Workshop on Evolution of Viruses
25–27 September 2023

Abstracts

Talk 1

Estimating viral mutation rates in the post-genomic era

Mark M. Tanaka, University of New South Wales, Australia

Abstract

Statistical methods to estimate mutation rates have been developed primarily for bacteria which replicate by binary fission. These methods have been applied to viruses even though viruses and bacteria can be very different in their modes of replication. The replication modes of viruses range from copying multiple genomes from a single template genome, which is referred to as the stamping machine, to geometric replication, where each genome produces a number of daughter genomes over multiple generations. Reproduction here is not necessarily binary. This range of replication modes is not explicitly considered in mutation rate assays. Here, we develop a framework that accounts for replication modes to infer mutation rates from sequence data. The method we propose can be applied to bacteria or viruses; it can infer the replication mode if it is unknown, and can also be applied if sampling of viral progeny genomes is incomplete. We also show that, with incomplete sampling, the replication mode affects the estimate of the mutation rate. By relying on sequence data, rather than phenotypes, our method merges the spirit of the Luria-Delbrück approach with the wealth of data available in the genomic era.

Talk 2

Evolutionary dynamics and lineage designation of SARS-CoV-2 genomes

Jian Lu, Peking University, China

Abstract

Coronavirus disease 2019 (COVID-19), the disease caused by infection with severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), has developed into a global pandemic that continues to pose an enormous threat to public health and the global economy. Since the first SARS-CoV-2 genome was released, thousands of genetic variants have been identified in SARS-CoV-2 strains isolated from worldwide patients. We and collaborators divided the SARS-CoV-2 genomes into two major lineages (L and S) based on variants at sites 8782 and 28144. To improve the tracing of the viral genomes' evolution during the development of the pandemic, we further divided the L lineage into two major sublineages (L1 and L2) using SNVs at sites 3037, 14408, and 23403. Subsequently, we categorized the genomes into 130 sublineages (37 in S, 35 in L1, and 58 in L2) based on marker SNVs at 201 additional genomic sites. This lineage/sublineage designation system has a hierarchical structure and reflects the relatedness among the subclades of the major lineages. We also provide a companion

website (www.covid19evolution.net) that allows users to visualize detailed sublineage information and to upload customized SARS-CoV-2 genomes for sublineage classification. We believe these efforts will improve our understanding of the evolutionary dynamics of SARS-CoV-2 genomes across varying temporal and spatial scales. Moreover, we analyzed 271 COVID-19 patients in the early outbreak in Wuhan and detected a significant difference in clinical severity between the L- and S-lineage patients, suggesting that SARS-CoV-2 lineage may provide useful clinical information for the management of COVID-19 and supports the argument that viral clades should be analyzed as a function of clinical severity.

Talk 3

Exploring the antibiotic resistance potential of the virome in infant metagenome

Anupam Gautam[1,2] and Daniel H. Huson[1,2]
1 *University of Tübingen, Germany*
2 *International Max Planck Research School "From Molecules to Organisms", Max Planck Institute for Biology Tübingen, Germany*

Abstract
Antibiotic resistance is a major global threat, and exploring the antibiotic resistance potential of the microbiome and virome is crucial for identifying and combating this issue. While many studies have focused on the antibiotic resistance potential of bacteria present in metagenomic samples, very few have delved into the virome resistance potential. In our study, we have conducted a comprehensive investigation of the antibiotic resistance potential of the human infant metagenome, including both the microbiome and virome, over a period of three years with a gap of four months between each sampling. We have explored the resistance potential of both normal metagenomic samples and virome samples and have attempted to correlate the development of the microbiome and virome with various factors such as antibiotic treatment and weight. Additionally, we report on the difficulty of mining resistance genes from sequenced reads and propose a bioinformatics and machine learning-based solution to this issue. Our findings have significant implications for understanding the role of the virome in antibiotic resistance and may inform future efforts to combat this growing global threat. This is joint work with the Ley lab at the Max-Planck-Institute for Biology Tübingen.

Talk 4

Drug Resistance Mutations in HIV

Olivier Gascuel
Institut de Systématique, Evolution, Biodiversité (ISYEB UMR7205 – CNRS, Muséum National d'Histoire Naturelle, SU, EPHE, UA), France

Abstract
Drug resistance mutations (DRM) appear in HIV under treatment pressure. Resistant variants can be transmitted to treatment-naive individuals, which can lead to rapid

virological failure and can limit treatment options. Consequently, quantifying the prevalence, emergence and transmission of drug resistance is critical to effectively treating patients and to shape health policies. In this keynote, I will introduce the field and the main issues and concerns. I will review recent computational approaches and in particular describe: (i) the machine learning methods intended to predict and explain the level of resistance of HIV variants from their sequence data; (ii) the phylogenetic methods used to survey the emergence and dynamics of resistant HIV transmission clusters. I will also present the results of several studies with contrasted results in high-income countries and in Africa. Finally, I will discuss possible research avenues and points that require special surveillance points to prevent treatment failure and HIV transmission.

Joint work with Luc Blassel and Anna Zhukova (Institut Pasteur, Paris, France)

References

Blassel L, Tostevin A, Villabona-Arenas CJ, Peeters M, Hué S, Gascuel O; UK HIV Drug Resistance Database. 2021. Using machine learning and big data to explore the drug resistance landscape in HIV. PLoS Computational Biology, 17(8):e1008873. doi: 10.1371/journal.pcbi.1008873

Blassel L, Zhukova A, Villabona-Arenas CJ, Atkins KE, Hué S, Gascuel O. 2021. Drug resistance mutations in HIV: new bioinformatics approaches and challenges. Current Opinion in Virology, 51:56-64. doi: 10.1016/j.coviro.2021.09.009

Zhukova A, Dunn D, Gascuel O on behalf of the UK HIV Drug Resistance Database & the Collaborative HIV, Anti-HIV Drug Resistance Network. 2023. Modeling Drug Resistance Emergence and Transmission in HIV-1 in the UK. *Viruses*. 2023; 15(6):1244. https://doi.org/10.3390/v15061244

Talk 5

A network approach for studying HIV molecular epidemiology

Tsz Ho Kwan, The Chinese University of Hong Kong, China

Abstract

Genetic similarity between two viral sequences could infer that the two individuals had a close transmission relationship. A phylogenetic network could be constructed based on the pairwise distance matrix as an alternative to a phylogenetic tree with edges connecting the nodes in the network denoting pairs of sequences with low genetic distances, hence high genetic similarity. Network analysis would then become possible to uncover the underlying structure of the network and the relationship between viral sequences to better understand the epidemiology of a pathogen. Particularly network metrics could be measured and compared among clusters or networks, and nodes situated in critical positions in the network could be identified to inform targeted public health policies. As a sexually transmitted infection, HIV transmission relied primarily on the community members' sexual behaviours and practices. On the other hand, their

networking profiles and preferences could not just shape their social network but also transmission network when the virus was circulating within the community. Supplemented with behavioural and clinical data, we investigated how such behaviours and partner-seeking patterns shaped the transmission dynamics and the resultant phylogenetic network of HIV in Hong Kong, in order that tailored prevention intervention could be designed to curb further transmission in the community.

Talk 6

Phylodynamics of SARS-CoV-2 in Hong Kong, 2020–2022

Vijaykrishna Dhanasekaran, The University of Hong Kong, Hong Kong, China

Abstract

Hong Kong implemented an elimination strategy, combining intermittent public health and social measures alongside increasingly stringent travel regulations, to control domestic transmission waves of SARS-Cov-2 until 2022. However, a surge of Omicron BA.2 infections in early 2022 resulted in a high per-capita death rate from COVID-19. This study aims to reconstruct the epidemic and evolutionary trajectory of SARS-CoV-2 variants and analyze the effects of intermittent public health and social measures on the evolution and epidemiology of SARS-CoV-2 in Hong Kong.

By analyzing comprehensive genome sequences and epidemiological data (Gu et al. 2022; Xie et al. 2023), we found that only five introductions were responsible for 90% of locally-acquired cases despite numerous importations. Notably, distinct community outbreaks were triggered by novel introductions rather than a resurgence of circulating strains. Stringent public health and social measures (PHSMs), along with comprehensive community surveillance, effectively suppressed epidemic growth. However, the prolonged implementation of preventative measures and festival gatherings led to adherence fatigue, hindering lineage termination and perpetuating community transmission. The largest fifth wave was triggered by a single cross-infection within a hotel quarantine. Phylodynamic analyses revealed extensive superspreading during the fifth wave, contributing to its rapid expansion. Furthermore, the impact of COVID-19 was unequally felt across Hong Kong's 18 districts.

This study provides insights into the introduction and transmission patterns of SARS-CoV-2 under an elimination strategy in Hong Kong. The findings offer valuable context for ongoing and future public health interventions. Additionally, the discordance between genomic and epidemiological data emphasizes the importance of integrating multiple disparate data sources to improve near real-time epidemic growth estimates, thus enhancing outbreak response policy. Further research is needed to refine these approaches and inform effective strategies for outbreak management.

Talk 7

Genomic epidemiology for estimation of serial intervals in COVID-19 transmission clusters

Jessica E Stockdale, Simon Fraser University, Canada

Abstract

Genomic data have proven extremely valuable in epidemiological investigations, with tracking of viral evolution providing detailed insight to underlying transmission processes. At the same time, the cost of genomic sequencing continues to decrease and many jurisdictions have moved to routine collection of pathogen sequences, particularly for SARS-CoV-2. This provides new opportunities for monitoring spatio-temporal trends in epidemiological quantities. Serial intervals – the time between symptom onset in an infector and infectee – are one such fundamental quantity, which help to define rates of transmission, estimates of reproductive numbers, and vaccination levels needed to prevent transmission.

I will present a new statistical approach for estimation of serial intervals using virus sequences. Although estimates of the serial interval are usually calculated from detailed contact tracing data, collection of such data can be expensive and as such estimates are often taken from small, early outbreaks. Our genomic approach takes incomplete sampling into account and is intended to be applied at a broader scale than typical contact-based approaches. We apply our methodology to data on COVID-19 from Victoria, Australia, and compare estimates of the serial interval from case clusters through time and across contact sites such as schools and healthcare.

Talk 8

Synthetic control methods for infectious disease epidemiology: applications to Wolbachia interventions

Jue Tao Lim, Nanyang Technological University, Singapore

Abstract

Large scale field trials to evaluate the effect of interventions to stave viral transmission are increasingly common. However, random assignment of interventions is often difficult to achieve. Here, we present the development and application of synthetic control methods (SCM) in evaluating the efficacies of interventions. We take *Wolbachia* releases as the key example, under the (1) introgression and (2) incompatible-insect technique approaches to reduce dengue virus transmission and/or suppress Aedes mosquito populations in 3 separate settings in dengue-endemic Southeast Asia and Latin America.

Spatially resolved dengue incidence and/or adult *Aedes* abundance data from Singapore, Malaysia and Brazil were used, along with a high-dimensional set of spatio-temporal set of environmental and anthropogenic covariates. Donor pools consisting of control units which are never-treated by *Wolbachia* interventions were used to construct synthetic controls using SCM. The canonical SCM was employed for single-site settings, while

partially-pooled and global SCM alternatives were employed for multiple-site settings. We used different linear combinations of covariates to account for confounding. Intervention efficacies, defined as the percentage reduction in dengue incidence or mosquito abundance, was compared under different settings and methodologies. Intervention efficacies ranged from 47 – 88% across all endpoints, and were found to be consistent across different settings, subgroups and *Wolbachia* strategies. Synthetic control methods were found to generate appropriate control groups as noted by good balance in the endpoint of interest and covariates between both intervention and synthetic control arms in the pre-intervention period. A further battery of robustness checks, such as placebo testing, confirm the validity of intervention efficacy estimates.

SCM can alleviate many problems which arise from non-randomized experimental settings. Alternatives can flexibly account for many confounders as well as staggered adoption settings where interventions were sequentially applied to different units across time. Our applications to *Wolbachia* releases demonstrate the high utility for this new class of vector control tools to stem both vector populations and dengue transmission across multiple setting.

Talk 9

Clinical research during the COVID-19 pandemic. What do we need to do better next time?

Barnaby Edward Young, National Centre for Infectious Diseases, Singapore

Abstract

The COVID-19 pandemic produced an avalanche of clinical research investigating topics such as SARS-CoV-2 transmission dynamics, variant immune escape and vaccine/treatment effectiveness. From a clinical point of view we had two main concerns - how can we prevent people from getting infected, and how can we identify early and best treat those at risk of a severe illness? In this talk A/Prof Barnaby Young from the National Centre of Infectious Diseases (NCID) will provide an overview of the clinical presentation and treatment of COVID-19; discuss some of the questions we had during the pandemic, and the studies we conducted to answer these; and present an overview of MOHs PREPARE programme that aims to make sure Singapore is ready for the next pandemic.

Talk 10

Levels of SARS-CoV-2 population exposure are considerably higher than suggested by seroprevalence

Siyu Chen, Oxford University, UK

Abstract

Accurate knowledge of prior population exposure has critical ramifications for preparedness plans for future SARS-CoV-2 epidemic waves and vaccine prioritization strategies. Serological studies can be used to estimate levels of past exposure and thus position populations in their epidemic timeline. To circumvent biases introduced by the decay in antibody titers over time, methods for estimating population exposure should account for seroreversion, to reflect that changes in seroprevalence measures over time are the net effect of increases due to recent transmission and decreases due to antibody waning. Here, we present a new method that combines multiple datasets (serology, mortality, and virus positivity ratios) to estimate seroreversion time and infection fatality ratios (IFR) and simultaneously infer population exposure levels. The results indicate that the average time to seroreversion is around six months, IFR is 0.54% to 1.3%, and true exposure may be more than double the current seroprevalence levels reported for several regions of England.

Talk 11

Detection and surveillance for genomic variation of SARS-CoV-2

Aiping Wu, Chinese Academy of Medical Sciences & Peking Union Medical College, China and Suzhou Institute of Systems Medicine, China

Abstract

The frequent variation and persistent pandemic of SARS-CoV-2 causes serious disease burden in human society. The detection and monitoring timely for genomic variation of SARS-CoV-2 is urgently needed not only for understanding viral evolution, but also for better disease control and prevention. Here, we have summarized our works about developing computational models for investigating SARS-CoV-2 evolution. We have reported an algorithm and R package named ‘sitePath’ to identify the parallel and fixed mutations in viral evolution. Used sitePath in SARS-CoV-2, 37 parallel and fixed mutations were identified as potentially adaptive mutations, in which 26 mutations had been proved with experimental evidence in references. We had also systematically analyzed the insertion/deletion mutation in SARS-CoV-2 evolution. We found at least 10 conserved hot regions for segment deletion among the whole viral genome, in which three regions were also observed in other coronaviruses. Then, we published a method named cov2Coinfect to detect the co-infected SARS-CoV-2 sequences based on deep sequencing data. With cov2Coinfect, we found 195 co-infected samples among over 500,000 sequenced infections. Furthermore, we are developing a recombinant-detection method named covRecomb to identify recombinant events from large-scale viral genomes and to construct the comprehensive recombinant picture in SARS-CoV-2

evolution. At last, by integrating the above methods together, we are building an online server named covEvolution to monitor the genomic variation and evolution of SARS-CoV-2, which could be used to help the disease surveillance and prevention.

Talk 12

Phylogenetic Supertree of Virus

Jie Feng, Lanzhou University, China

Abstract

The Supertree method combines a set of source phylogenetic trees to produce one comprehensive phylogenetic tree reasonably. The source phylogenetic trees employed for supertree construction can be consistent or inconsistent or partly overlapped based on different genes or phenotypes. Supertree method exhibits its technical superiority for phylogenetic analysis of creatures that lack compatible data for analysis using a single optimization criterion. It can use the full phylogenetic dataset that is available and combine data in various forms, including DNA or amino acid sequences, morphology, immunological distances, etc., to produce the overall finest supertree. The Supertree methods have been widely used for phylogenetic analysis of creatures with large size of genomes, including mammals, birds, fishes, etc. However, the supertree method is rarely used for the phylogenetic analysis of viruses. Our recent study applied the MRP (matrix representation with parsimony) pseudo-sequence supertree analysis to study the origin and evolution of SARS-CoV-2. Compared with other phylogenetic analysis methods, the supertree method showed more resolution power for the phylogenetic analysis of coronaviruses. As we know our study is the first one to use the approach of phylogenetic supertree analysis for phylogenetic inference of SARS-CoV-2. Taken together, the MRP pseudo-sequence supertree provided more information on the SARS-CoV-2 evolution inference relative to the normal phylogenetic tree based on full-length genomic sequences. These findings have implications for studying the origin of viruses and help to control viral pandemics.

Talk 13

The value of high coverage whole genome sequencing of SARS-CoV-2 in Denmark

Mark P. Khurana, University of Copenhagen, Denmark

Abstract

During the period of the Delta SARS-CoV-2 variant in Denmark, the sequencing coverage of PCR positive infections frequently surpassed 70-80%, which is high enough to consider the set of whole viral genomes as broadly “whole epidemic” representative. Using the full set of aligned whole genome sequences from the Statens Serum Institut (SSI) in Denmark, in combination with metadata linked to individuals through their social security numbers, we explore a suite of methods to build robust phylogenies tracing the ancestral

evolutionary process and attempt to answer questions about how useful sequencing almost every infection is.

More specifically, we focus on the following questions: (a) To what degree population wide phylogenies are robustly resolvable and how many phylogenies are congruent with each other (b) What is the “coreset” or reduced sample size needed to maintain full population wide genetic diversity (c) Can metadata from the registries be used for hypothetical survey design to define a sequencing coreset (d) Can approximate Bayesian methods be applied to allow for the use of complex molecular clocks and estimates of ancestral effective population sizes?

Talk 14

Genomic characterization and epidemiology of P.1 and Delta SARS-CoV-2 Variants
Swapnil Mishra, National University of Singapore, Singapore

Abstract
In this talk, I will retrospectively talk about how we (our previous work) characterized both P.1 and Delta variants around the time of their emergence. The main focus of the talk will be efforts around sequencing, and understanding the evolution and spread of these variants during the initial period. In particular, I will focus on how we integrate both the genomic and mortality data in a two-category dynamical model. The two-category dynamical model allowed us to tease apart the inherent increased transmissibility and/or immune evasion. We stress that the exact values are still very hard to pinpoint but these models allowed us to look at the role played by waning immunity, increased transmission, and immune escape in the establishment of P.1 and Delta variant.

Talk 15

The origin tracing and tracking of emerging influenza A virus and SARS-CoV-2
Yifei Xu, Shandong University, China

Abstract
SARS-CoV-2 Omicron caused a large wave of COVID-19 cases in China in spring 2022. Shandong was one of the most affected regions during this epidemic, yet was also among those areas that were able to quickly contain the transmission. We aimed to investigate the origin, genetic diversity, and transmission patterns of Omicron epidemic in Shandong under a dynamic clearance strategy. We generated 1,149 Omicron sequences, performed phylogenetic analysis, and interpreted results in the context of available epidemiological information. We observed that there were multiple introductions of distinct Omicron sub-lineages into Shandong from foreign countries and other regions in China, while a small number of introductions led to majority of local cases. We found evidence suggesting that some local clusters were potentially associated with foreign imported cases. Superspreading events and cryptic transmissions contributed to the rapid spread

of this epidemic. We identified a BA.1.1 genome with the R493Q reversion mutation in the spike receptor binding domain, potentially associated with an escape from vaccine and Omicron infection elicited neutralizing immunity. Our findings illustrated how the dynamic clearance strategy constrained this epidemic's size, duration, and geographical distribution.

Talk 16

Zika virus transmission dynamics and characterisation of naturally occurring mutations in a paediatric cohort during the 2016 Nicaragua epidemic

October M. Sessions, National University of Singapore, Singapore

Abstract

Nicaragua experienced a large Zika epidemic in 2016, with up to 50% of the population in Managua infected. With the domesticated Aedes aegypti mosquito as its vector, it is widely assumed that Zika virus transmission occurs within the household and/or via human mobility. We investigated these assumptions by using viral genomes to trace Zika transmission spatially. We analysed serum samples from 119 paediatric Zika cases participating in the long-standing Paediatric Dengue Cohort Study in Managua, which was expanded to include Zika in 2015. Contrary to our initial hypothesis, our findings suggest that community transmission, often involving long geographical distances, played a much more important role in epidemic spread than within-household transmission. Interestingly, our genomic analysis revealed that some mutations occurred with regularity throughout the cohort. To facilitate characterisation of these positions, the Nicaraguan ZIKV genome was reverse engineered into an infectious clone and the specific mutations were cloned individually via site-directed mutagenesis. Three of the mutations we identified display a significantly increased growth phenotype in human hepatoma cells in comparison to the wild type virus. Two of these mutations are located in the RdRp region of the viral NS5 protein and one in the viral NS4B. The NS5-R515K mutation is located proximal to the RdRP replication site and the NS5-H770P mutation is adjacent to one of the known T cell epitopes. The NS4B-M70I mutation is within the IFN suppression domain. In adult female Aedes aegypti, the NS5 mutations grew similarly to the wild type while the NS4B mutant displayed an attenuated phenotype suggesting that this particular mutation is human specific.