Source identification in two criminal cases using phylogenetic analysis of HIV-1 DNA sequences

Diane I. Scaduto  
*Baylor College of Medicine*

Jeremy M. Brown  
*The University of Texas at Austin*

Wade C. Haaland  
*Baylor College of Medicine*

Derrick J. Zwickl  
*The University of Texas at Austin*

David M. Hillis  
*The University of Texas at Austin*

See next page for additional authors

Follow this and additional works at: [https://digitalcommons.lsu.edu/biosci_pubs](https://digitalcommons.lsu.edu/biosci_pubs)

**Recommended Citation**


This Article is brought to you for free and open access by the Department of Biological Sciences at LSU Digital Commons. It has been accepted for inclusion in Faculty Publications by an authorized administrator of LSU Digital Commons. For more information, please contact ir@lsu.edu.
Source identification in two criminal cases using phylogenetic analysis of HIV-1 DNA sequences

Diane I. Scaduto\textsuperscript{a,b}, Jeremy M. Brown\textsuperscript{c,1}, Wade C. Haaland\textsuperscript{a,b}, Derrick J. Zwickl\textsuperscript{c,2}, David M. Hillis\textsuperscript{c,3}, and Michael L. Metzker\textsuperscript{a,b,d}

\textsuperscript{a}Human Genome Sequencing Center, \textsuperscript{b}Department of Molecular and Human Genetics, and \textsuperscript{c}Cell and Molecular Biology Program, Baylor College of Medicine, Houston, TX 77030; and \textsuperscript{d}Section of Integrative Biology and Center for Computational Biology and Bioinformatics, University of Texas, Austin, TX 78712

This contribution is part of the special series of Inaugural Articles by members of the National Academy of Sciences elected in 2008.

Contributed by David M. Hillis, October 20, 2010 (sent for review September 22, 2010)

Phylogenetic analysis has been widely used to test the a priori hypothesis of epidemiological clustering in suspected transmission chains of HIV-1. Among studies showing strong support for relatedness between HIV samples obtained from infected individuals, evidence for the direction of transmission between epidemiologically related pairs has been lacking. During transmission of HIV, a genetic bottleneck occurs, resulting in the paraphyly of source viruses with respect to those of the recipient. This paraphyly establishes the direction of transmission, from which the source can then be inferred. Here, we present methods and results from two criminal cases, \textit{State of Washington v Anthony Eugene Whitfield}, case number 04-1-0617-5 (Superior Court of the State of Washington, Thurston County, 2004) and \textit{State of Texas v Philippe Padiou}, case numbers 219-82276-07, 219-82277-07, 219-82278-07, 219-82279-07, 219-82280-07, and 219-82276-07 (219th Judicial District Court, Collin County, TX, 2009), which provided evidence that direction can be established from blinded case samples. The observed paraphyly from each case study led to the identification of an inferred source (i.e., index case), whose identity was revealed at trial to be that of the defendant.

HIV transmission | phylogeny | forensics | evolution | molecular epidemiology

DNA profiling technology has been successfully used to link suspects to crime scenes; identify victims of accidents, disasters, and wars; and exonerate wrongly convicted prisoners (1). Stable human genetic variation allows for definitive identification of individuals because our genome remains relatively unchanged (2) during our lifetime as a result of efficient DNA repair systems (3). In contrast, individuals infected with HIV-1 contain a dynamically evolving population of related genomes. Factors contributing to the expansion of multiple viral lineages are high mutation (4–6) and recombination rates (7–9), coupled with an estimated replicative production of $10^{8}$ to $10^{10}$ virions per day (10–12). This expansion is offset by lineage extinction from the production of defective nonreplicating virions (13), the effectiveness of the host’s immune system, and the efficacy of highly active antiretroviral therapy (14). Although viral dynamics limit an investigator from using the common practice of matching DNA profiles, phylogenetic methods are ideally suited for determining the HIV pattern of descent in cases of suspected transmission between individuals.

The case of a Florida dentist was a high-profile investigation involving the phylogenetic relationships of HIV-1 in different individuals and establishing that viral sequences from the dentist and six of his patients were more closely related to each other than to unrelated controls (15, 16). Other phylogenetic studies have provided support for the transmission of HIV-1 from a French surgeon (17) and a French nurse (18) to their respective patients while receiving care in the hospital. Investigators have also established that phylogenetic methods can provide support against allegations of suspected transmission events. For example, the Centers for Disease Control and Prevention investigated the contention that a second Florida dentist infected 24 patients during invasive procedures and rejected the a priori hypothesis of suspected transmission based on phylogenetic analysis (19). Similarly, molecular evidence dismissed the assertion that a Baltimore surgeon (20) and a UK obstetrician/gynecologist (21) infected their respective patients while providing care. These studies establish an important touchstone of objectivity for the use of phylogenetic methods in providing strong support for or against allegations of suspected HIV-1 transmission events.

An early criminal case that used molecular evidence in support of an alleged rape and deliberate transmission of HIV-1 to a female victim occurred in Sweden (22). Other studies supporting criminal charges of HIV-1 transmission have been reported in Sweden (23) as well as in Australia (24, 25), Belgium (26), Denmark (27), Germany (28), and Scotland (29). Our group reported a US criminal case involving a gastroenterologist who was convicted of purposefully infecting his former girlfriend with blood or blood products obtained from a patient under his care (30). Phylogenetic analysis revealed that HIV-1 sequences obtained from the victim (the former girlfriend of the physician) were most closely related to those from the physician’s patient. Unlike the studies described above, our phylogenetic analysis also provided evidence about the direction of transmission and supported a transmission route from the physician’s patient to the victim. The identities of the case pair, however, were revealed to us at the start of the molecular investigation; thus, the study was not conducted with a blinded design.

Providing molecular evidence for the direction of transmission (source $\rightarrow$ recipient) would further strengthen the a priori hypothesis under investigation. This is possible if a paraphyletic relationship (i.e., a subset of source viral sequences is more closely related to all recipient sequences than to other source sequences) is observed in the phylogenetic tree. Despite the large population of related HIV genomes in infected individuals, paraphyly is the result of a significant genetic bottleneck when establishing
Phylogenetic Analysis. For the Washington case, all Washington case sequences were monophyletic in the pol and env gene regions with respect to BLAST-selected GenBank controls (Figs. 1 and 2, respectively), supported by significant Bayesian posterior probabilities and maximum likelihood (ML) bootstrap proportions (Table 1). A similar finding was observed when Washington case sequences from the pol and env gene regions (Figs. S1 and S2, respectively) were analyzed using HIV-1 sequences obtained from local controls (Table 1). The env sequence alignment contained many gaps, which have been shown to potentially cause error in phylogenetic inference (35). The env sequence alignment was also analyzed using both control sets by removing those sites containing gaps (Figs. S3 and S4), which gave results consistent with those of the entire dataset (Table 1). These data provide strong statistical evidence that all Washington case HIV-1 sequences are more closely related to each other than to either BLAST-selected GenBank or local controls.

For the Texas case, all Collin County samples were monophyletic in the pol and env gene regions with respect to BLAST-selected GenBank controls (Figs. 3 and 4, respectively), supported by significant Bayesian posterior probabilities and ML bootstrap proportions (Table 1). The env sequence alignment was further analyzed using BLAST-selected GenBank controls by removing sites containing gaps (Fig. S5), which gave results consistent with those of the entire dataset (Table 1). After trial, one additional individual, CC08, was analyzed by phylogenetic methods. All case samples, including CC08, remained monophyletic in both gene regions with respect to BLAST-selected GenBank controls (Figs. S6–S8). These data provide strong statistical evidence that all Collin County HIV-1 sequences form a monophyletic clade with respect to BLAST-selected GenBank controls.

Direction of Transmission. If paraphyly among case samples is observed in the phylogenetic tree, the direction of transmission can be inferred. For the Washington case, pol phylogenetic trees showed a monophyletic cluster of HIV sequences sampled from each individual (Fig. 1 and Fig. S1). Viral sequences from WA04cd, however, exhibited a paraphyletic relationship in env phylogenetic trees with those from WA01yn, WA03pe, WA05vt, and WA06tk, wherein the most recent common ancestor of sequences from WA04cd is shown by a filled red circle (Fig. 2 and Figs. S2–S4). Sequences from WA02qd diverged before the most recent common ancestor of WA04cd sequences when the analysis was based on BLAST-selected GenBank controls (Fig. 2), but WA04cd sequences were paraphyletic with respect to WA02qd sequences when analyzed with local controls (Fig. S2) or when gaps were removed in sequence alignments with either control set (Figs. S3 and S4). Based on these analyses, we inferred that the
direction of transmission occurred from WA04cd to all other Washington case individuals. At trial, the identity of sample WA04cd was revealed to be that of Anthony Eugene Whiffen.

For the Texas case, viral sequences from CC01 were paraphyletic in the pol phylogenetic trees with respect to those from all other Collin County samples, wherein the filled red circle indicates the most recent common ancestor of sequences from CC01 (Fig. 3 and Fig. S6). A similar finding was observed in the env trees, except that CC05 (Fig. 4 and Figs. S5, S7, and S8) and CC08 (Figs. S7 and S8) could not be definitively identified as transmission recipients based on the paraphyly of CC01 sequences. From these analyses, we inferred that the direction of transmission occurred from CC01 to all other Collin County case individuals. At trial, the identity of sample CC01 was revealed to be that of Philippe Padieu.

Discussion

State of Washington v Anthony Eugene Whiffen and State of Texas v Philippe Padieu represent the second and third US criminal cases, respectively, using phylogenetic analysis as part of the overall evidence. These independent cases support the

Table 1. Bayesian posterior probabilities and ML nonparametric bootstrap proportions supporting monophyly of the transmission groups in the Washington and Texas cases

<table>
<thead>
<tr>
<th>Gene regions</th>
<th>pol</th>
<th>env</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Method</strong></td>
<td>Bayesian posterior probability</td>
<td>ML bootstrap proportion</td>
</tr>
<tr>
<td>Washington case</td>
<td>GenBank controls (entire dataset)</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>Local controls (entire dataset)</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>GenBank controls (gaps removed)</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td>Local controls (gaps removed)</td>
<td>N/A</td>
</tr>
<tr>
<td>Texas case</td>
<td>GenBank controls (entire dataset)</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>GenBank controls (gaps removed)</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td>GenBank controls with CC08 (entire dataset)</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>GenBank controls with CC08 (gaps removed)</td>
<td>N/A</td>
</tr>
</tbody>
</table>

N/A, not analyzed because the pol gene region did not contain gaps.
use of phylogenetic analysis to test a priori transmission hypotheses, both by linking epidemiologically related individuals and by providing evidence of the direction of transmission between individuals. From these analyses, an index case was inferred for each study based on the observed paraphyly. In addition, the identity of anonymously coded case samples at trial, the inferred index case was identified as the defendant in each case.

The paraphyly of source sequences with respect to recipient sequences is expected to decline over time. Loss of diversity can occur within individuals as a result of lineage extinction (13), as well as elimination of some variants by the host’s immune system and antiretroviral therapy (14). Such loss is expected eventually to lead to monophyly of the surviving viral lineages within the source individual (30). This is consistent with the loss of paraphyly in the env phylogenetic trees for viral sequences from CC05 and CC08 with respect to those from CC01, both of whom reported an earlier sexual relationship with the defendant than the other Collin County partners described (SI Text). In addition, recombination among viral sequences within the source individual will degrade support for particular paraphyletic relationships over time. Therefore, strongly supported paraphyly can provide evidence to infer direction of transmission between pairs of epidemiologically related individuals; however, a lack of paraphyly cannot be used to refute a possible transmission route.

Phylogenetic analysis can be informative regarding epidemiological relationships among and transmission direction between individuals, although caution should be exercised in conducting such analysis. In particular, alternative sources of infection should carefully be considered and experiments should test a priori hypotheses of those relationships. Linking sequences from a database, with no a priori evidence of relationships, is likely to result in many missed intermediate links of a transmission chain. The interpretation of phylogenetic trees regarding hypothesized transmission scenarios should therefore be weighed appropriately. Moreover, phylogenetic trees remain statistical estimates, subject to several key assumptions, and do not carry the same degree of certainty as human DNA profiling technology, which does not require the need to model sequence changes over time and considers only two hypotheses (i.e., matching and nonmatching). Also noteworthy is that the molecular evidence cannot provide any support for the motivation behind the acts of exposure or transmission of HIV-1. Similar to a Louisiana case (30), the phylogenetic data for both the Washington and Texas cases represented part of the overall evidence that was presented at trial, with additional facts being presented regarding the means, motive, and opportunity for transmission of HIV. In each case, the defendant was charged with intentionally exposing and, in some instances, infecting his female partners with HIV, with the motivation of each defendant being weighed alongside other evidence presented at trial.

In 2004, Anthony Eugene Whitfield was convicted on 17 counts of first-degree assault with sexual motivation, 2 counts of witness tampering, and 3 counts of no-contact order violations. Of the 17 victims, 5 were infected with HIV. The prison terms for the first-degree assaults were ordered to be served consecutively, totaling 178 y and 1 mo, with the remaining counts to be served concurrently. In 2009, Philippe Padieu was convicted on six counts of aggravated assault with a deadly weapon, receiving five 45-y and one 25-y prison terms, to be served concurrently. In 2009, Philippe Padieu was convicted on six counts of aggravated assault with a deadly weapon, receiving five 45-y and one 25-y prison terms, to be served concurrently.

The recent enactment of national laws that criminalize transmission of HIV in more than a dozen African countries has led...
to considerable controversy (36–38). Although the goal of the Action in West Africa (AWARE-HIV/AIDS) is to improve health and contribute to political stability and economic prosperity within the region (39), some argue that criminalization of HIV is likely to have a negative impact on public health and human rights. Those opposing criminalization propose that prosecution should be limited only to those charged with malicious intent under the provisions of general criminal laws and not those specific to HIV. In North America and Europe, public health and legal professionals have also raised concerns that the criminalization of HIV provides a disincentive to voluntary disclosure of HIV status as well as uptake of HIV testing (29, 40–42). In the United States, 37 states have specific HIV laws (including Washington but not Texas) (43), and in 2009, 35 US news reports across 21 states described individuals being involved in HIV-related crimes (Table S1). The most common charge from these reports was exposure of HIV to sexual partners without disclosing known status (Table S1). Finding the appropriate balance of responsibility of those living with HIV and those not, the protection of human rights for all, and the legislative policies that promote public awareness and education while prosecuting those charged with wrongful acts will be challenging but important in reducing the transmission of HIV globally.

Materials and Methods

Protocols, Chain of Custody, and Receipt of Anonymously Labeled Blood Samples. Before the commencement of each molecular investigation, a case study protocol was provided to and reviewed by the prosecution and forwarded to the defense team. On approval of the protocol, the prosecution then coded blood samples drawn from case individuals and maintained their secrecy throughout each molecular investigation, labeling them as WA01yn, WA02q, WA03pe, WA04cd, WA05vt, and WA06tk for the Washington case and CC01 through CC07 for the Texas case. Chain of custody was strictly maintained for each blood sample, drawn by experienced medical personnel and witnessed by law officers. Only one blood sample was delivered to Baylor College of Medicine and processed at a time. On completion of the analysis, all biological materials associated with the sample were removed before receiving the next blood sample (SI Text and Table S2).

HIV-1 Sequences from Case Samples and Controls. Similar to the Louisiana case protocol (30), genomic DNA was extracted from each blood sample. PCR was used to amplify two products from the pol and env gene regions of the HIV-1 genome, as previously described (30, 44). Approximately 20 molecular clones derived from these regions were sequenced using the automated Sanger method (45). HIV-1 sequences from each case sample were examined against those previously obtained from the Louisiana case (30) and the laboratory sequence pNL4-3 and did not reveal any evidence of laboratory contamination (SI Text).

Pairwise distances were calculated to identify the two most divergent HIV-1 sequences from molecular clones for each case sample (Tables S3 and S4). Those sequences were then used to search GenBank (release 137 and 138 for the Washington case and release 166 for the Texas case) using BLAST (46) to identify the best-matching HIV-1 DNA sequences according to score significance. We rationalized that best-matching HIV-1 DNA sequences would increase the probability of refuting the a priori hypothesis that case samples were involved in an alleged HIV-1 transmission chain. To maximize the number of unrelated but high-scoring HIV-1 sequences from GenBank, those related to the same research study were excluded, even when significant BLAST scores were obtained. Approximately 20 GenBank sequences were selected for each study (SI Text). Blood samples from 21 unrelated HIV-infected individuals were also collected from the Puget Sound, Tacoma, Centralia, and Olympia, WA areas and used as local controls in the Washington case. Of the 21 samples, 1 failed to yield genomic DNA and 2 failed PCR (SI Text). The two most divergent sequences for both gene regions from 18 local controls were used in the phylogenetic analysis. Given our experience that BLAST-selected GenBank sequences provided appropriate control sets for the Louisiana (30) and Washington cases (SI Text and Table S5), local controls were not obtained for

![Fig. 4.](image_url)
the Texas case. Multiple DNA sequence alignments for each gene region of rion (AIC) was used to choose a model for each gene region as a whole and sequence evolution with for analyzing our data, 24 models were considered, ranging from simple alism and statistical tractability (47). To identify the model most appropriate tant, and to perform well, it must strike a balance between biological re-
sults conservative, because support for particular paraphyletic relationships would decrease in recombinant sequences. Briefly, the methods used in the subsequent analysis of the data are summarized. Model selection is impor-
tant, and to perform well, it must strike a balance between biological re-
subsequent analysis of the data are summarized. Model selection is impor-