COMMENTARY

The Power and Pitfalls of HIV Phylogenetics in Public Health

James I. Brooks, MD, FRCPC,[1,2] Paul A. Sandstrom, PhD[1]

ABSTRACT

Phylogenetics is the application of comparative studies of genetic sequences in order to infer evolutionary relationships among organisms. This tool can be used as a form of molecular epidemiology to enhance traditional population-level communicable disease surveillance. Phylogenetic study has resulted in new paradigms being created in the field of communicable diseases and this commentary aims to provide the reader with an explanation of how phylogenetics can be used in tracking infectious diseases. Special emphasis will be placed upon the application of phylogenetics as a tool to help elucidate HIV transmission patterns and the limitations to these methods when applied to forensic analysis. Understanding infectious disease epidemiology in order to prevent new transmissions is the *sine qua non* of public health. However, with increasing epidemiological resolution, there may be an associated potential loss of privacy to the individual. It is within this context that we aim to promote the discussion on how to use phylogenetics to achieve important public health goals, while at the same time protecting the rights of the individual.

KEY WORDS: Phylogenetics; HIV; public health; criminalization of HIV; molecular epidemiology

La traduction du résumé se trouve à la fin de l'article.

Can J Public Health 2013;104(4):e348-e350.

Classification of living organisms, and the relationship among species, is one of the most enduring pursuits in human beings' efforts to understand their place in the world. Formalized in the early 18th century by Carl Linnaeus, the phylogeny, or relationship of plant and animals has historically been based upon morphological features. With the advent of rapid and inexpensive DNA sequencing and the development of bioinformatics tools for comparing the genetic sequences from different organisms, we are now in the era of molecular phylogeny. Molecular characterization of both the relationship between species, and among organisms within the same species, has significant implications for public health.

Phylogenetic analysis employs the cumulative similarities and weighted polymorphisms found in collections of sequences to build a graphical representation of genetic relatedness. An illustrative example of how a phylogenetic analysis can be shown to infer relationships is presented in Figure 1. Suppose that viral sequences from seven different infected individuals are aligned on the right. One can easily observe that while all seven viruses are identical at the majority of base-pair positions, single nucleotide differences occurring at two positions allow grouping into three phylogenetically related clusters. Each of these monophyletic clusters is in turn linked through an inferred "most recent common ancestors" at the branch points between clusters. In order to accurately resolve the relative relatedness of sequences under investigation, it is critical that care is taken when selecting a suitable set of background sequences for inclusion in the analysis. Similar to the way in which using cats as a reference for comparing different dogs would highlight only the similarities in all canine breeds, using only distantly related reference sequences may cause investigational sequences to appear more highly related. While the basic methods of comparing genetic sequences and quantifying the certainty of the predicted relationships was developed in the 1960s, with the advent of widely available DNA sequencing, ever more sophisticated methods were developed. In the past few years, powerful desktop computers and freely available bioinformatics programs have allowed phylogenetics to become an important tool in increasing our understanding of the epidemiology of infections.[1]

Application of phylogenetic tools in the context of epidemiology have led to startling revelations and helped to answer and illuminate previously unanswerable questions. The paradigm of the management of TB changed after the publication of an article in the Lancet in 1999.[2] It was traditionally assumed that people with TB disease who had no organisms seen on a sputum smear were relatively non-infectious. In this report, Behr and colleagues used DNA fingerprinting on all of the TB cases in San Francisco between 1991 and 1996, to identify clusters of related infections. Through the use of molecular epidemiology, the authors determined that nearly 20% of all TB transmissions were contracted from those patients previously categorized as smear-negative and "non-infectious". Along similar lines, early in the recent H1N1 influenza epidemic, Smith et al.[3] used phylogenetics to determine the origins of the emergent strain. They were able to show that the pandemic variant was the result of reassortment of existing swine influenza viruses that had been circulating for a decade or more. This analysis revealed that the havoc caused by H1N1 resulted from recombination of mundane circulating viruses rather than the emergence of a novel variant such as H5N1 (bird flu).

Author Affiliations

1. National HIV & Retrovirology Laboratory, National Microbiology Laboratory, Public Health Agency of Canada, Ottawa, ON
2. Department of Medicine, University of Ottawa, Ottawa, ON

Correspondence: Dr. James Brooks, National HIV & Retrovirology Laboratory, Rm 3172, Building #6, 100 Eglantine Driveway, AL0603A2, Tunney's Pasture, Ottawa, ON K1A 0K9, Tel: 613-946-0120, Fax: 613-957-1163, E-mail: james.brooks@phac-aspc.gc.ca

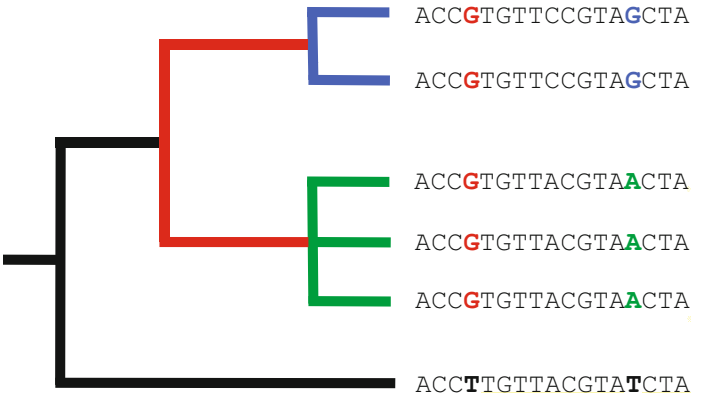
Conflict of Interest: None to declare.

Phylogenetics is ideally suited for examining HIV transmission patterns which have proved challenging to study due to veiled modes of transmission (sex and intravenous drug use) and the interval between infection and diagnosis. Owing to rapid error-prone replication, HIV-1 transmitting through a network of individuals acquires a genetic signature allowing the sequences to serve as proxies for epidemiological links among these individuals.[4] While not providing information on routes or direction of transmission, the phylogenetic relationships can be used to infer an association within a transmission chain. Two examples of the insights obtained through HIV phylogenetics in public health are the paradigm that primary HIV infections play a disproportionate role in driving the epidemic,[5] and that drug-resistant HIV is capable of being transmitted and of establishing infection.[6] By using anonymized population-based sequence data, HIV phylogenetics may significantly enhance traditional epidemiologic surveillance.

HIV phylogenetics has also been used as evidence in two recent criminal prosecutions in Canada involving transmission of HIV between an HIV-infected man and, in each case, several women. Phylogenetic tools were employed by the prosecution to support the case that the complainants were infected with a virus similar to that of the accused. It is important to note the limitations of phylogenetics as it applies to criminal cases.[7] In these specific cases, the relationship between the accused and the complainants, and the coincidental times of HIV infection, were already established using standard criminal investigation methods. In isolation, demonstrating clustering of individual infections on a phylogenetic tree does not resolve the direction or timing of transmission events. The possibility that sampling missed individuals involved in a transmission chain means that it is impossible to infer by phylogenetics alone, the source or sources of the individual infections. Paradoxically, and most pertinent to criminal cases, is the power of phylogenetics to reliably prove the absence of any relationship. For example, if the analysis fails to demonstrate an association between viral sequences from the accused and complainants, then this information provides conclusive evidence against any transmission linkage. There are also methodological considerations in determining the reliability of sequence associations found using phylogenetic analysis. Laboratory protocols for forensics require strict attention to: chain of custody; work flow; lab design; elimination of operator bias; the absence of specimen mix-up; and ultra-stringent contamination control.[8]

An obvious concern to health care providers, who order individual HIV drug resistance tests, is the potential use of these data to elucidate putative transmission chains.[9] One could envision scenarios wherein drug resistance genotyping data derived from a specific patient, who has been counseled with respect to transmission risks, might be phylogenetically linked to a cluster of new HIV infections. A possible conclusion might be that this individual under treatment, who is aware of his/her infection, is transmitting the virus through unsafe practices. Due to the limitations of phylogenetics described above, finding two HIV sequences that are similar does not prove a direct physical connection between the individuals infected with those viruses. Imagine the historical scenario where two individuals are infected with near identical strains of HIV transmitted through the blood supply. Not only was there never intimate contact between these individuals, but in fact they may never have been within 1000 km of each other. The ability of phylogenetics to support apparent transmission chains is only valid in the context of establishing, through traditional evidence-gathering means, the actual relationship of individuals infected with similar viruses.

Figure 1. Hypothetical phylogenetic tree



The relationship among sequences is represented by the colour coding of the bases in the sequence which corresponds to the colour of ancestral lineage that is associated with that sequence. For example, the red G shared among the first 5 sequences indicates a common lineage as shown by the red lines in the figure. These sequences are then separated into different taxa based upon further base changes associated with distinct lineages, shown in blue and green.

Between the potential power of phylogenetics to exclude the association of sequences and the inability of phylogenetics to prove the direct association between HIV-infected individuals, where does the utility of these tools lie? In a simple example, one can envision an anonymized HIV sequence database curated by a blinded third party that is monitored for the association between sequences. If a new sequence is found to be associated with two or more sequences in the database, then public health could be notified of the simple fact that the last sequence submitted was found to be part of a cluster. These data may be useful to augment contact tracing investigations; as the database consists of anonymous sequences and can be re-randomized daily, this use in and of itself presents a very small risk of breaching confidentiality. More importantly, this example demonstrates how phylogenetics can trigger a public health investigation that may result in a reduction in onward HIV transmission.

As the analytic tools within the field of HIV become more refined, the considerations surrounding the application of these methods become more complex. For the foreseeable future, a single sequence obtained from routine drug resistance testing is expected to reveal little about the timing of infection. However, there are increasingly complex methods of analysis utilizing next generation sequencing (NGS) that reveal much more about the history of the virus. Using NGS results from specimens obtained at two different time points, a date of infection could be determined that was highly correlated with the estimated time of infection derived using other methods.[10] In addition, evidence of the virus' history, in terms of the imprint from the immune system of previous hosts, is detectable in the viral sequence.[11] One can imagine that as these techniques mature, the association between viruses, inferred by the metadata encoded in viral sequence, could lead to more robust association being established among viral sequences.

These advances in analysis should not preclude the use of phylogenetics but instead serve to impress upon users the importance of appropriate safeguards, such as thorough sequence anonymization.

Phylogenetics can serve as a powerful adjunct to investigation of HIV transmission in the field of public health and elsewhere; however, using phylogenetics in HIV raises numerous ethical concerns, such as: What is being consented to at time of diagnosis and treatment? Does agreeing to HIV testing mean that you are consenting for the state to identify your most intimate contacts? What about the use of this information in civil or workman's compensation cases where the burden of proof is lower? In order to protect their privacy, individuals may avoid actions that benefit their health, such as HIV diagnosis or drug resistance testing. Not only could this behaviour be detrimental to the individual, but public health could be compromised due to the potential for an increase in new infections and increased transmission of resistant viruses. A rational, informed discussion on these issues is vital to ensuring that we do not throw out public health benefits with a poorly considered approach to the phylogenetics bathwater.[12]

REFERENCES

1. Hall BG, Barlow M. Phylogenetic analysis as a tool in molecular epidemiology of infectious diseases. *Ann Epidemiol* 2006;16(3):157-69.
2. Behr M, Warren S, Salamon H, Hopewell P, de Leon AP, Daley C, et al. Transmission of Mycobacterium tuberculosis from patients smear-negative for acid-fast bacilli. *Lancet* 1999;353(9151):444-49.
3. Smith GJD, Vijaykrishna D, Bahl J, Lycett SJ, Worobey M, Pybus OG, et al. Origins and evolutionary genomics of the 2009 swine-origin H1N1 influenza A epidemic. *Nature* 2009;459(7250):1122-25.
4. Hué S, Clewley JP, Cane PA, Pillay D. HIV-1 pol gene variation is sufficient for reconstruction of transmissions in the era of antiretroviral therapy. *AIDS* 2004;18(5):719-28.
5. Brenner BG, Roger M, Routy JP, Moisi D, Ntemgwa M, Matte C, et al. High rates of forward transmission events after acute/early HIV-1 infection. *J Infect Dis* 2007;195(7):951-59.
6. Kaye M, Chibo D, Birch C. Phylogenetic investigation of transmission pathways of drug-resistant HIV-1 utilizing pol sequences derived from resistance genotyping. *J Acquir Immun Defic Syndr* 2008;49(1):9-16.
7. Pillay D, Rambaut A, Geretti AM, Brown AJL. HIV phylogenetics. *BMJ* 2007;335(7618):460-61.
8. 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(6):382-87.
9. Brown AE, Murphy G, Rinck G, Clewley JP, Hill C, Parry JV, et al. Implications for HIV testing policy derived from combining data on voluntary confidential testing with viral sequences and serological analyses. *Sex Transm Infect* 2009;85:4-9.
10. Poon AFY, McGovern RA, Mo T, Knapp DJHF, Brenner B, Routy JP, et al. Dates of HIV infection can be estimated for seroprevalent patients by coalescent analysis of serial next-generation sequencing data. *AIDS* 2011;25(16):2019-26.
11. Brumme ZL, John M, Carlson JM, Brumme CJ, Chan D, Brockman MA, et al. HLA-associated immune escape pathways in HIV-1 subtype B Gag, Pol and Nef proteins. *PLOS ONE* 2009;4(8):e6687.
12. Hecht FM, Wolf LE, Lo B. Lessons from an HIV transmission pair. *J Infect Dis* 2007;195(9):1239-41.

Received: January 16, 2013
Accepted: July 24, 2013

RÉSUMÉ

La phylogénie est l'application d'études comparatives de séquences génétiques afin d'inférer des liens évolutionnaires entre des organismes. Cet outil peut être utilisé comme une forme d'épidémiologie moléculaire pour améliorer la surveillance classique des maladies transmissibles dans la population. Des études phylogénétiques ont donné lieu à la création de nouveaux paradigmes dans le domaine des maladies transmissibles, et notre commentaire vise à expliquer l'utilisation possible de la phylogénie pour retracer des maladies infectieuses. Nous insistons surtout sur l'application de la phylogénie en tant qu'outil pour élucider les grandes tendances de transmission du VIH et sur les contraintes de ces méthodes lorsqu'elles sont appliquées à l'analyse médico-légale. En santé publique, il est indispensable de comprendre l'épidémiologie des maladies infectieuses pour prévenir les nouvelles transmissions. Toutefois, l'augmentation de la résolution épidémiologique pourrait s'accompagner d'une perte de confidentialité pour les particuliers. Dans ce contexte, nous voulons lancer un débat sur l'utilisation de la phylogénie pour atteindre d'importants objectifs de santé publique tout en protégeant les droits des particuliers.

MOTS CLÉS : phylogénie; VIH; santé publique; criminalisation du VIH; épidémiologie moléculaire