Can Directionality of HIV Transmission be Predicted by Next-Generation Sequencing Data?

Huldrych F. Günthard[1,2] and Roger Kouyos[1,2]

[1]Division of Infectious Diseases and Hospital Epidemiology, University Hospital Zurich, and [2]Institute of Medical Virology, University of Zurich, Switzerland

(See the Major Article by Rose et al, on pages 1406–413.)

The use of phylogenetic methods has become increasingly helpful, in addition to classical epidemiological methods, in providing information on the structure and dynamics of human immunodeficiency virus type 1 (HIV) epidemics. Key issues addressed by phylogenetic methods include the transmission of drug resistance [1, 2], the role of different infection phases for transmission [3, 4], interactions between different transmission groups [5, 6], the spread of HIV between countries [7–9], as well as the impact of HIV genomes on set point viral load variation [10, 11] and other phenotypes [12]. Most of these analyses have been performed using population-based Sanger sequences, which are being generated at a large scale because of routine genotypic resistance testing and, in some cohorts containing biobanks, by retrospective sequencing. With the ongoing switch of routine genotypic resistance testing to next-generation sequencing (NGS), the potential benefits and scope of phylogenetic analyses of HIV transmission are likely to increase further [13].

Estimation of transmission directionality emerged as a research question once HIV sequences became available [14, 15]. However, correct prediction of directionality has turned out to be a major challenge for many reasons, including imperfect sampling times of index subjects and their partners, allowing for within-host evolution, lack of background sequences from larger HIV-infected populations, technical challenges that do not allow generation of all HIV variants present in a given index subject or partner, dying out of variants over time in the index subject, superinfection and recombination events over time, and potential differences between genes studied.

For many research questions, directionality is not central because identifying potential transmission pairs or clusters is sufficient for their assessment [16]. Nevertheless, for detailed transmission studies of early pathogenesis events, particularly for legal cases, correct prediction of the directionality is of high importance. Once NGS became available, the expectation was that prediction of directionality most likely could be improved because the identification of a larger number of genetic variants should help improve the quality of the phylogenetic trees and might provide additional topological signals associated with directionality of transmission [13].

In this issue of *The Journal of Infectious Diseases,* Rose et al [17] present an interesting study in which they tested whether prediction of directionality could be improved by using NGS. They performed 454 NGS sequencing of a 323-nucleotide *env* fragment in 33 known index patients with chronic HIV infection and their partners who subsequently got infected during the HPTN 052 study [18, 19]. For index patients, 2 samples were available: one close to the seroconversion event of the partner (hereafter, "SC samples") and an earlier one (hereafter, "early index samples"). The authors used 2 phylogenetic methods: the single-tree method and a bootstrap method. They showed that transmission directionality could be correctly predicted when using the single-tree method in 22 of 33 pairs (67%) when using SC samples and in 23 pairs (74%) when using early index samples. For 4 pairs (12%–13%), prediction was incorrect. In the remaining pairs, the predicted transmission directionality was equivocal or inconsistent. Thus, the method provided a correct estimate of transmission directionality for 85% of pairs for which it claimed to provide an unequivocal estimate (23 of 27 pairs involving early index samples and 22 of 26 pairs involving SC samples). By using the bootstrap method, transmission directionality was correctly predicted for 18 pairs (55%) involving SC samples and for 24 pairs (73%) involving early index samples. Transmission directionality was predicted incorrectly in 7 pairs (21%) involving SC samples and in 4 pairs involving early index samples.

At first glance, these results are rather disappointing because the analyses based on this impressive set of sequence data, combined with a perfect patient population of prospectively studied serodiscordant couples, were not able to better predict transmission directionality; in particular, a wrong prediction was made

Received 20 December 2018; editorial decision 20 December 2018; accepted 21 December 2018; published online December 24, 2018.

Correspondence: H. F. Günthard, Division of Infectious Diseases and Hospital Epidemiology, University Hospital Zurich, University of Zurich, Zurich, Switzerland (huldrych.guenthard@usz.ch).

DOI: 10.1093/infdis/jiy737

in 15% of cases. It should, however, be noted that, as a positive side effect of this limited predictability of transmission directionality, some of the ethical problems associated with NGS data might be less serious than expected. In particular, the results showed that, despite the larger information content of NGS sequences, which can, in principle, allow the identification of transmitters [13], inferences about directionality are associated with substantial uncertainty. Thus, these results should reduce the fears that NGS sequences will create qualitatively new privacy concerns [20, 21].

Why is prediction of transmission directionality not more reliable? First, this study considered a fragment of *env*, the most variable region of the HIV genome. Owing to its high diversity, *env* might be prone to increased polymerase chain reaction amplification bias and more convergent evolution than if more-conservative genes, such as *pol*, were chosen for such analysis. Second, the fragment was quite short, and if the same analysis were to be repeated using multiple genes and full-length NGS methods with high coverage, the accuracy of transmission directionality would most likely increase. However, generating haplotypes for full-length sequences on the basis of NGS data is highly challenging and might not easily solve the problem [22]. Third, in 60%–80% of HIV infections, the founder population is monoclonal or oligoclonal, and that population might not be found anymore in the index cases because of within-host evolution and extinction in the course of infection [23].

In conclusion, although this was a very well-conducted study to test the predictability of transmission directionality, it showed that, for individual cases, reliable prediction is still not possible on the basis of sequence data alone (ie, in the absence of sound epidemiological data). Performance of this method could potentially be improved if the analysis included a diversity measure, based on the logic that, because the HIV population in index patients with chronic infection is, in most cases, much more diverse than that in their partners, knowledge of the relative diversity could enhance the reliability of predicting the transmission directionality.

Notes

Financial support. This work was supported by the Swiss National Science Foundation (grants 179 571 [to H. F. G.] and BSSGI0_155851 [to R. D. K.]); the Clinical Research Priority Program Viral Infectious Diseases Project, University of Zurich (to H. F. G.); the Zurich Primary HIV Infection Study (to H. F. G.); the National Institutes of Health (grant to H. F. G.); the Swiss HIV Cohort Study (to H. F. G.); Systems.X (to H. F. G.); the Yvonne Jacob Foundation (to H. F. G.); and Gilead Sciences and Roche (unrestricted research grant to H. F. G.).

Potential conflicts of interest. H. F. G. has received unrestricted research grants from Gilead Sciences and Roche, fees for data and safety monitoring board membership from Merck, and fees for consulting/advisory board membership from Gilead Sciences, Sandoz, and Mepha. R. D. K. reports receipt of personal fees from Gilead Sciences and grants from the Swiss National Science Foundation, all outside the submitted work. 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. Mbisa JL, Fearnhill E, Dunn DT, Pillay D, Asboe D, Cane PA; UK HIV Drug Resistance Database. Evidence of self-sustaining drug resistant HIV-1 lineages among untreated patients in the United Kingdom. Clin Infect Dis **2015**; 61:829–36.
2. Drescher SM, von Wyl V, Yang WL, et al.; Swiss HIV Cohort Study. Treatment-naive individuals are the major source of transmitted HIV-1 drug resistance in men who have sex with men in the Swiss HIV Cohort Study. Clin Infect Dis **2014**; 58:285–94.
3. Ratmann O, van Sighem A, Bezemer D, et al.; ATHENA observational cohort. Sources of HIV infection among men having sex with men and implications for prevention. Sci Transl Med **2016**; 8:320ra2.
4. Marzel A, Shilaih M, Yang WL, et al.; Swiss HIV Cohort Study. HIV-1 transmission during recent infection and during treatment interruptions as major drivers of new infections in the Swiss HIV Cohort Study. Clin Infect Dis **2016**; 62:115–22.
5. 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.
6. Marzel A, Kusejko K, Weber R, et al.; Swiss HIV Cohort Study. The cumulative impact of harm reduction on the Swiss HIV epidemic: cohort study, mathematical model, and phylogenetic analysis. Open Forum Infect Dis **2018**; 5:ofy078.
7. Paraskevis D, Pybus O, Magiorkinis G, et al.; SPREAD Programme. Tracing the HIV-1 subtype B mobility in Europe: a phylogeographic approach. Retrovirology **2009**; 6:49.
8. Gilbert MT, Rambaut A, Wlasiuk G, Spira TJ, Pitchenik AE, Worobey M. The emergence of HIV/AIDS in the Americas and beyond. Proc Natl Acad Sci U S A **2007**; 104:18566–70.
9. Wertheim JO, Leigh Brown AJ, Hepler NL, et al. The global transmission network of HIV-1. J Infect Dis **2014**; 209:304–13.
10. Alizon S, von Wyl V, Stadler T, et al.; Swiss HIV Cohort Study. Phylogenetic approach reveals that virus genotype largely determines HIV set-point viral load. PLoS Pathog **2010**; 6:e1001123.
11. Blanquart F, Wymant C, Cornelissen M, et al.; BEEHIVE collaboration. Correction: Viral genetic variation accounts for a third of variability in HIV-1 set-point

viral load in Europe. PLoS Biol **2017**; 15:e1002608.
12. Kouyos RD, Rusert P, Kadelka C, et al.; Swiss HIV Cohort Study. Tracing HIV-1 strains that imprint broadly neutralizing antibody responses. Nature **2018**; 561:406–10.
13. Wymant C, Hall M, Ratmann O, et al. PHYLOSCANNER: inferring transmission from within- and between-host pathogen genetic diversity. Mol Biol Evol **2017**; 35(3):719–733. doi:10.1093/molbev/msx304.
14. Ou CY, Ciesielski CA, Myers G, et al. Molecular epidemiology of HIV transmission in a dental practice. Science **1992**; 256:1165–71.
15. Yirrell DL, Robertson P, Goldberg DJ, McMenamin J, Cameron S, Leigh Brown AJ. Molecular investigation into outbreak of HIV in a Scottish prison. BMJ **1997**; 314:1446–50.
16. Bachmann N, Turk T, Kadelka C, et al.; Swiss HIV Cohort Study. Parent-offspring regression to estimate the heritability of an HIV-1 trait in a realistic setup. Retrovirology **2017**; 14:33.
17. Rose R, Hall M, Redd AD, et al.; Phylogenetic methods inconsistently predict direction of HIV transmission among heterosexual pairs in the HPTN052 cohort. J Infect Dis. **2019**; 220:1406–413.
18. Cohen MS, Chen YQ, McCauley M, et al.; HPTN 052 Study Team. Antiretroviral therapy for the prevention of HIV-1 transmission. N Engl J Med **2016**; 375:830–9.
19. Eshleman SH, Hudelson SE, Redd AD, et al. Treatment as prevention: characterization of partner infections in the HIV prevention trials network 052 trial. J Acquir Immune Defic Syndr **2017**; 74:112–6.
20. Abecasis AB, Pingarilho M, Vandamme AM. Phylogenetic analysis as a forensic tool in HIV transmission investigations. AIDS **2018**; 32:543–54.
21. Coltart CEM, Hoppe A, Parker M, et al.; Ethics in HIV Phylogenetics Working Group. Ethical considerations in global HIV phylogenetic research. Lancet HIV **2018**; 5:e656–66.
22. Beerenwinkel N, Günthard HF, Roth V, Metzner KJ. Challenges and opportunities in estimating viral genetic diversity from next-generation sequencing data. Front Microbiol **2012**; 3:329.
23. Leitner T, Romero-Severson E. Phylogenetic patterns recover known HIV epidemiological relationships and reveal common transmission of multiple variants. Nat Microbiol **2018**; 3:983–8.