

What genomic data could tell us

[00:00:01.26] Hi. My name is Philippe Lemey, and I work at the Rega Institute, University of Leuven, Belgium. It's a pleasure to contribute to this course, specifically to the section-- what genomic data could tell us? And I will briefly introduce what role phylodynamics could play in leveraging the information contained in pathogen genome sequences.

[00:00:24.73] Phylodynamics is a research discipline that unifies the epidemiological and evolutionary dynamics of pathogens. And the central principle of phylodynamics is that viruses evolve so rapidly-- with evolutionary rates that are several orders of magnitude larger than for our own genes, as illustrated here-- which allows epidemiological processes to leave an imprint in the viral genomes. And we can use statistical techniques rooted in phylogenetics to recover information about these processes.

[00:01:05.46] Phylodynamics dynamics was originally coined in a paper by Bryan Grenfell and colleagues, published in 2004, to describe the interaction between genetic diversity, epidemic dynamics, and natural selection, mostly in the context of immunology.

[00:01:25.50] Let me start by illustrating the relationship between transmission processes and phylogeny and I will do this with the transmission history on the left here, where the vertical lines represent transmission events and the numbers here represent sampling events. Now the underlying phylogeny, or the relatedness between the sample genomes, should ideally look like the graph on the right. However, our ability to reconstruct this pattern of relatedness and, hence, this part of this transmission history, depends on how many mutations or substitutions are thrown into the genomes throughout this transmission history.

[00:02:19.12] When many substitutions occur, specifically when they occur at a higher rate than the transmission rate, then our phylogenetic reconstruction will generally be well-resolved and we will do a good job in recovering the pattern of relatedness, which is now measured in substitution units. However, when the substitution rate is lower than the transmission rate, as is the case, for example, for COVID-19, much less substitutions will accumulate in the genomes and there will be less opportunity to reconstruct the short-term transmission process in great detail.

[00:03:06.70] Now using phylodynamic techniques that builds on these phylogenetics concepts, there are a number of questions we can address about pathogen evolution and epidemiology. And I will focus here on the epidemiological questions. One of these questions is when an epidemic started? And this is part of a broader set of questions about the timescale of the transmission process. Importantly, our standard phylogenetic techniques reconstruct the pattern of relatedness in units of substitution. And now we want to obtain an estimate on a calendar time scale.

[00:03:51.70] So we need to convert our trees in substitution units into one in time units. This can be done if we have an estimate of the rate at which substitutions accumulate in genome sequences. Now for rapidly evolving viruses, we often see that sequences sampled later in time show more divergence from the root as compared to sequences sampled earlier in time. This is also what we-- what is illustrated in this regression plot of divergence as a function of sampling time, where the slope essentially represents an estimate of the evolutionary rate.

[00:04:40.99] Now we can formally incorporate this information about sampling time in what we call a dated deep molecular clock model and use this to estimate our tree in time units. Now in such a tree, every node is now associated with a time estimate. And the time to the most recent common ancestor represents the origin of all the sampled genomes.

[00:05:11.43] Other phylodynamic questions focus on the rate of viral spread. This is something that can be addressed with traditional modelling of epidemiological data, but phylodynamics analysis of genomic data may offer an interesting complementary approach here. This is because the shape of a tree reflects how the size of the sampled population has changed over time. We can formally make the connection between tree shape and changes in population size using a mathematical theory called the coalescent.

[00:05:52.63] Now intuitively we can see that when it takes a lot of time for lineages to find their common ancestor or to coalesce, then this is because the population size is relatively large. While when lineages coalesce fast, then this is because the population sizes are small. Now other types of models exist in addition to the coalescent that rely on trees to learn about epidemic processes and some of these have been explicitly parameterized, for example, in terms of the basic reproductive rate, a key parameter of transmission.

[00:06:38.88] Another aspect that we can learn a lot about using phylodynamic analysis is how transmission manifests spatially. For this we can embed simple spatial models in our phylogenies and reconstruct how viruses have migrated or dispersed between locations over time, for example, considering a set of discrete locations. You can also use a similar model for locations represented by spatial coordinates and, hence, perform phylogeographic analyses in continuous space.

[00:07:26.24] All the models I have discussed here have been made available in an integrated fashion in a coherent Bayesian statistical software. So a single analysis may provide insights into all of these processes. The drawback, however, is that this may require some degree of expertise and that the analysis may be time-consuming, so not necessarily applicable in a real time fashion. That is why also more approximate heuristics have been developed and some of these are available in popular tools, such as Nextstrain.

[00:08:04.17] Now all of this has contributed to the popularity of phylodynamics. And as we have witnessed during the COVID-19 pandemic, the types of epidemiological questions we have discussed have been extensively addressed for SARS-CoV-2, thanks to a massive genomic sequencing effort. So we think phylodynamics will only grow in importance as an academic discipline, but also as one that can ultimately contribute to the response against infectious disease spread.