

Phylodynamic analysis of pathogens in the genomics era: statistical tools and applications

Marco Salemi, Ph.D.

Abstract

During the past decade a new framework has been developed called phylodynamics that combines phylogenetic analysis and coalescence theory to correlate the epidemiology and evolutionary behavior of pathogens. The time-scale of epidemic spread usually provides ample time for evolving pathogens to accumulate informative mutations in their genomes. Reconstructing the evolutionary history from full genome viral or bacterial sequences has provided a fundamental understanding of the dynamics underlying epidemics of pathogens such as HIV, Ebola, *Vibrio cholerae* and MRSA. By using a statistical model, known as the relaxed molecular clock, it is possible to infer the genealogy and evolutionary time-scale of the sampled sequences. Spatial diffusion also leaves a measurable footprint in sampled sequences; Bayesian phylogeography can successfully be used to reconstruct the gene flow (migration/spread) of specific pathogens and test hypotheses of spatial spread. Such tools have been applied also to study the intra-host evolution of pathogens, resulting in key insights on the relationship between genomic evolution and pathogenesis.