

# Statistical inference of virus phylogenetics

Oliver Ratmann

Imperial College, London

The infectious disease dynamics of many viral pathogens like influenza, norovirus and coronavirus are inextricably tied to their evolution. This interaction between evolutionary and ecological processes complicates our ability to understand the infectious disease behavior of rapidly evolving pathogens. Most statistical methods for the analysis of these "phylogenetics" require that the likelihood of the data can be explicitly calculated. Currently, this is not possible for many phylogenetic models, so that questions on the interaction between viral variants cannot be well-addressed within this framework. Simulation-based statistical methods circumvent likelihood calculations. Considering inter-pandemic human influenza A virus subtype H3N2, we illustrate the effectiveness of these methods to fit and assess complex phylogenetic models against both sequence and surveillance data. We find that combining molecular genetic and epidemiological data is key to estimate phylogenetic parameters reliably. Moreover, the information in the available data taken together is enough to expose quantitative model inconsistencies. Methods such as ABC which can combine sequence and surveillance data appear to be well-suited to fit and assess mechanistic hypotheses on the phylogenetics of RNA viruses.

Introductory reading: Grenfell et al. (2004) *Science* 303:327-332, Unifying the epidemiological and evolutionary dynamics of pathogens