

Phylogenetics of an HIV epidemic in Southeastern Michigan

Erik Volz *

erikvolz@umich.edu

We analyze 632 clade B HIV-1 sequences isolated from MSM in Detroit, Michigan between 2004 and 2010. Each sequence is linked to behavioral and demographic characteristics of the patient from whom it was isolated. Epidemiological analysis was based on both heuristic clustering methods and a rigorous coalescent framework. Phylogenetic clustering of the sample was calculated by a variety of measures and over a range of thresholds. Thresholds for clustering were defined in terms of tree-based cophenetic distance as well as alignment-free methods that don't rely on estimation of a phylogeny. Correlation of demographic and behavioral variables with co-clustered taxa was estimated, and the different clustering methods yielded broadly concordant estimates of population structure. Clustering patterns were further compared to self-reported sexual histories from the National Health and Behavioral Surveillance survey. We find that assortativity coefficients calculated from phylogenetic clustering patterns and independent survey data were very similar. The strongest assortativity was found by race (38.2%) followed by age (16%).

We also consider the problem of estimating demographic history of an epidemic from a combination of standard data sources (case reports) and genetic sequences of the virus. This analysis is based on both the standard Bayesian skyline estimate of effective population size and a population genetics model specifically designed for infectious disease epidemics. We specified a model for HIV transmission dynamics using ordinary differential equations. This forward-time description of epidemic dynamics was adapted into a coalescent framework, allowing direct estimation of transmission rates from genetic data. Estimates are obtained for early (one per ~ 32.05 days) and late HIV transmission rates (one per ~ 1373.61 days), the reproduction number (3.31), and behavioral variables, such as the duration of risk behavior. The fraction of HIV transmissions that occur during the first 60 days of infection changes over the course of the epidemic (initially 56%). Bayesian skyline estimates were largely in agreement with the deterministic model during the initial period of exponential growth, but gave divergent predictions for epidemic dynamics following the expansion phase.

*Department of Epidemiology, School of Public Health, University of Michigan, 1415 Washington Heights, 4647 SPH Tower, Ann Arbor, MI 48109-2029, USA.