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

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

[a]Human Genome Sequencing Center, [d]Department of Molecular and Human Genetics, and [b]Cell and Molecular Biology Program, Baylor College of Medicine, Houston, TX 77030; and [c]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, *State of Washington v Anthony Eugene Whitfield*, case number 04-1-0617-5 (Superior Court of the State of Washington, Thurston County, 2004) and *State of Texas v Philippe Padieu*, case numbers 219-82276-07, 219-82277-07, 219-82278-07, 219-82279-07, 219-82280-07, and 219-82705-07 (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 inferring 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 → 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 pro-

Author contributions: D.I.S., W.C.H., and M.L.M. performed research; D.I.S., J.M.B., W.C.H., D.J.Z., D.M.H., and M.L.M. analyzed data; J.M.B., D.J.Z., D.M.H., and M.L.M. designed research; and J.M.B., D.M.H., and M.L.M. wrote the paper.

Conflict of interest statement: This paper reports the results of two criminal investigations, one in Washington and one in Texas. M.L.M. and D.M.H. were retained as expert witnesses for the Washington case, and M.L.M. was retained as an expert witness for the Texas case. The cases are concluded, and there are no continuing financial interests of any of the authors.

Data deposition: The sequences reported in this paper have been deposited in the GenBank database (accession nos. HQ537787–HQ538418).

[1]Current address: Department of Integrative Biology, University of California, Berkeley, CA, 94720.

[2]Current address: Department of Ecology and Evolutionary Biology, University of Kansas, Lawrence, KS 66045.

[3]To whom correspondence should be addressed. E-mail: dhillis@mail.utexas.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1015673107/-/DCSupplemental.

ductive infection in a recipient (31). Several studies support the high probability of this pattern by demonstrating that the majority (>75%) of productive infections are derived from a single virus (32–34). Following initial infection, the rapid rate of evolution of HIV leads to increased diversity of HIV sequences within a newly infected individual. If HIV sequences are sampled from the source and recipient shortly after a transmission event, sequences from the source are expected to be paraphyletic with respect to all recipient sequences. The paraphyly provides support for the direction of transmission and, in a criminal case, could be used to identify the index case (i.e., source).

These findings led us to design a study to investigate whether a source could be identified using the criterion of paraphyly when relatedness between case individuals was examined using phylogenetic methods. Criteria necessary for the study were (*i*) the identities of case individuals being blinded to investigators, (*ii*) the handling of case samples being separated both temporally and spatially to eliminate the possibility of cross-contamination, and (*iii*) the allegation being made from multiple transmissions from a single source. Here, we present the molecular evidence used in two US criminal cases: *State of Washington v Anthony Eugene Whitfield*, case number 04-1-0617-5 (Superior Court of the State of Washington, Thurston County, 2004) and *State of Texas v Philippe Padieu*, case numbers 219-82276-07, 219-82277-07, 219-82278-07, 219-82279-07, 219-82280-07, and 219-82705-07 (219th Judicial District Court, Collin County, TX, 2009). For each case, the observed paraphyly in the phylogenetic analysis led to the identification of an index case, which, at trial, was revealed to be that of the defendant.

The Washington case was based on circumstantial evidence that the defendant intended to inflict "great bodily harm" by administering, exposing, or transmitting HIV to 17 female partners through unprotected sexual relations. Court records revealed that the defendant learned of his HIV status in April 1992. Between 1999 and 2004, he engaged in more than 1,000 oral, vaginal, and anal acts of unprotected sex with his female partners. The defendant never informed them of his HIV status and denied having any disease when asked. Five of the 17 female partners tested positive for HIV between May 2003 and March 2004, 2 of whom claimed that the defendant was their only sexual partner since 1999 (*SI Text*). The six case individuals formed the basis of the a priori hypothesis that the suspected transmission of HIV was from one source to multiple recipients.

The Texas case was based on circumstantial evidence that the defendant intentionally, knowingly, and recklessly caused "serious bodily injury" by exposing six female partners to HIV through unprotected sexual contact. Court records revealed that the defendant learned of his HIV status on September 12, 2005. The defendant never informed any of his sexual partners that he was HIV-positive, stating to them that he had tested negative for the virus. The six partners tested positive for HIV between April 2006 and March 2007, four of whom claimed that the defendant was their only sexual partner or that other partners had tested negative for the virus after their diagnosis (*SI Text*). The seven case individuals formed the basis of the a priori hypothesis that the suspected transmission of HIV was from one source to multiple recipients.

Results

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.

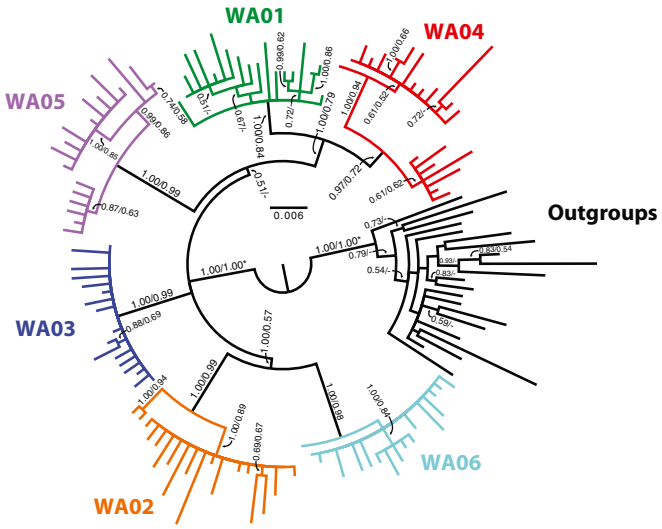

Fig. 1. Washington case: ML tree for the *pol* gene dataset using BLAST-selected GenBank controls. Branches are labeled with support values (Bayesian posterior probability/ML bootstrap proportion). Support values <0.5 are denoted as "–" or are not shown. Asterisks indicate that branches shown on either side of the root represent the same bipartition. Clades of viral sequences from different individuals are colored differently. No paraphyly was found in this tree for sequences from any case individual.

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

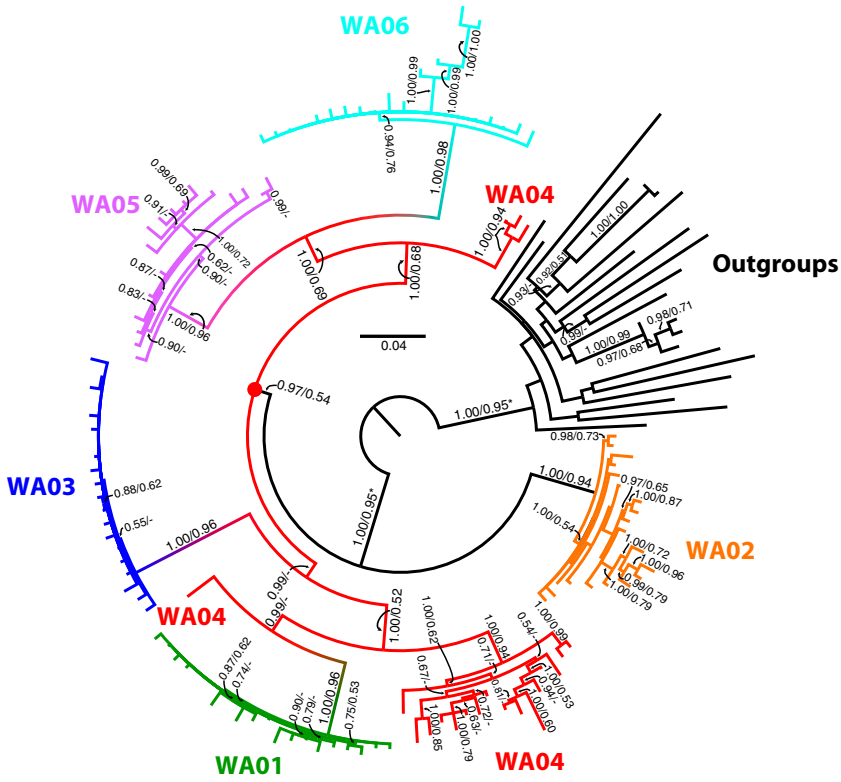

Fig. 2. Washington case: ML tree for the *env* gene dataset using BLAST-selected GenBank controls. Support values and clade coloring are as in Fig. 1. Color gradients along branches represent putative transmission events from WA04cd to other individuals.

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 Whitfield.

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 Whitfield 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

	Gene regions			
	pol		*env*	
Method	Bayesian posterior probability	ML bootstrap proportion	Bayesian posterior probability	ML bootstrap proportion
Washington case				
GenBank controls (entire dataset)	1.00	1.00	1.00	0.95
Local controls (entire dataset)	1.00	0.88	1.00	1.00
GenBank controls (gaps removed)	N/A	N/A	1.00	0.93
Local controls (gaps removed)	N/A	N/A	1.00	0.94
Texas case				
GenBank controls (entire dataset)	1.00	0.98	1.00	1.00
GenBank controls (gaps removed)	N/A	N/A	1.00	1.00
GenBank controls with CC08 (entire dataset)	1.00	0.99	1.00	1.00
GenBank controls with CC08 (gaps removed)	N/A	N/A	1.00	0.99

N/A, not analyzed because the *pol* gene region did not contain gaps.

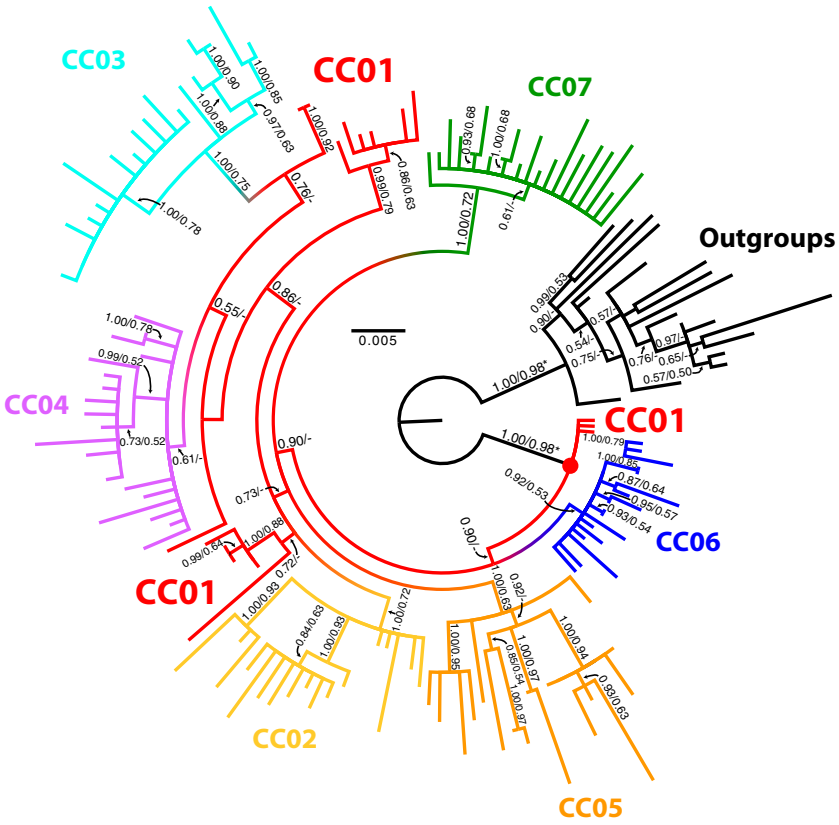

Fig. 3. Texas case: ML tree for the *pol* gene dataset using BLAST-selected GenBank controls. Support values and clade coloring are as in Fig. 1. HIV-1 sequences from CC01 (shown in red) are paraphyletic with respect to viral sequences from all other case individuals (CC02, CC03, CC04, CC05, CC06, and CC07). Color gradients along branches represent putative transmission events from CC01 to other individuals.

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. On revealing 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.

The recent enactment of national laws that criminalize transmission of HIV in more than a dozen African countries has led

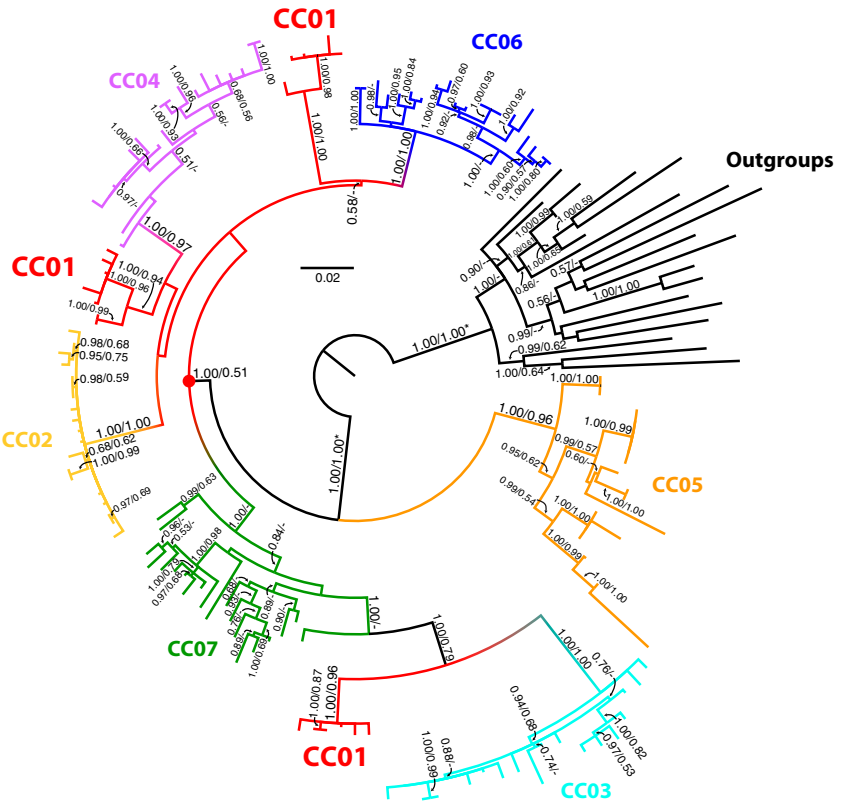

Fig. 4. Texas case: ML tree for the *env* gene dataset using BLAST-selected GenBank controls. Support values and clade coloring are as in Fig. 1. HIV-1 sequences from CC01 (shown in red) are paraphyletic with respect to viral sequences from individuals CC02, CC03, CC04, CC06, and CC07. Color gradients along branches represent putative transmission events from CC01 to other individuals. Additionally, sequences from CC07 are paraphyletic with respect to sequences from individual CC03 and some sequences from CC01.

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, WA02qd, 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

the Texas case. Multiple DNA sequence alignments for each gene region of cases and controls were aided by their corresponding protein alignments.

Phylogenetic Analysis. Details of the analysis originally presented in each court trial and the subsequent methods used here are outlined in *SI Text*. Recombination was not modeled explicitly in our datasets, making our results 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 important, and to perform well, it must strike a balance between biological realism and statistical tractability (47). To identify the model most appropriate for analyzing our data, 24 models were considered, ranging from simple (Jukes–Cantor) to complex (GTR + Γ + I; the general time-reversible model of sequence evolution with Γ-distributed rate variation across sites and an estimated proportion of invariable sites) (48). The Akaike Information Criterion (AIC) was used to choose a model for each gene region as a whole and for subsets corresponding to different codon positions. Statistical phylogenetic estimates were then conditioned on the model of sequence evolution assumed during analysis. Bayesian analyses were conducted both using the single AIC-chosen model across the entire dataset and multiple partitioned models, wherein separate AIC-chosen models were applied to independent codon positions. Bayes factors were used to compare unpartitioned with partitioned models, the latter of which were strongly supported. In one case (*env* gene dataset for the Washington case using local controls), support for the partitioned model was only modest [2ln(BF) = 6.3]. Nonetheless, we report our results using partitioned models to maintain consistency between analyses using BLAST-selected GenBank and local controls and to reduce error (i.e., modest overpartitioning generally induces fewer errors than modest underpartitioning) (49). ML estimation and ML nonparametric bootstrapping were also performed using partitioned models identical to those used in the Bayesian analyses.

ACKNOWLEDGMENTS. We thank Jodilyn Erikson-Muldrew and Dave Bruneau from the Thurston County District Attorney's Office and Diana T. Yu from the Thurston County Health Department for their help in the Washington case. We also thank Lisa Milasky King, Samme Glasby, Curtis Howard, and Roberto Chacon from the Collin County District Attorney's Office for their help in the Texas case. J.M.B. was supported by a Donald D. Harrington fellowship from the University of Texas, a National Science Foundation graduate research fellowship, and a National Science Foundation postdoctoral research fellowship in biological informatics (DBI 0905867). W.C.H. was supported by a National Science Foundation GK12 fellowship (DGE 0086397).

1. Jobling MA, Gill P (2004) Encoded evidence: DNA in forensic analysis. *Nat Rev Genet* 5: 739–751.
2. Hastings PJ, Lupski JR, Rosenberg SM, Ira G (2009) Mechanisms of change in gene copy number. *Nat Rev Genet* 10:551–564.
3. Misteli T, Soutoglou E (2009) The emerging role of nuclear architecture in DNA repair and genome maintenance. *Nat Rev Mol Cell Biol* 10:243–254.
4. Preston BD, Poiesz BJ, Loeb LA (1988) Fidelity of HIV-1 reverse transcriptase. *Science* 242:1168–1171.
5. Boyer JC, Bebenek K, Kunkel TA (1992) Unequal human immunodeficiency virus type 1 reverse transcriptase error rates with RNA and DNA templates. *Proc Natl Acad Sci USA* 89:6919–6923.
6. Mansky LM, Temin HM (1995) Lower *in vivo* mutation rate of human immunodeficiency virus type 1 than that predicted from the fidelity of purified reverse transcriptase. *J Virol* 69:5087–5094.
7. Zhuang J, et al. (2002) Human immunodeficiency virus type 1 recombination: Rate, fidelity, and putative hot spots. *J Virol* 76:11273–11282.
8. Jung A, et al. (2002) Recombination: Multiply infected spleen cells in HIV patients. *Nature* 418:144.
9. Rhodes T, Wargo H, Hu W-S (2003) High rates of human immunodeficiency virus type 1 recombination: Near-random segregation of markers one kilobase apart in one round of viral replication. *J Virol* 77:11193–11200.
10. Wei X, et al. (1995) Viral dynamics in human immunodeficiency virus type 1 infection. *Nature* 373:117–122.
11. Ho DD, et al. (1995) Rapid turnover of plasma virions and CD4 lymphocytes in HIV-1 infection. *Nature* 373:123–126.
12. Perelson AS, Neumann AU, Markowitz M, Leonard JM, Ho DD (1996) HIV-1 dynamics in vivo: Virion clearance rate, infected cell life-span, and viral generation time. *Science* 271:1582–1586.
13. Coffin JM (1995) HIV population dynamics in vivo: Implications for genetic variation, pathogenesis, and therapy. *Science* 267:483–489.
14. Rambaut A, Posada D, Crandall KA, Holmes EC (2004) The causes and consequences of HIV evolution. *Nat Rev Genet* 5:52–61.
15. Ou C-Y, et al. (1992) Molecular epidemiology of HIV transmission in a dental practice. *Science* 256:1165–1171.
16. Hillis DM, Huelsenbeck JP (1994) Support for dental HIV transmission. *Nature* 369: 24–25.
17. Blanchard A, Ferris S, Chamaret S, Guétard D, Montagnier L (1998) Molecular evidence for nosocomial transmission of human immunodeficiency virus from a surgeon to one of his patients. *J Virol* 72:4537–4540.
18. Goujon CP, et al. (2000) Phylogenetic analyses indicate an atypical nurse-to-patient transmission of human immunodeficiency virus type 1. *J Virol* 74:2525–2532.
19. Jaffe HW, et al. (1994) Lack of HIV transmission in the practice of a dentist with AIDS. *Ann Intern Med* 121:855–859.
20. Holmes EC, Zhang LQ, Simmonds P, Rogers AS, Brown AJL (1993) Molecular investigation of human immunodeficiency virus (HIV) infection in a patient of an HIV-infected surgeon. *J Infect Dis* 167:1411–1414.
21. Arnold C, Balfe P, Clewley JP (1995) Sequence distances between *env* genes of HIV-1 from individuals infected from the same source: Implications for the investigation of possible transmission events. *Virology* 211:198–203.
22. Albert J, Wahlberg J, Leitner T, Escanilla D, Uhlén M (1994) Analysis of a rape case by direct sequencing of the human immunodeficiency virus type 1 *pol* and *gag* genes. *J Virol* 68:5918–5924.
23. Leitner T, Albert J (2000) Reconstruction of HIV-1 transmission chains for forensic purposes. *AIDS Rev* 2:241–251.
24. Birch CJ, et al. (2000) Molecular analysis of human immunodeficiency virus strains associated with a case of criminal transmission of the virus. *J Infect Dis* 182:941–944.
25. Kaye M, Chibo D, Birch C (2009) Comparison of Bayesian and maximum-likelihood phylogenetic approaches in two legal cases involving accusations of transmission of HIV. *AIDS Res Hum Retroviruses* 25:741–748.
26. Lemey P, et al. (2005) Molecular testing of multiple HIV-1 transmissions in a criminal case. *AIDS* 19:1649–1658.
27. Machuca R, Jørgensen LB, Theilade P, Nielsen C (2001) Molecular investigation of transmission of human immunodeficiency virus type 1 in a criminal case. *Clin Diagn Lab Immunol* 8:884–890.
28. Banaschak S, Werwein M, Brinkmann B, Hauber I (2000) Human immunodeficiency virus type 1 infection after sexual abuse: Value of nucleic acid sequence analysis in identifying the offender. *Clin Infect Dis* 31:1098–1100.
29. Bird SM, Brown AJL (2001) Criminalisation of HIV transmission: Implications for public health in Scotland. *BMJ* 323:1174–1177.
30. Metzker ML, et al. (2002) Molecular evidence of HIV-1 transmission in a criminal case. *Proc Natl Acad Sci USA* 99:14292–14297.
31. Shattock RJ, Moore JP (2003) Inhibiting sexual transmission of HIV-1 infection. *Nat Rev Microbiol* 1:25–34.
32. Keele BF, et al. (2008) Identification and characterization of transmitted and early founder virus envelopes in primary HIV-1 infection. *Proc Natl Acad Sci USA* 105: 7552–7557.
33. Haaland RE, et al. (2009) Inflammatory genital infections mitigate a severe genetic bottleneck in heterosexual transmission of subtype A and C HIV-1. *PLoS Pathog* 5: e1000274.
34. Abrahams MR, et al.; CAPRISA Acute Infection Study Team; Center for HIV-AIDS Vaccine Immunology Consortium (2009) Quantitating the multiplicity of infection with human immunodeficiency virus type 1 subtype C reveals a non-poisson distribution of transmitted variants. *J Virol* 83:3556–3567.
35. Lemmon AR, Brown JM, Stanger-Hall K, Lemmon EM (2009) The effect of ambiguous data on phylogenetic estimates obtained by maximum likelihood and Bayesian inference. *Syst Biol* 58:130–145.
36. Pearshouse R (2007) Legislation contagion: The spread of problematic new HIV laws in Western Africa. *HIV AIDS Policy Law Rev* 12:1, 5–11.
37. Jürgens R, et al. (2009) Ten reasons to oppose the criminalization of HIV exposure or transmission. *Reprod Health Matters* 17:163–172.
38. UNAIDS (2008) Criminalization of HIV transmission: Policy briefAvailable at http://data.unaids.org/pub/Manual/2008/jc1601_policy_brief_criminalization_long_en.pdf. Accessed November 1, 2010.
39. Family Health International (2008) AWARE-HIV/AIDS, 2003-2008 Closeout report. Available at http://www.fhi.org/en/CountryProfiles/WestAfrica/index.htm. Accessed November 1, 2010.
40. Lowbury R, Kinghorn GR (2006) Criminal prosecution for HIV transmission. *BMJ* 333: 666–667.
41. Pillay D, Rambaut A, Geretti AM, Brown AJL (2007) HIV phylogenetics. *BMJ* 335: 460–461.
42. Burris S, Cameron E (2008) The case against criminalization of HIV transmission. *JAMA* 300:578–581.
43. Lamba Legal (2010) State criminal statutes on HIV exposure. Available at http://www.lambdalegal.org/our-work/publications/general/state-criminal-statutes-hiv.html. Accessed November 1, 2010.
44. Metzker ML, Ansari-Lari MA, Liu X-M, Holder DJ, Gibbs RA (1998) Quantitation of mixed-base populations of HIV-1 variants by automated DNA sequencing with BODIPY dye-labeled primers. *Biotechniques* 25:446–447, 450–452, 454 passim.
45. Metzker ML (2005) Emerging technologies in DNA sequencing. *Genome Res* 15: 1767–1776.
46. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215:403–410.
47. Sullivan J, Joyce P (2005) Model selection in phylogenetics. *Annu Rev Ecol Evol Syst* 36: 445–466.
48. Swofford DL, Olsen GJ, Waddell PJ, Hillis DM (1996) *Molecular Systematics*, eds Hillis DM, Moritz C, Mable BK (Sinauer Associates, Inc., Sunderland, MA), pp 407–514.
49. Brown JM, Lemmon AR (2007) The importance of data partitioning and the utility of Bayes factors in Bayesian phylogenetics. *Syst Biol* 56:643–655.