HIV-1 Transmission Networks in a Small World

Pleuni S. Pennings,[1] Susan P. Holmes,[2] and Robert W. Shafer[3]

[1]Department of Biology, [2]Department of Statistics, and [3]Department of Medicine, Stanford University, Stanford, California

(See the major article by Wertheim et al on pages 304–13)

Keywords. HIV; phylogeny; transmission; network.

Phylogenetics, the study of relatedness among homologous genetic sequences, is an integral part of virology research. Phylogenetic methods are used to reconstruct ancestral relationships among a set of viral sequences and to provide a framework within which hypotheses about virus evolution can be tested. By accounting for the nonindependence or shared ancestry of sampled sequences, the phylogenetic context makes it possible to distinguish between founder effects and natural selection as explanations for the spread of virus variants [1]. The recognition that circulating virus lineages coalesced at some time in the past makes it possible to use the dates and locations of viral sequences to reconstruct the spatiotemporal dynamics of an epidemic [2].

Viruses such as human immunodeficiency virus type 1 (HIV-1) are measurably evolving and, thus, generate epidemics in which there is a correspondence between the transmission network and phylogenetic branching [3]. Viruses that are phylogenetically clustered are likely to be connected by shorter transmission chains than viruses that are phylogenetically more distant from one another. Factors such as geography, the method of virus transmission, and the risk factors of infected persons also influence the parameters of a phylogenetic tree [3]. When combined with epidemiological data, phylogenetic cluster analyses can inform public health interventions by determining which epidemiological factors are associated with increased virus transmission [4, 5].

The parameters responsible for virus evolution and spread and the extent to which virus phenotypic characteristics correlate with shared ancestry can be estimated from large sequence datasets by subsampling up to several hundred sequences at a time and integrating or averaging the results obtained from the resulting trees [6]. Repeated subsampling makes large phylogenetic analyses tractable and yields parameter estimates that are independent of any single phylogeny.

The HIV-1 reverse transcriptase (RT) gene has been sequenced more often than any other virus gene because RT sequences are often used in clinical settings to help guide the use of antiretroviral therapy. In this issue of *The Journal of Infectious Diseases*, Wertheim and colleagues leveraged the vast amount of publicly available HIV-1 RT sequence data to analyze the relatedness of previously published HIV-1 RT sequences from >80 000 individuals worldwide [7]. To analyze this large number of sequences, the authors dispensed with a phylogenetic approach and instead clustered sequences solely by their genetic distances without considering their ancestral history. Viruses were represented as nodes and were connected to one another if the genetic divergence of their sequences was no greater than 1%.

Wertheim and colleagues found that 13 300 sequences, approximately one-sixth of the sequences in the dataset, were connected to at least 1 other sequence and that the mean number of connections per sequence was 3.8. The 1% threshold for connecting sequences from different individuals was sensitive enough to identify many of the inferred transmission clusters reported in the previously published studies from which the sequences were obtained. Most of these clusters were known from smaller regional cohorts of HIV-1–infected individuals. The authors' global analysis made it uniquely possible for them to identify >200 connections among sequences in different countries.

As in previous studies of transmission clusters, Wertheim et al found that the degree of connectivity of viruses from different individuals was heterogeneous with some nodes connected to many more nodes than other nodes. Such heterogeneous connectivity patterns characterize most biological and social networks [8]. Consistent with the decreased replication fitness of viruses with drug resistance

---

Received 30 July 2013; accepted 31 July 2013; electronically published 22 October 2013.

Correspondence: Robert W. Shafer, MD, Stanford University, Division of Infectious Diseases, Department of Medicine, Lane L143, 300 Pasteur Dr, Stanford, CA 94305 (rshafer@stanford.edu).

The Journal of Infectious Diseases 2014;209:180–2

DOI: 10.1093/infdis/jit525

mutations, sequences containing drug resistance mutations were less likely to be part of a cluster than sequences without drug resistance mutations.

The article by Wertheim and colleagues warrants commentary to address the use of a network (rather than a tree) to study HIV-1 transmission dynamics and to review the epidemiological conclusions that can be drawn by analyzing populations of HIV-1 sequences from different individuals either in isolation or in combination with temporal, geographic, epidemiological, and clinical data.

A network is derived from a matrix of pairwise distances between each pair of nodes or sequences in a dataset. For large numbers of sequences, a network can be created much more rapidly than a tree because no attempt is made to take into account the patterns of shared ancestry of the individual sequences. One limitation of creating a network directly from a set of sequences, however, is that many more connections may be inferred than could have possibly existed in the real transmission network, a problem that may not occur if a phylogenetic analysis had been performed first. For example, if multiple infections happen in a short time span, several people may be infected with very similar viruses. The viral sequences from these people would all be connected by the method of Wertheim and colleagues, leading to many more edges in the thus constructed network than exist in reality. Indeed, with the exception of a study published by several authors of the study reviewed here [9], previous HIV-1 network analyses used a preliminary phylogenetic analysis either to define the connections between sequences or to characterize the internal structure of a set of connected sequences [4, 5, 10].

Both phylogenetic and network analysis can be used to estimate the cluster size distribution in a population of virus sequences. As in the study by Wertheim et al, the distribution of cluster sizes in populations of HIV-1 sequences is usually found to be heterogeneous, consistent with different rates of transmission for different subpopulations or individuals [4, 5, 10]. This heterogeneity suggests that interventions to prevent transmission should be targeted toward those with the highest risk of transmitting the virus to others. Not surprisingly, studies in which sequence data are combined with behavioral risk factors have shown that sequences from individuals with the same risk factors are likely to cluster with one another [10, 11].

Both phylogenetic and network analyses of HIV-1 sequences combined with information about the duration of infection have shown that sequences from recently infected individuals are more likely to cluster with sequences from other recently infected individuals [4, 12]. This clustering is consistent with the high HIV-1 transmissibility known to occur during acute infection [13]. In agreement with these studies, the authors' previously cited study [9] combined sequences and clinical data from 3700 individuals in 5 US clinics and found that sequences that were part of a cluster (defined as having a genetic distance no more than 1.5% from 1 or more other sequences) were more likely to be from younger antiretroviral-naive individuals with high plasma HIV-1 RNA levels than were sequences that were not part of a cluster. These and other studies support interventions directed toward reducing the risks of secondary transmission events by individuals with acute infection [14].

Although the complete knowledge of a specific transmission network is informative, such perfect knowledge is not required to understand the general processes responsible for the spread of a virus [15]. Neither phylogenetic nor nonphylogenetic network analyses can prove that HIV-1 transmission occurred directly between 2 individuals [16]. Although 2 individuals may carry HIV-1 strains with very similar sequences, these sequences will not necessarily be unique; very similar sequences could be found in viruses from other persons within the same transmission network. The notion that sequences alone can identify specific transmission events between individuals is a misconception that has the potential to jeopardize the continued public benefit that results from the open publication of pathogen genomes [17]. Hypotheses about direct transmission, therefore, make sense only when sequence analysis is combined with contact tracing. Few studies, however, contain both contact data and HIV-1 sequence data, and those that do have been too small to provide insight into the population-level factors responsible for HIV-1 spread.

By identifying the large number of connections between individuals in different countries, Wertheim and colleagues demonstrate the novel insights that can be gained by analyzing publicly available sequence data from a variety of published studies. A more complete understanding of the relative importance of different global transmission routes will require further studies and new methods that can incorporate information on the intensity of sampling in different regions, different risk groups, and different years. Nonetheless, the data presented here demonstrate that even a virus such as HIV-1, which requires intimate contact for transmission, can form small-world networks linking virus variants from geographically remote regions [18].

Notes

Financial support. P. S. P. is supported by a long-term fellowship of the Human Frontier Science Program (LT000591/2010-L). S. P. H. is supported by the National Institutes of Health (NIH; grant R01 GM 86884). R. W. S. is supported by the National Institute of Allergy and Infectious Diseases, NIH (grant AI068581).

Potential conflicts of interest. All authors: No reported conflicts.

All authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Conflicts that the editors consider relevant to the content of the manuscript have been disclosed.

References

1. Drummond AJ, Nicholls GK, Rodrigo AG, Solomon W. Estimating mutation parameters, population history and genealogy simultaneously from temporally spaced sequence data. Genetics **2002**; 161:1307–20.

2. Moya A, Holmes EC, Gonzalez-Candelas F. The population genetics and evolutionary epidemiology of RNA viruses. Nat Rev Microbiol **2004**; 2:279–88.
3. Pybus OG, Rambaut A. Evolutionary analysis of the dynamics of viral infectious disease. Nat Rev Genet **2009**; 10:540–50.
4. Volz EM, Koopman JS, Ward MJ, Brown AL, Frost SD. Simple epidemiological dynamics explain phylogenetic clustering of HIV from patients with recent infection. PLoS Comput Biol **2012**; 8:e1002552.
5. Leigh Brown AJ, Lycett SJ, Weinert L, Hughes GJ, Fearnhill E, Dunn DT. Transmission network parameters estimated from HIV sequences for a nationwide epidemic. J Infect Dis **2011**; 204:1463–9.
6. Drummond AJ, Rambaut A. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol Biol **2007**; 7:214.
7. Wertheim JO, Leigh Brown AJ, Hepler NL, et al. The global transmission network of HIV-1. J Infect Dis **2014**; 209:304–13.
8. Keller EF. Revisiting “scale-free” networks. BioEssays **2005**; 27:1060–8.
9. Aldous JL, Pond SK, Poon A, et al. Characterizing HIV transmission networks across the United States. Clin Infect Dis **2012**; 55:1135–43.
10. Lewis F, Hughes GJ, Rambaut A, Pozniak A, Leigh Brown AJ. Episodic sexual transmission of HIV revealed by molecular phylodynamics. PLoS Med **2008**; 5:e50.
11. Kouyos RD, von Wyl V, Yerly S, et al. Molecular epidemiology reveals long-term changes in HIV type 1 subtype B transmission in Switzerland. J Infect Dis **2010**; 201: 1488–97.
12. Brenner BG, Roger M, Routy JP, et al. High rates of forward transmission events after acute/early HIV-1 infection. J Infect Dis **2007**; 195:951–9.
13. Wawer MJ, Gray RH, Sewankambo NK, et al. Rates of HIV-1 transmission per coital act, by stage of HIV-1 infection, in Rakai, Uganda. J Infect Dis **2005**; 191:1403–9.
14. Goodreau SM, Cassels S, Kasprzyk D, Montano DE, Greek A, Morris M. Concurrent partnerships, acute infection and HIV epidemic dynamics among young adults in Zimbabwe. AIDS Behav **2012**; 16: 312–22.
15. Welch D, Bansal S, Hunter DR. Statistical inference to advance network models in epidemiology. Epidemics **2011**; 3:38–45.
16. Bernard EJ, Azad Y, Vandamme AM, Weait M, Geretti AM. HIV forensics: pitfalls and acceptable standards in the use of phylogenetic analysis as evidence in criminal investigations of HIV transmission. HIV Med **2007**; 8:382–7.
17. Pybus OG, Fraser C, Rambaut A. Evolutionary epidemiology: preparing for an age of genomic plenty. Philos Trans R Soc Lond B Biol Sci **2013**; 368:20120193.
18. Danon L, Ford AP, House T, et al. Networks and the epidemiology of infectious disease. Interdiscip Perspect Infect Dis **2011**; 2011: 284909.