2,227) for grade 3 (severe). The genome-wide association study identified two genetic markers potentially associated with severe HFMD. The results highlight that active surveillance and understanding pathogen evolution are essential to inform public health in prioritizing the development of intervention strategies. Efforts to unravel the evolutionary process of Vietnamese CV-A10 and CV-A16 in relation to global strains are ongoing. An independent cohort is needed to replicate the preliminary findings of the genome-wide association study.