Evolution in the Courtroom: Using Phylogenetics to Investigate Legal Claims of HIV Transmission

Benjamin Toups and Jeremy M. Brown

Department of Biological Sciences and Museum of Natural Science
Louisiana State University, Baton Rouge, LA, 70803

DNA sequences have become ubiquitous across the biological sciences and are even embedded in the public psyche, perhaps most famously in the context of forensic science. A human being's DNA changes very little over his or her lifetime, and this inherent stability lends itself well to positively identifying individuals using DNA samples. However, not all genomes are so stable, even over short timespans. One particularly dramatic example is human immunodeficiency virus (HIV-1). Unlike the human genome, the HIV-1 genome has an extraordinarily high mutation rate. This, in combination with recombination, rapid proliferation, and strong selection exerted by host immune systems, leads to exceptionally fast rates of evolution. The result of these interacting processes is a population of diverse and dynamically evolving HIV-1 genomes in the host, which is one reason why the virus is so difficult to eradicate. HIV-1's rapid rate of evolution also prevents the use of standard DNA fingerprinting techniques that rely on stable, unchanging genomes to connect the infections in different individuals, but such rapid evolution does lend itself particularly well to phylogenetic analysis.

Molecular phylogenetics uses (portions of) genomes to reconstruct evolutionary relationships. Often, these relationships result from divergence over millions of years of gradual evolution. However, pathogens like HIV-1 evolve so quickly that we can reconstruct their history over timescales of only a few years or even months. Several previous studies have established that phylogenetic analysis of HIV-1 genomes could be used to test hypotheses of person-to-person transmission, perhaps most famously in a case regarding transmission of HIV-1 from a Florida dentist to multiple patients in the late 1980s (e.g., Ou et al. 1992, Hillis and Huelsenbeck 1994). More recent applications of this approach (e.g., Metzker et al. 2002) have raised the standards of such analyses further by employing carefully constructed *a priori* hypotheses, more appropriate statistical methods, corroboration with detailed epidemiological information, and replication. As a result, phylogenetic analyses have now been presented as evidence in courtroom proceedings.

One shortcoming of early investigations into HIV-1 transmission was a lack of explicit tests regarding the direction of transmission. However, the biology of HIV-1 suggests the use of a particular phylogenetic signature for establishing directionality - paraphyly. Paraphyly occurs when a subset of a source's viral sequences is more closely related to recipient sequences than to other source sequences. In the context of HIV-1, this pattern results from a genetic bottlenecking event that commonly occurs during transmission where only a single virus

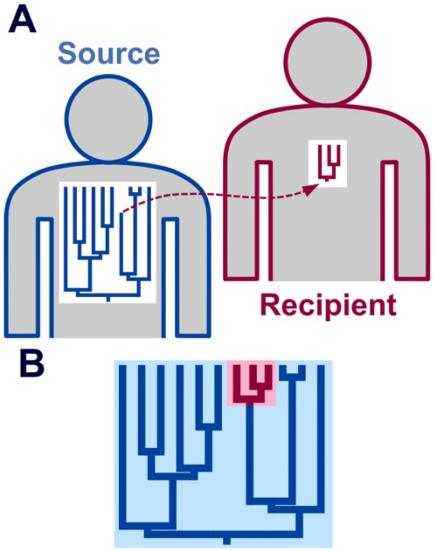

Figure 1 – (A) A hypothetical transmission scenario showing the phylogenetic trees relating viral sequences in a source and a recipient individual. The red dashed line depicts a transmission event where one viral lineage from the source establishes a productive infection in the recipient. Sampled lineages from each individual are represented by the tips at the top of the trees. The structure of these trees is a result of within-host diversification of HIV-1. (B) The overall phylogenetic tree with lineages sampled from both source and recipient. If their identities were unknown, the nested relationship of recipient lineages within those from the source would indicate the identity of the source. This nested relationship, where some source lineages are more closely related to those from the recipient, is known as paraphyly.

from the source successfully establishes a productive infection in a recipient (**Fig. 1**). Some of the remaining viruses in the source are closely related to the one that was transferred, while others are not. However, time is of the essence when sampling case individuals. As time passes, the source and recipient sequences will continue to evolve, eventually leading to loss of the paraphyletic signature.

In 2010, Scaduto et al. took phylogenetic investigation of HIV-1 transmission one step further by using the signature of paraphyly to predict who the source of infection might be among a set of individuals. These researchers conducted analyses for two court cases (*State of Washington vs. Anthony Eugene Whitfield, State of Texas vs. Phillipe Padieu)* in which one individual, the defendant in each case, was accused of transmitting the virus to multiple victims. Scaduto et al. (2010) investigated both the levels of relatedness among case sequences compared to non-case sequences, as well as the potential presence of direction-indicating paraphyly. To eliminate any biases, they only had access to blinded samples where the identities of all individuals were unknown to the researchers. In both cases, samples from the *pol* and *env* genes in the viral genomes were used to conduct phylogenetic analyses.

In the Washington case, results from the *env* gene showed that all viral sequences from each case individual were each other's closest relatives, except for one – a case sample labeled WA04. Some viruses from WA04 were more closely related to those from other case samples than they were to each other. WA04's viruses were paraphyletic with respect to four of the five other case samples, providing evidence not only that these infections were epidemiologically linked, but that transmission had occurred from WA04 to these four other individuals. At trial, WA04 was identified as the defendant. The group performed similar analyses in the

Texas case, in which a tree inferred from the *pol* gene **(Fig. 2)** showed paraphyly of the viral genomes of one individual, CC01, with respect to viral genomes from all six other case individuals (CC02-CC07; CC08 was added after trial). As with the Washington case, this paraphyletic pattern supported the hypothesis of transmission from individual CC01 to all other case individuals, and once again at trial it was revealed that CC01 was the defendant.

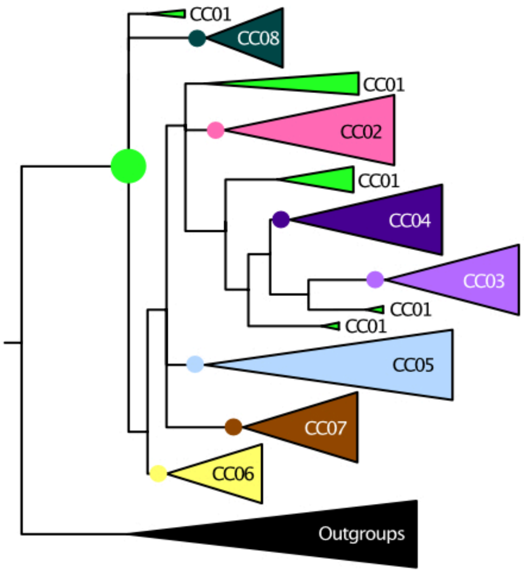

Figure 2 – Phylogeny inferred from the *pol* gene in the Texas case. The most recent common ancestor of each individual's viral sequences is shown using a dot that is colored by individual. All case samples are more closely related to each other than to samples not related to the case (outgroups). Subsets of CC01 sequences are more closely related to sequences from other case individuals than to other sequences from CC01, demonstrating the paraphyly used to infer directionality of transmission in this case.

For all the strengths of phylogenetics to address hypotheses of viral transmission, the approach is still only able to provide information about the direction of transmission and cannot speak to motive or means. In both cases investigated by Scaduto et al (2010), other circumstantial and epidemiological evidence was presented in tandem with the phylogenetic conclusions, which led to convictions of both defendants. While this study demonstrates the power of phylogenetic tools in forensic studies of HIV-1 transmission, such approaches are widely applicable to many other rapidly evolving pathogens and in other contexts, including epidemiology and public health.

References

Hillis DM, Huelsenbeck JP. 1994. Support for dental HIV transmission. *Nature.* 369:24-25.

Metzker ML, Mindell DP, Liu X-M, Ptak RG, Gibbs RA, Hillis DM. 2002. Molecular evidence of HIV-1 transmission in a criminal case. *Proc. Natl. Acad. Sci. USA.* 99:14292–14297.

Ou C-Y, Ciesielski CA, Myers G, Bandea CI, Luo C-C, Korber BTM, Mullins JI, Schochetman G, Berkelman RL, Economou AN, Witte JJ, Furman LJ, Satten GA, Macinnes KA, Curran JW, Jaffe HW, Laboratory Investigation Group, Epidemiologic Investigation Group. 1992. Molecular epidemiology of HIV transmission in a dental practice. *Science* 256:1165–1171.

Scaduto DI, Brown JM, Haaland WC, Zwickl DJ, Hillis DM, Metzker ML. 2010. Source identification in two criminal cases using phylogenetic analysis of HIV-1 DNA sequences. *Proc. Natl. Acad. Sci. USA* 107:21242–21247.