serological testing; increased genetic information from diverse sources should enable PCR design that will reliably detect a wide range of strains thus facilitating diagnosis and epidemiological analysis of outbreaks.

A64 Overview of virus metagenomics classification tools

Sam Nooij,[1,2] Dennis Schmitz,[1,2] Annelies Kroneman,[1] Harry Vennema,[1] and Marion Koopmans[1,2]

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

The use of next-generation sequencing for discovery of viruses has yielded vast amounts of known and putative viral reads. The computational analysis of the reads, however, is quite a challenge, in particular the classification of reads to viral taxa. This is illustrated by the many computational tools that have been devised and new tools that appear monthly. These provide opportunities for other researchers, but the large numbers make it hard for virologists to pick a tool that suits their own study. To facilitate this choice and guide users to through the forest of computational pipelines, we have surveyed publications describing fifty tools, inventoried their approaches and scored their methods, user-friendliness, validation, and other performance criteria for diagnostics, outbreak source tracing, virus discovery, and virome profiling. The tools have variable approaches to the classification of viral reads, but they all rely on searching (i.e., homology (44/50 pipelines) and composition search (8/50)) through reference databases (e.g., nucleotide databases, protein databases, or virus-specific databases). Some pipelines include quality control/pre-processing of reads (23/50), filtering non-viral reads (20/50), and de novo assembly (18/50) before the search, and checking and correcting classifications with phylogenetic or statistical methods after the search step (8/50). Furthermore, some are tailor-made for particular studies; others are more generally applicable. Few tools provide a graphical user interface, and when they do they are often online, which increases the ease of use. Reported runtimes vary greatly—from several minutes per sample to days; newer tools are often faster than old ones. Moreover, some tools have been validated in wet-lab experiments or compared to other tools with in silico benchmark tests. The overview of pipelines is presented on the COMPARE website (https://compare.cbs.dtu.dk/inventory#pipeline). A decision tree is provided separately to help virologists with any level of bioinformatics expertise select suitable analysis tools. The next step will be to benchmark the most promising tools with the COMPARE and VIROGENESIS projects to better assess their performance for diagnostics and surveillance studies.

A65 Mitochondrial DNA studies of Lisbon immigrants from Portuguese speaking African countries

A. Amorim,[1,2,3,4] A. Afonso Costa,[1,5] C. Vieira da Silva,[1] T. Ribeiro,[1] M. J. Porto,[1] N. Taveira,[3,6] and T. Fernandes[4]

[1]Instituto Nacional de Medicina Legal e Ciências Forenses, Portugal, [2]Faculdade de Ciências da Universidade de Lisboa, Lisboa, Portugal, [3]Instituto Superior de Ciências da Saúde Egas Moniz, Almada, Portugal, [4]Escola de Ciências e Tecnologia da Universidade de Évora, Évora, Portugal, [5]Faculdade de Ciências da Saúde da Universidade da Beira Interior, Covilhã, Portugal and [6]Faculdade de Farmácia da Universidade de Lisboa, Lisboa, Portugal

Since the end of the 1970s, Portugal has had an important role in migratory movements, becoming a destination for immigrants of a wide range of nationalities, mainly from African countries. According to PORDATA, until the end of 2014 there were ~40,000 immigrants from Cape Verde, 20,000 from Angola, 18,000 from Guinea-Bissau, and 3,000 from Mozambique living in Portugal, and of those, >80 per cent live in the Lisbon region. This may be one of the main contributors to genetic variation of Lisbon residents in the present and the future. Mitochondrial DNA (mtDNA) has features that make it desirable for forensics, namely, high copy number, lack of recombination, and matrilineal inheritance. These features are also important in evolutionary and population studies. We aim to characterize mtDNA diversity in immigrants from Portuguese Speaking African Countries (PALOP) living in Lisbon and their potential contribution to genetic variation of Lisbon population. Blood samples were collected from 439 PALOP immigrants living in Lisbon, of which 173 immigrants from Angola, 103 immigrants from Cape Verde, eighty-three immigrants from Mozambique and eighty immigrants from Guinea-Bissau, from January 2000 to December 2016. The control region of the mtDNA was amplified using two pairs of primers—L15971/H016 and L16555/H639, and sequenced by BigDye Terminator v.3.1 Cycle Sequence (AB). Sequenced products were detected in a sequencer Genetic Analyzer 3130 (AB). Finally the results were analysed by Sequencing Analysis v.5.2 software and also compared with Revised Cambridge Reference Sequence (rCRS) using SeqScape v.3 (AB) software. The haplogroups were determined based on Phylotree, build 17. Genetic distances and other genetic parameters were calculated with Arlequin software ver.3.5 and analysed and represented with PhyML 3.0. For each sample, the complete sequence of the control region was obtained. The comparison of the sequences obtained with the rCRS, among the 439 analysed individuals, allowed the identification of 319 different haplotypes, corresponding to 164 different haplogroups distributed by ten macrohaplogroups. Macrohaplogroup L was the most common with 386 haplotypes followed by U with fifteen haplotypes, H with twelve haplotypes, M and T with six haplotypes, K with five, R with four, X and J with two and HV with one. PALOP's immigrants presented a high number of unique haplotypes, most of them belonging to macrohaplogroup L, originating from sub-Saharan regions of Africa. This macrohaplogroup is uncommon in European and Portuguese populations. Consistent with this, phylogenetic analysis showed the establishment of two distinct groups, one composed of the Portuguese population and another of the African populations. In comparing the different immigrant populations living in Lisbon, the genetically closest community to the Portuguese population is Mozambique and the furthest is Cape Verde, followed by Guinea-Bissau and Angola. Our results show that the PALOP immigrants living in Lisbon are genetically heterogeneous. The increase in genetic diversity in Lisbon due to immigrants from PALOP countries may have a major impact on haplotypic and allelic frequencies, on which all forensic and medico-legal investigations are based.

A66 Multi-drug-resistant *Klebsiella pneumoniae* strains circulating in hospital setting: Whole-genome sequencing and Bayesian phylogenetic analysis for outbreak investigations

E. Cella,[1,2] M. Ciccozzi,[1,3] M. Fogolari,[1] T. Azarian,[4] M. Prosperi,[5] G. Dicuonzo,[1] M. Salemi,[6] and S. Angeletti[1]

[1]Unit of Clinical Pathology and Microbiology, University Campus Bio-Medico of Rome, Italy, [2]Department of Public Health and Infectious Diseases, Sapienza University of Rome, Rome, Italy, [3]Department of Infectious Parasitic and Immune-Mediated Diseases, Istituto Superiore di Sanità, Rome, Italy, [4]Department of Epidemiology, Center for Communicable Disease Dynamics, Harvard's T.H. Chan School of Public Health, Boston, MA, [5]Department of Epidemiology, University of Florida, Gainesville, FL, USA and [6]Department of Pathology, Immunology, and

Laboratory Medicine, Emerging Pathogens Institute, University of Florida, Gainesville, FL, USA

Carbapenems-resistant Enterobacteriaceae infections are increasing worldwide representing an emerging public health problem. The application of phylogenetic and phylodynamic analyses to bacterial whole-genome sequencing data have become essential in the epidemiological surveillance of multi-drug-resistant nosocomial pathogens. Between January 2012 and February 2013, twenty-one multi-drug-resistant *K. pneumoniae* strains, were collected from patients hospitalized among different wards of the University Hospital Campus Bio-Medico. Epidemiological contact tracing of patients and Bayesian phylogenetic analysis of bacterial whole-genome sequencing data were used to investigate the evolution and spatial dispersion of *K. pneumoniae* in support of hospital infection control. The epidemic curve of incident *K. pneumoniae* cases showed a bimodal distribution of cases with two peaks separated by forty-six days between November 2012 and January 2013. The time-scaled phylogeny suggested that *K. pneumoniae* strains isolated during the study period may have been introduced into the hospital setting as early as 2007. Moreover, the phylogeny showed two different epidemic introductions in 2008 and 2009. Bayesian genomic epidemiology is a powerful tool that promises to improve the surveillance and control of multi-drug-resistant pathogens in an effort to develop effective infection prevention in healthcare settings or constant strains reintroduction.

A67 Use of next-generation whole-genome sequencing to understand drug-resistant *Mycobacterium tuberculosis* in Botswana

T. Iketleng,[1,2] T. Mogashoa,[1] B. Mbeha,[3] L. Letsibogo,[3] J. Makhema,[1] S. Moyo,[1] T. de Oliveira,[2] and S. Gaseitsiwe[1]

[1]Botswana Harvard AIDS Institute Partnership for HIV Research and Education, Gaborone, Botswana, [2]University of KwaZulu-Natal, South Africa and [3]Botswana National TB Reference Laboratory, Gaborone, Botswana

The need to curb the emergence and spread of *M. tuberculosis* drug resistance requires exploration of new methods especially in high incidence countries. Effective treatment informed by timely and accurate drug susceptibility testing is critical to control drug-resistant tuberculosis. Next-generation whole-genome sequencing is increasingly becoming attractive and could potentially provide a fast and comprehensive determination of drug susceptibility that could inform timely treatment decisions. The aim of this study is to use next-generation whole-genome sequencing analysis to understand *M. tuberculosis* drug resistance, population structure, and transmission dynamics in Botswana. The study will be a retrospective cross sectional study using isolates from adults diagnosed with drug-resistant pulmonary tuberculosis by culture based drug susceptibility testing at Botswana National Tuberculosis Reference Laboratory in Gaborone, Botswana. A total of 150 isolates with a spectrum of resistance ranging from mono-resistance to multidrug-resistant tuberculosis (MDR-TB) will be selected for next-generation whole-genome sequencing using the Illumina MiSeq® system (Illumina, San Diego, California). FASTQ files or Unmapped Binary Alignment Map (BAM) files of the reads generated will be assembled and aligned using Geneious, Genome Analysis Toolkit (GATK), and SamTools. Alignment output for each sample will then be indexed, sorted, and merged into a single alignment file using SamTools. SNP calling will be done using Genius, GATK, and SamTools; the SNPs detected by at least two methods will be deemed real and compared to a TB drug resistance mutation list (http://pfgrc.jcvi.org/). The list includes 1,086 SNPs identified in thirty *M. tuberculosis* genes.

A68 Bats as a source of zoonotic spillover: Investigating viruses in enteric bat samples from Viet Nam

A. O'Toole,[1,2] C. Bogaart,[1,2] L. Lu,[1,2] J. Ashworth,[1,2] and M. E. Woolhouse[1,2]

[1]University of Edinburgh, Edinburgh, UK and [2]VIZIONS Consortium

Bats, belonging to the order Chiroptera, are the natural reservoir hosts for an array of zoonotic viruses. Aspects of bat ecology, behavior, and physiology, such as migrating great distances, roosting in close association in large numbers and variation in metabolic rate and core body temperature during sustained flight, make them a unique concern for viral zoonotic transfer. As part of the Viet Nam Iniative on Zoonotic Infections (VIZIONS) project, 169 enteric bat samples were collected from two sites in the Dong Thap province in Viet Nam, ~100 km apart, and Illumina sequenced at the Sanger Institute. Based on host mtDNA sequence present in these enteric samples, we identified *Scotophilus kuhlii* as the host species for > 97 per cent of the 196 samples, with remaining samples of the *Myotis*, *Murina*, or *Pipistrellus* genera. Significant quantities of Alphacoronavirus, Rotavirus, and Mamastrovirus reads were identified in the enteric bat samples using Kraken. We confirmed significant mixing and jumping of Alphacoronavirus between the two locations, using two independent analyses: Bayesian Tip-associated Significance testing, which confirms no significant clustering of virus with respect to location, and host-state reconstruction analysis, which predicted a mean of seventeen host-jumps between the two locations. These findings suggest that 100 km is negligible for *Scotophilus kuhlii* in terms of viral transfer.