Relationships were consistently weak between true and estimated rates of extinction and of diversification. Overall, we suggest that results from this approach should be interpreted with considerable caution.

A63 Quantifying the dynamics of evolutionary rates through time

J. V. Membrebe,[1] G. Baele,[1] M. A. Suchard,[2,3] and P. Lemey[1]

[1]Department of Microbiology and Immunology, Rega Institute, KU Leuven, Leuven, Belgium, [2]Departments of Biomathematics and Human Genetics, David Geffen School of Medicine at UCLA, University of California, Los Angeles, CA 90095, USA and [3]Department of Biostatistics, UCLA Fielding School of Public Health, University of California, Los Angeles, CA 90095, USA

The availability of evolutionary rate estimates in recent years led to the observation that they may depend on the time-scale on which they are measured. Specifically, RNA virus evolutionary rates are frequently estimated to be low towards the past and high towards the present. This time-dependent rate phenomenon (TDRP) has important implications for evolutionary studies as it could severely bias divergence time estimates. While recent studies are providing insights into the relationship between viral evolutionary rate and time, formal probabilistic models to draw inference under TDRP scenarios remain lacking. Here, we adopt epoch-modelling to develop a Bayesian model of discrete rate changes through time in an unknown evolutionary history and combine this with a log-linear parameterization of rates as a function of times in the past. We provide an implementation for nucleotide substitution rates as well as for nonsynonymous rates change in a codon substitution model. Using a foamy virus dataset for which internal node calibrations can be applied based on host-virus co-divergence, we estimate a significant decline in evolutionary rates as a function of time into the past for nucleotide substitutions as well as for non-synonymous substitutions in a codon model. We also estimate a deep evolutionary history for primate Lentiviruses by combining an HIV-1 group M node calibration and a biogeographic calibration for viruses in drill monkeys in the TDRP model. Our analyses lead to the conclusion that studies of evolutionary timescales require a reconsideration of substitution rates, in either codon and nucleotide substitution model, as a dynamic feature of molecular evolution.

A64 Viral sequence classification using deep learning algorithms

David Nieuwenhuijse, Bas Oude Munnink, My Phan, and Marion Koopmans

[1]Department of Viroscience, Erasmus Medical Center, Rotterdam, the Netherlands

Sewage samples have a high potential benefit for surveillance of circulating pathogens because they are easy to obtain and reflect population-wide circulation of pathogens. These type of samples typically contain a great diversity of viruses. Therefore, one of the main challenges of metagenomic sequencing of sewage for surveillance is sequence annotation and interpretation. Especially for high-threat viruses, false positive signals can trigger unnecessary alerts, but true positives should not be missed. Annotation thus requires high sensitivity and specificity. To better interpret annotated reads for high-threat viruses, we attempt to determine how classifiable they are in a background of reads of closely related low-threat viruses. As an example, we attempted to distinguish poliovirus reads, a virus of high public health importance, from other enterovirus reads. A sequence-based deep learning algorithm was used to classify reads as either polio or non-polio enterovirus. Short reads were generated from 500 polio and 2,000 non-polio enterovirus genomes as a training set. By training the algorithm on this dataset we try to determine, on a single read level, which short reads can reliably be labeled as poliovirus and which cannot. After training the deep learning algorithm on the generated reads we were able to calculate the probability with which a read can be assigned to a poliovirus genome or a non-poliovirus genome. We show that the algorithm succeeds in classifying the reads with high accuracy. The probability of assigning the read to the correct class was related to the location in the genome to which the read mapped, which conformed with our expectations since some regions of the genome are more conserved than others. Classifying short reads of high-threat viral pathogens seems to be a promising application of sequence-based deep learning algorithms. Also, recent developments in software and hardware have facilitated the development and training of deep learning algorithms. Further plans of this work are to characterize the hard-to-classify regions of the poliovirus genome, build larger training databases, and expand on the current approach to other viruses.

A65 Characterization of endolysin gene of bacteriophages infecting *Listeria* spp. isolated from dairy industry wastewater

Blanco Fernández,[1,2] M. E. Barrios,[1,2] R. V. Cammarata,[1,2] C. Torres,[1,2] and V. A. Mbayed[1,2]

[1]Cátedra de Virología, Facultad de Farmacia y Bioquímica, Universidad de Buenos Aires, Buenos Aires, Argentina and [2]Consejo Nacional de Investigaciones Científicas y Tecnológicas, CONICIT, San José, Costa Rica

Bacteriophages and their endolysins, enzymes that degrade the cell walls of bacteria, are emerging as alternative tools to detect and inhibit growth of pathogen bacteria. *Listeria monocytogenes* is a foodborne pathogen that causes listeriosis, a serious invasive disease that affects both humans and a wide range of animals. *Listeria* spp. are ubiquitous in the dairy farm environment and could be present in dairy-processing plants and wastewater. All Listeria-specific bacteriophages found to date are members of the Caudovirales, of the Siphoviridae or Myoviridae families. Myophages infecting Listeria have been recently classified by the ICTV in the Spounavirinae subfamily, as well as in the P100 virus genus. The aim of this work was to isolate *Listeria* spp. bacteriophages and their endolysin codifying genes from wastewater of a dairy industry. Wastewater with and without treatment was sampled during the course of a year, and isolation of bacteriophages was performed after an enrichment step using as hosts *L. innocua, L. ivanovii,* and *L. monocytogenes* serotypes 1/2a, 1/2b, and 4b. Bacteriophages infecting *L. innocua* and *L. ivanovii* were isolated ($n = 24$) from 3 out of 12 samples. Bacteriophages were purified, and the host range was determined using spot test and EOP against five collection strains and several field isolates of *Listeria* spp. Two bacteriophages of narrow and broad host range, vB_Lino_VEfB7, and vB_Liva_VAfA18, were selected for further characterization. High titer stocks of bacteriophages were purified by centrifugation with ammonium acetate, and morphological information on the purified bacteriophages was obtained by negative staining and transmission electronic microscopy. Their morphology, size, and contractile tails indicated that these bacteriophages belonged to the Myoviridae family. Bacteriophage genomes were extracted using phenol-chloroform, followed by ethanol precipitation, and tested by digestion with RNAsa A and DNAse I. RFLP was performed, digesting genomes with restriction enzymes HindIII and NcoI. Consistent with the morphological findings, bacteriophages contained dsDNA genomes but showed different RFLP patterns. A PCR designed to amplify conserved domains of endolysins—PGRP and CwlA—was applied to characterize this gene. Another PCR was designed to amplify the complete endolysin gene, and the complete sequence of this gene was obtained and analyzed. Substitution model selection and a maximum likelihood phylogenetic tree of the endolysin gene was carried out using IQ-Tree software. The sequences of the endolysin gene indicated that the codified enzyme is an N-acetyl-muramoyl-L-alanine amidase, related to A511 and P100 species of the recently described P100virus genus. Further evolutionary analyses are needed to evaluate their belonging to this species or their taxonomy within this genus.

A66 Tracing the evolutionary history of an emerging Salmonella 4,[5],12:i:- clone in the United States

E. Elnekave,[1] S. Hong,[1] A. Taylor,[2] D. Boxrud,[2] A. Rovira,[3] and J. Alvarez[4]

[1]Department of Veterinary Population Medicine, College of Veterinary Medicine, University of Minnesota, Saint Paul, MN, USA, [2]Minnesota Department of Health, St Paul, MN, USA, [3]Veterinary Diagnostic Laboratory, College of Veterinary Medicine, University of Minnesota, St Paul, MN, USA and [4]Centro de Vigilancia Sanitaria Veterinaria (VISAVET), Universidad Complutense de Madrid, Madrid, Spain

Salmonellosis is one of the leading causes of foodborne disease worldwide, with an estimated one million cases a year in the United States. Salmonella 4,[5],12: i:-, a monophasic variant of *Salmonella typhimurium*, is an emerging serovar that has been

associated with multiple foodborne outbreaks throughout the world, mostly attributed to pig and pig products. Recently, we have demonstrated that two distinct groups of Salmonella 4,[5],12:i:- circulate in the USA and Europe, with the majority of isolates recovered during recent years belonging to an emerging multidrug-resistant clade (Elnekave et al. 2018). We applied Bayesian phylodynamic reconstruction to uncover the evolutionary history of this clade. We used a dataset of whole-genome sequences of 1446 4,[5],12:i:- isolates from different sources (livestock, human, food products, and others) from the USA ($n = 752$) and Europe ($n = 694$), collected between 2008 and 2017 and belonging to the Multilocus Subtype 34, which was predominant in the emerging clade (Elnekave et al. 2018). A subset ($n = 110$) of Salmonella 4,[5],12:i: isolates was then randomly selected after stratifying by location and year of isolation in order to achieve balanced sampling. Evidence of temporal signal was confirmed by looking at root-to-tip divergences using TempEst. Evolutionary hypotheses using strict and relaxed-clock models were tested using BEAST for a variety of demographic models and assuming a general time reversible substitution model. Model selection was performed by estimating Bayes Factors using path sampling and stepping-stone sampling. The selected model was then used for applying discrete trait models comparing different scenarios of transmission between locations (i.e. bidirectional symmetric/asymmetric or unidirectional). Our preliminary phylodynamic inference results indicate that the origin of this subtype was in Europe and dates back to 1990 (HPD 95%: 1984–2001). We report an exponential growth rate of 0.362 per year, which corresponds to a doubling time of 1.43 years. Our results suggest that this subtype was introduced to the US in the year 2000 (HPD 95%: 1994–2006). Phylodynamic analysis suggests that the recent increase in isolation of Salmonella 4,[5],12:i:- from different sources in the USA may be due to the exponential expansion of an emerging clone which originated in Europe and then expanded to the USA. The emergence and expansion of this serovar is of great public health importance due to the high prevalence of multidrug resistance traits found in USA isolates from this group and especially due to the presence of plasmid-mediated resistance genes for quinolones and extended spectrum cephalosporins, key antimicrobials used for the treatment of invasive Salmonella infections.

A67 Bloodstream infections by carbapenem-resistant *Klebsiella pneumoniae* subsp. *pneumoniae*: Bayesian phylogenetic analysis of whole genomes

Eleonora Cella,[1] Davide Leoni,[2] Walter Mirandola,[3] Carla Fontana,[3] Loredana Sarmati,[2] Massimo Andreoni,[2] Silvia Angeletti,[4] and Massimo Ciccozzi[1]

[1]Unit of Medical Statistic and Molecular Epidemiology, University Campus Bio-medico of Rome, Rome, Italy, [2]Clinical Microbiology Laboratories, Polyclinic of Tor Vergata, Viale Oxford 81, Rome 00133, Italy, [3]Clinical of Infectious Diseases, Department of System Medicine, Tor Vergata University, Rome, Italy and [4]Unit of Clinical Laboratory Science, University Campus Bio-medico of Rome, Rome, Italy

Bloodstream infection (BSI) caused by carbapenemase-producing Enterobacteriaceae (CPE) is a major public health concern, particularly in the hospital setting. The rapid detection of resistance patterns is of paramount importance for establishing the proper antibiotic regime. In addition, in countries where CPE are endemic, it is also important to evaluate genetic relationship among the isolates in order to trace pathogen circulation and to improve the infection control programs. This study is an application of a rapid blood culture (BC) workflow consisting of fast reporting of Gram stain results, rapid pathogen identification (using MALDI TOF technology), and a molecular assay for the detection of the major genes conferring resistance, all of them performed directly from positive BCs. The application of phylogenetic and phylodynamic analyses to bacterial whole-genome sequencing (WGS) data have become essential in the epidemiological surveillance of multidrug-resistant nosocomial pathogens. We analyzed 40 strains of *Klebsiella pneumoniae* subsp. *pneumoniae* (KP) carrying blaKPC (KP-KPC), randomly selected among 147 CPE identified from BCs collected from consecutive patients from 2013 to 2016. The number of BSIs-related CPE were 23, 31, 43, and 50 in 2013, 2014, 2015, and 2016, respectively. Among 147 CPE isolates, 143 were KP and four were *Escherichia coli* (EC). The gene blaKPC was detected in 117 strains of KP and in four strains of EC. Other carbapenemase genes, such as blaVIM and blaOXA-48, were detected in four and nine different isolates of KP, respectively. Moreover, 13 KP strains carried two resistance genes: twelve vehicled blaKPC plus blaVIM and one blaKPC plus blaOXA-48. Phylogenetic analysis of bacterial WGS data was used to investigate the evolution and spatial dispersion of KP in support of hospital infection control. The maximum likelihood tree showed two main clades statistically supported, with statistical support for several subclusters within as well. The minimum spanning tree showed mixing between sequences from different years and wards with only few specific groups. Bayesian analyses are ongoing, as the aid of Bayesian genomic epidemiology in combination with active microbial surveillance is highly informative regarding the development of effective infection prevention in healthcare settings or constant strain reintroduction.

A68 Characterization and comparative genomic analysis of two *Bacillus megaterium* lytic bacteriophages

A. Sharaf,[1] M. Oborník,[2,3] A. Hammad,[4] S. El-Afifi,[5] and E. Marei[5]

[1]Genetic Department, Faculty of Agriculture, Ain Shams University, Cairo, Egypt, [2]Biology Centre CAS, Institute of Parasitology, České Budějovice, Czech Republic, [3]Faculty of Science, University of South Bohemia, České Budějovice, Czech Republic, [4]Department of Microbiology, Faculty of Agriculture, Minia University, Minia, Egypt and [5]Department of Agricultural Microbiology, Virology Laboratory, Ain Shams University, Cairo, Egypt

Next-generation sequencing technologies provide unique possibilities for the comprehensive assessment of the environmental diversity of bacteriophages. Many Bacillus bacteriophages have been isolated, but very few *Bacillus megaterium* bacteriophages have been characterized. Here, we describe biological characteristics and whole-genome sequences and their annotations for two new isolates of the *B. megaterium* bacteriophages (BM5 and BM10), isolated from Egyptian soil samples and representing two different groups according to their host range and amplified fragment length polymorphism profiles. Both phages have been displaying different thermal inactivation points (82 and 78 °C) and pH tolerance range (5–9.2 and 5–8.4 pH) while having the same longevity *in vitro* (192 h). Electron microscopy observation has proved that both phages belonged to the Myoviridiae family. Furthermore, growth analyses indicated that phages BM5 and BM10 have a shorter latent period (20 and 25 min) and smaller burst size (103 and 117 PFU) than is typical for Bacillus phages. The genome sizes of phages BM5 and BM10 were 165,031 bp and 165,213 bp, respectively, with a modular organization. Bioinformatic analyses of BM5 and BM10 genomes enabled assignments of putative functions to 97 and 65 putative ORFs, respectively. Comparative analysis of BM5 and BM10 genome structures with other *B. megaterium* bacteriophages revealed relatively high levels of sequence and organizational identity. Both genomic comparisons and phylogenetic analyses support the conclusion that the sequenced phages (BM5 and BM10) belong to different sub-clusters (L5 and L7) within L cluster and display different lifestyles (lysogenic and lytic). Sequenced phages encode proteins associated with Bacillus pathogenesis. BM5 does not contain any tRNA sequences, while BM10 genome codes for 17 tRNAs.

A69 Phylodynamics of language evolution

K. Hoffmann and D. Kühnert

Max Planck Institute for the Science of Human History, Jena, Germany

We used Bayesian evolutionary analysis to study linguistic data and infer phylogenetic trees of language evolution. Languages were encoded as binary strings indicating the presence or absence of members of cognate classes, the equivalence of classes of words with similar meaning, and shared ancestry. These strings formed the alignment data used to compute the posterior likelihood of a tree with respect to Bayes' formula. Informative priors are crucial for testing hypotheses regarding the age of common ancestry and divergence times and should include as much available information as possible. Here, we investigated the birth–death process as a method to construct tree priors specifically suitable for modeling the evolution of cognate data. To test these models, we will use a dataset of the languages from Vanuatu, an island nation featuring world's highest language density.