A new SIV lineage was identified in Allen's swamp monkeys (*Allenopithecus nigroviridis*). Three new STLV-1 subtypes were identified in Allen's swamp monkeys (Allenopithecus nigroviridis), blue monkeys (*Cercopithecus mitis*), red-tailed guenons (*Cercopithecus ascanius schmidti*), and agile mangabeys (*Cercocebus agilis*). SIV and STLV prevalences varied according to species and geographic region. Our study illustrates clearly, even on a small sample size from a limited number of geographic areas, that our knowledge on the genetic diversity and geographic distribution of simian retroviruses is still limited, and that humans continue to be exposed to relative high proportions of infected NHP bushmeat.

A62 Molecular characterization of emerging variants of bovine leukemia virus

Aneta Pluta Marzena Rola-Łuszczak and Jacek Kuźmak

OIE Reference Laboratory for EBL, Department of Biochemistry, National Veterinary Research Institute, Pulawy, Poland

Bovine leukemia virus (BLV) is the etiological agent of the enzootic bovine leucosis. The BLV transcriptional promoter is located in 5′ long terminal repeat (LTR), which is composed of the U3, R, and U5 regions. One of the limiting factors in the study of BLV is the lack of sequence data in relation to geographical origin of strains and the potential effects of genetic variability on virus infectivity and disease progression. We would like to address this question by analysing LTR variability of BLV variants isolated from emerging cases of infection, recorded in already-cured herds. Our hypothesis is that the genetic variability in the LTR sequences of these isolates can be associated with transcriptionally down regulated provirus variants. DNA samples were isolated from PBLs of sixty cattle serologically positive for BLV, which were classified as a newly infected animals in herds having BLV free status. Additionally, twenty-five archival DNA samples were collected from East European countries: Poland, Russia, Ukraine, Moldova, and Croatia, representing endemic cases of BLV infection. Fusion PCR was developed and used for amplification of 531-bp fragment corresponding to the full length BLV LTR. Amplified fragments were sequenced and sequences were aligned using Muscle. The following steps of phylogeographic analysis were performed: (1) creation of trait information for the sequences, different molecular clock models, tree priors and parameters for the MCMC chain, (2) computation of Bayesian Markov chain Monte-Carlo simulations of 100 million steps in BEAST, (3) creation of a maximum clade credibility (MCC) tree in TreeAnnotator, (4) creation a Google Earth file of the MCC tree using SPREAD. The Shimodaira–Hasegaewa (SH) test was used to simultaneously compare sets tree topologies based on the partial env gene and LTR sequences. Comparative analysis of eighty-five sequences and sequences available from GenBank allowed the analysis of a new mutations in Geneious Pro 5.3. Sequence analysis of the BLV variants from new infections revealed 96.4 to 99.5 per cent homology, when compared to the archival DNA samples. Detailed analysis of LTR sequences showed ninety-one different mutations dispersed mainly along U3 and U5 regions. Out of these mutations, thirty-three were located in regulatory elements: Ebox1, TRE2, CRE2, Ebox2, NF-kB, PU Box, GRE, Ebox3, TATA Box, CAP site, DAS, and IRF binding site. The phylogenetic analysis of full length LTR sequences revealed the existence of three main groups (G1–G3) distinguished into two to five subgroups. Analysis of sequences belonging to G3 group and those from certain subgroups of G1 and G2 groups showed that they comprise distinct sets of BLV variants, which evolved separately. The topology of the phylogenetic maximum likelihood tree based on the env gene was in general similar to that of the maximum likelihood tree topology based on the LTR sequences. However, full matching between both trees was not been seen and revealed some conflicting nodes, according to the results of the SH test. In addition, information about the location of each sample was assigned to each sequence, and a discrete phylogeographic analysis was performed. A posterior probability distribution of trees was summarized as an MCC consensus tree with posterior probabilities of geographic areas plotted at interior nodes. The final tree was then visualized in FigTree software and the KML file successfully visualized geographical migration of BLV variants in Poland and neighboring countries. The phylogenetic analysis of LTR-related sequences analysed in this study revealed that they can be classified into major clades. LTR sequences were suitable to perform phylogeographic analysis of BLV variants circulating in Poland. This analysis can be an alternative to env gene based phylogenetic methods. Detailed analysis of LTR sequences showed ninety-one point mutations dispersed mainly along U3 and U5 regions of the LTR. About thirteen of these mutations were located within promotor contained in the U3 region of BLV and these could be associated with diminished transcriptional activity.

A63 Equine infectious anaemia virus in Great Britain: Molecular characterisation of cases from 1975 to 2012

B. Choudhury,[1] D. Robinson,[1,2] and F. Steinbach[1,2]

[1]Animal and Plant Health Agency (APHA)—Weybridge, New Haw, Surrey KT15 3NB, UK and [2]University of Surrey (UNIS), Guildford Surrey GU2 7XH, UK

Equine infectious anaemia virus (EIAV) is the aetiological agent of equine infectious anaemia (EIA), a notifiable disease within the EU and to the OIE. In Great Britain, EIA cases were reported in 1975–1978; then followed a disease-free period until 2010–2012, when six positive animals were identified. As a member of the family Retroviridae and genus *Lentivirus*, EIAV shares the lentiviral trait of considerable genetic variation which has hampered development of molecular diagnostics. In addition, there is a lack of publically available sequence data, hence our intention was to conduct molecular characterisation of the British cases. A combination of Sanger and next-generation sequencing (NGS) was employed in order to determine viral sequences. Using Sanger sequencing only small genomic fragments could be recovered, hence efforts were refocused on NGS. Phylogenetic analysis and evaluation of sequence diversity of the British sequences were assessed using the MEGA5 software. Full-genome sequences were obtained from symptomatic cases in 1975, 2010, and 2012. Phylogenetic analysis of these sequences revealed that each British case formed its own branch on the tree as did sequences from America, China, Ireland, and Japan. Almost an equal distance was observed between each of the isolates, with nucleotide homology of 75–79 per cent. Nucleic acid identity of the full-genome sequences varied between individual genes and ranged from 46 per cent (P9) to 98 per cent (protease). EIAV's variable nature made the use of primer walking sequencing strategies laborious, whereas attaining full genome via NGS was relatively straightforward. However, it is not without its problems as significant viral load is required to overcome the high host background typical of clinical extractions rendering sequencing of asymptomatic cases problematic. As the British asymptomatic cases provided limited or no sequence, investigation is ongoing into strategies to elucidate these sequences. Currently EIAV diagnosis relies solely on

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