Molecular evidence of HIV-1 transmission in a criminal case

Michael L. Metzker*†, David P. Mindell‡, Xiao-Mei Liu*§, Roger G. Ptak¶‖, Richard A. Gibbs*, and David M. Hillis**

*Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030; ‡Department of Ecology and Evolutionary Biology and Museum of Zoology, University of Michigan, Ann Arbor, MI 48109-1079; ¶School of Dentistry, Biologic and Materials Sciences, University of Michigan, Ann Arbor, MI 48109; and **Section of Integrative Biology and Center for Computational Biology and Bioinformatics, University of Texas, Austin, TX 78712

Edited by Walter M. Fitch, University of California, Irvine, CA, and approved September 4, 2002 (received for review May 2, 2002)

A gastroenterologist was convicted of attempted second-degree murder by injecting his former girlfriend with blood or blood-products obtained from an HIV type 1 (HIV-1)-infected patient under his care. Phylogenetic analyses of HIV-1 sequences were admitted and used as evidence in this case, representing the first use of phylogenetic analyses in a criminal court case in the United States. Phylogenetic analyses of HIV-1 reverse transcriptase and *env* DNA sequences isolated from the victim, the patient, and a local population sample of HIV-1-positive individuals showed the victim's HIV-1 sequences to be most closely related to and nested within a lineage comprised of the patient's HIV-1 sequences. This finding of paraphyly for the patient's sequences was consistent with the direction of transmission from the patient to the victim. Analysis of the victim's viral reverse transcriptase sequences revealed genotypes consistent with known mutations that confer resistance to AZT, similar to those genotypes found in the patient. *A priori* establishment of the patient and victim as a suspected transmission pair provided a clear hypothesis for phylogenetic testing. All phylogenetic models and both genes examined strongly supported the close relationship between the HIV-1 sequences of the patient and the victim. Resampling of blood from the suspected transmission pair and independent sequencing by different laboratories provided precaution against laboratory error.

In recent years, DNA testing has been widely used in the judicial system, mainly in violent crimes to link a perpetrator to the scene of the crime. Human DNA is generally stable, allowing the techniques of DNA fingerprinting to be used in analyzing multiple polymorphic markers for the purpose of excluding suspected individuals. Assessing phylogeny for HIV type 1 (HIV-1) strains, however, is more complex than human DNA testing because of the dynamic nature and rapid rates of HIV-1 change (1–4). However, this high rate of change among HIV sequences permits an application of phylogenetic methods, and several case studies have been described that investigated the relatedness of HIV-1 strains for the purpose of examining suspected viral transmission events. Probably the most well-known and scrutinized study is the "Florida dentist" case, which concluded that six patients became infected with HIV-1 while receiving care from an HIV-1-positive dentist (5–9). Other studies that have supported suspected transmissions of HIV-1 between individuals are the "Swedish rape case" (10) and the "French orthopedic surgeon" case (11). In addition, one published study rejected a hypothesis of transmission between a Baltimore surgeon and one of his patients (12).

Because of the rapid rate of evolution of HIV-1, phylogenetic analysis of HIV-1 DNA sequences is a powerful tool for the identification of closely related viral strains that may be used to infer the transmission between individuals. In the case of the State of Louisiana vs. Richard J. Schmidt, the prosecution argued successfully that the methods of genomic DNA isolation, PCR, DNA sequencing, and phylogenetic analysis of HIV-1 DNA sequences to characterize HIV-1-positive samples identified by criminal investigation met the judicial standards of evidence admissibility. These standards include the facts that the methods are subject to empirical testing, are subject to peer review and publication, can be assessed for error, and are generally accepted in the scientific community (13). This case was the first time that phylogenetic analysis has been used as evidence in a United States criminal proceeding. Here we present the phylogenetic evidence that constituted part of the prosecution's case that resulted in the conviction of the Louisiana gastroenterologist on the charge of attempted second-degree murder.

Materials and Methods

Criminal Investigation. The prosecution's case was based on circumstantial evidence indicating that on August 4, 1994, a Lafayette, LA, gastroenterologist made a mixture of blood or blood-products from two patients under the doctor's care, one infected with HIV-1 and the other with hepatitis C, and infected his former girlfriend by intramuscular injection. Our efforts for the criminal investigation involved only the molecular analysis of HIV-1 sequences, which represented only one part of the prosecution's case against the physician.

Risk factors associated with HIV-1 infection for the victim were determined through the course of the criminal investigation. From 1984 to 1995, the victim reported having sexual contacts with seven men, including the doctor, all of whom were interviewed by local law enforcement agents. The seven men were tested between the years of 1995 and 1998, and were found to be negative for HIV-1 (Keith A. Stutes, Assistant District Attorney, 15th Judicial District, Lafayette, LA, personal communication). The victim was a nurse from the Lafayette area who had no documented reports of needle sticks, but did report in the mid-1980s that an HIV-1-infected patient splashed saliva onto her skin while spitting into a pan. The victim was tested after the incident and was found to be negative for HIV-1. Moreover, the victim had donated blood to a local blood bank on several occasions, for which she was tested and found negative for HIV-1 in October 1992, May 1993, and April 1994 (Keith A. Stutes, personal communication).

In January 1995, however, the victim tested positive for HIV-1, and subsequently accused the physician of infecting her with HIV-1 from an intramuscular injection during an argument in early August 1994. In the subsequent investigation, law enforcement agents identified an HIV-1-infected patient whose blood had been drawn in the doctor's office on August 4, 1994. The patient's name was one of the last recorded entries into a reportedly missing log notebook that was discovered after a search warrant had been obtained to search the physician's office. This patient's blood draw was recorded in a different manner from the usual operating procedures of the

This paper was submitted directly (Track II) to the PNAS office.

Abbreviations: HIV-1, HIV type 1; RT, reverse transcriptase; BCM, Baylor College of Medicine; MIC, University of Michigan; PBMC, peripheral blood mononuclear cell; AZT, 3′-azido-3′-deoxythymidine; BP, bootstrap proportions; MCMC, Markov-chain Monte Carlo.

Data deposition: The sequences reported in this paper have been deposited in the GenBank database (accession nos. AY156734–AY156907).

†To whom correspondence should be addressed. E-mail: mmetzker@bcm.tmc.edu.

§Present address: Department of Virus and Cell Biology, Merck Research Laboratories, West Point, PA 19486.

‖Present address: Southern Research Institute, Frederick, MD 21701-4756.

doctor's office. The patient was a homosexual male who was infected with HIV-1 in 1990. Thus, the evidence gathered formed the basis of the *a priori* hypothesis of suspected transmission of HIV-1 from the patient to the victim and justified the use of phylogenetic analysis to investigate the relatedness of HIV-1 DNA sequences for the suspected transmission pair.

Sample Handling, PCR, and DNA Sequencing. The experimental data were derived from two independent laboratories, first performed at Baylor College of Medicine (BCM) and confirmed at the University of Michigan (MIC). On September 28, 1995, a Lafayette detective delivered whole blood drawn from the patient and the victim to BCM for DNA analysis, and LA control blood samples (see below) were delivered to BCM between March 1996 and June 1996. Eighteen months later, the first anonymously labeled blood sample (BS1) was delivered on March 26, 1997 to the Michigan group. After sequences had been obtained from BS1, the BS2 sample was delivered on April 25, 1997. The strategy of independent testing in two different laboratories from two separate blood draws each for the patient and victim allowed us to exclude the possibility that the relatedness of HIV-1 DNA sequences derived from the patient and the victim were the result of laboratory error, sample mix-up, or cross-contamination. Moreover, we designed the second study to remove the potential for investigator bias by conducting experiments in a blinded fashion, in which the identities of the samples were not revealed to the Michigan group until the analyses of both samples were completed. In addition, BS1 and BS2 samples were analyzed at different times, separated by thorough cleaning of all laboratory equipment and surfaces. In all instances, peripheral blood mononuclear cells (PBMC) isolation from whole blood, PCR setup, genomic DNA addition, and DNA sequencing were each performed in physically separated laboratory areas, where reagents were premixed and aliquoted with dedicated positive displacement pipettes to minimize the risk of contamination (15, 16).

Genomic DNA from each individual was used to PCR-amplify a 858-bp *env* gene fragment and a 1,147-bp reverse transcriptase (RT) gene fragment as described (17, 18). Comparative phylogenetic analyses are important in that they strengthen the interpretation of relatedness between the key individuals because these genes exhibit different biological functions, are targeted by different selective pressures (i.e., host versus drug, respectively), and are known to undergo different rates of evolution. Briefly, ≈200 ng of genomic DNA from the patient, victim, or any LA controls were handled and amplified separately by coamplifying the *pol* and *env* genes regions with PCR1/PCR2 and PCR5/PCR6 primer pairs to yield 1,231- and 1,288-bp fragments. A "hot start" technique was used for all PCRs to increase product specificity by using AmpliWax beads according to the manufacturer's protocol. A second round of hot start PCR was performed to amplify separately a 1,147-bp *pol* fragment by using PCR3B/PCR4B primer pairs and a 858-bp *env* fragment by using PCR7B/PCR8B primer pairs. PCR products from the patient, victim, and the LA controls were directly sequenced. Seven and two different PCR products from the RT gene were directly sequenced for the patient (P1–P7) and victim (V1 and V2) samples, respectively. To further delineate the extent of *env* genetic diversity of the suspected transmission pair, 50 molecular clones were isolated and sequenced at BCM from their respective PCR products. The molecular clones for the patient and victim were designated P01 through P50 and V01 through V50.

Ideally, the within-host sampling size should be representative of the diversity of the gene under investigation, which becomes difficult to standardize because genetic diversity depends on several factors including unknown host factors, stage of the disease, specific combinations of HIV-1 drug therapies, and their effectiveness (i.e., emergence of drug-resistant strains). The selection of 50 molecular clones represented a reasonable compromise for sampling diversity of the patient and victim HIV-1 isolates and represented greater than three times as many clones analyzed as other studies (5, 10–12). For example, molecular clone sequences P04, P16, P22, P32, and P41 were identical to each other as were sequences P06, P17, P19, P31, and P44, which suggests oversampling by the identification of redundant sequences. On the other hand, all molecular clones obtained from the victim represented unique sequences.

Independently, the Michigan group obtained several molecular clones for both HIV-1 gene fragments from separate blood draws each for the patient and victim. Custom dipyrromethene boron difluoride (BODIPY) dye primers (BCM) or dye terminator (MIC) chemistries were used to sequence both strands for an internal 784-bp *env* fragment and an internal 689-bp RT fragment by using a Perkin–Elmer Applied Biosystems 377 DNA sequencer. DNA sequencing reads were then aligned, and any ambiguities were manually edited when needed to obtain an accurate consensus sequence from each molecular clone or PCR product (17).

To safeguard against laboratory contamination of previously characterized HIV-1 samples produced at BCM, the patient and victim HIV-1 DNA sequences were compared by phylogenetic analyses to all available BCM HIV-1 DNA sequences. Consensus sequences for the patient and victim were found to be closely related to each other and not to any previously characterized HIV-1 DNA sequences (data not shown).

HIV-1 DNA Controls. We reasoned that the most appropriate controls for the current study were HIV-1-infected individuals from the Lafayette area. Sampling from the same metropolitan area where the two key individuals reside can provide the greatest opportunity for identifying an alternative HIV-1 transmission between these individuals. The use of local sequences also allowed us to consider the possibility that HIV-1 DNA sequences derived from the suspected transmission pair were representative of an HIV-1 strain ubiquitous in the Lafayette metropolitan area. This rationale is supported by molecular studies that have shown global geographic subtype stratification of HIV-1 sequences (19). We did not consider race, gender, or risk factors for the selection of the local controls because of a lack of evidence suggesting a correlation between these factors and specific HIV-1 DNA sequences.

An ideal among host sampling size should be obtained from a reasonable percentage of the local population, which is representative of the total number of HIV-1 infected individuals. Several issues, including knowing the total number of infected individuals and the willingness of individuals to participate in such studies, however, confound an accurate determination of this number for any analysis. For this study, 32 blood samples were obtained from available HIV-1-infected individuals from the local metropolitan area. PBMCs from whole blood for controls LA03 and LA11, however, could not be isolated successfully. Sequence analysis of direct DNA sequence data from the *env* gene revealed laboratory contamination of pNL4-3 plasmid sequences for LA01 and LA19 controls (20); these sequences were removed from further analysis.

Of the LA controls, 28 samples gave positive HIV-1 DNA sequences that were unique relative to published database sequences, though not all gave PCR products for both gene regions. For example, PCR products for LA21 and LA23 for the *env* region and LA09, LA15, and LA20 for the RT region could not be obtained and were excluded from their respective analyses. The control group represented a diverse spectrum of infected individuals with respect to date of infection, $CD4^+$ cell count, and clinical status (Table 1). To characterize the relevance of the gp120 LA controls, molecular clone sequences derived from the patient and victim were compared with HIV-1 DNA sequences in GenBank by using BLAST (21). In all cases, multiple clones from the same published reports gave significant BLAST scores. To represent the broadest spectrum of related public HIV-1 sequences from independent publications, only the highest scoring clone sequences were selected: GenBank accession numbers AF025750, U95403, M21098, M79352, U63632, M80663, U84880, D12582, L14574,

Table 1. Summary of LA control group sample sources

Risk factors			
Homosexual	57%	(16/28)	
Heterosexual	18%	(5/28)	
Blood transfusion	11%	(3/28)	
Bi-sexual	7%	(2/28)	
IV drug user	4%	(1/28)	
Sharps	4%	(1/28)	
Date infected			
1983–1989	57%	(16/28)	
1990–1992	36%	(10/28)	
Unknown	7%	(2/28)	
$CD4^+$ cell counts			AIDS
>500	36%	(10/28)	(2/10)
200–500	25%	(7/28)	(3/7)
<200	25%	(7/28)	(6/7)
ND	14%	(4/28)	(4/4)

Risk factors and dates infected were obtained by anonymous questionnaire. ND, not determined.

M95292, M38429, U69584, U79719, U43096, M17449, Y13716, U16778, L08655, and M63929. The pNL4-3 plasmid sequence (M19921), which did not meet the requirements as stated above, was also included because of its usage at BCM.

Phylogenetic Analyses. We used phylogenetic analyses of the gp120 and RT sequences to examine relationships among the patient, victim, and LA control viral DNA sequences. The analyses that formed the basis of the results we presented in court were conducted by using the optimality criteria of parsimony and minimum evolution using maximum-likelihood distances (22). These approaches were used because they were accepted by the court in a pretrial hearing as meeting the criteria for admissibility of evidence. Analyses based on direct likelihood evaluations of the sequence data were not computationally feasible at the time of the pretrial hearing or court case. Recent developments of Markov-chain Monte Carlo (MCMC) approaches have, however, made a Bayesian analysis under a likelihood model feasible. Therefore, we conducted additional posttrial analyses with MCMC Bayesian analysis (23, 24) by using the Metropolis-coupled MCMC algorithm implemented in the program MRBAYES (25). Our Bayesian analysis was based on a General-Time-Reversible model of sequence evolution, with γ-distributed rate heterogeneity among sites and a calculated proportion of invariable sites (GTR+Γ+I; ref. 22). This model of sequence evolution was chosen based on results from the program MODELTEST (version 3.06; ref. 26). For each gene, we ran the MCMC searches for 5,000,000 generations, and sampled solutions once every 100 generations. After 2,500,000 generations, we determined that the searches had reached equilibrium by plotting the values for the likelihood scores and the various parameters of the model. We therefore used the samples from the final 2,500,000 generations to compute 95% confidence intervals for the model parameters shown in Tables 2 and 3, and to assess the posterior probabilities of the relationships between the victim and patient sequences.

For the parsimony and minimum evolution analyses, we searched for optimal trees with PAUP (versions 4.0b4–10; ref. 28), and used tree-bisection-reconnection branch-swapping from random-order stepwise addition of starting trees for the parsimony analyses and neighbor-joining for distance analyses. Our minimum evolution analyses conducted for the trial were based on eight different maximum-likelihood distances, including a range of models from JC to HKY+Γ (28–31). For each of four basic models (JC, F81, K2P, and HKY), we calculated distances with and without a γ correction for rate heterogeneity among sites. All analyses (parsimony, minimum evolution under all distances, and Bayesian) produced highly congruent results that did not differ with respect to the relevant relationships between the victim and patient sequences (see *Phylogenetic Results*).

Table 2. Means and 95% confidence intervals for parameters of the GTR + Γ + I model for gp120 sequences

Parameter	Mean	95% Confidence interval
C–T substitution rate	5.03	3.60–7.03
C–G substitution rate	0.97	0.57–1.54
A–T substitution rate	0.75	0.52–1.07
A–G substitution rate	3.87	2.91–5.10
A–C substitution rate	2.34	1.60–3.34
Frequency of A	0.40	0.37–0.43
Frequency of C	0.15	0.13–0.17
Frequency of G	0.23	0.21–0.25
Frequency of T	0.22	0.20–0.25
α (shape of Γ distribution)	0.53	0.43–0.68
Proportion of invariable sites	0.08	0.01–0.18

Data based on MCMC sampling (25). The rate of all substitution classes is shown relative to that of the G–T substitution class.

For the parsimony and minimum evolution analyses, nonparametric bootstrapping (32) was used to test the *a priori* hypothesis of a relationship between the victim and patient sequences. The generally accepted standard for rejecting a null hypothesis (in this case, the null hypothesis is that the sequences obtained from the victim are not most closely related to sequences obtained from the patient) is $P < 0.05$. In forensic studies, however, there is no widely accepted standard for the meaning of *beyond a reasonable doubt*. Under a wide range of conditions, bootstrap proportions (BP) have been shown to represent a conservative estimate of phylogenetic confidence (33–35), and 1-BP was used as a conservative estimate of p (the probability of type I error) in a test of the *a priori* hypothesis (36). Because of the importance of estimating the strength of the results, we conducted as many bootstrap replications as were computationally feasible for each analysis. For parsimony analyses, we examined 100,000 bootstrap replicates, whereas for the more computationally intense maximum-likelihood distance analyses (in which large numbers of pairwise distances had to be recalculated for each replicate), we conducted from 1,000 (gp120) to 10,000 (RT) replicates.

Bayesian posterior probabilities, calculated from a consensus analysis of the samples of solutions from the MCMC searches of tree and parameter space, have been shown to be less biased measures of phylogenetic accuracy compared with nonparametric bootstrapping proportions (37). Although this approach was not available for use at the trial, it represents the current standard of practice for assessing the reliability of phylogenetic analyses, especially under a likelihood model. Therefore, we used our samples of 25,000 solutions (one solution sampled every 100 generations from

Table 3. Means and 95% confidence intervals for parameters of the GTR + Γ + I model for the RT sequences

Parameter	Mean	95% Confidence interval
C–T substitution rate	110.36	23.04–195.53
C–G substitution rate	17.59	2.82–42.02
A–T substitution rate	7.62	1.34–17.32
A–G substitution rate	83.01	16.29–171.17
A–C substitution rate	16.60	3.41–35.62
Frequency of A	0.40	0.36–0.43
Frequency of C	0.17	0.14–0.19
Frequency of G	0.20	0.17–0.23
Frequency of T	0.23	0.20–0.26
α (shape of Γ distribution)	0.94	0.38–1.94
Proportion of invariable sites	0.50	0.29–0.63

Data are based on MCMC sampling (25). The rate of all substitution classes is shown relative to the rate of the G–T substitution class.

Fig. 1. Phylogenetic analysis of the gp120 region using a minimum evolution criterion and maximum likelihood distances assuming an HKY+Γ model of evolution. Nucleotide alignment was based on the protein alignment in Fig. 3. P.ENV and V.ENV are DNA sequences for provirus PCR products from the patient and victim, respectively. Sequence names beginning with LA denote viral sequences from control HIV-1 infected individuals from the Lafayette, LA, metropolitan area. The same pattern of relationships (monophyly of all patient and victim sequences) was obtained with all phylogenetic methods (parsimony, minimum evolution, and Bayesian) and all models of evolution examined. In addition to the 100% bootstrap support of this relationship for the minimum evolution analyses, the parsimony bootstrap support and the Bayesian posterior support were also 100%.

the last 2,500,000 generations of the MCMC analysis) to compute posterior probabilities for the relationship between the victim and patient sequences.

Results

Sequences Examined. The gp120 gene was initially characterized by DNA sequence analysis because of its high genetic variability and its important role in governing antibody neutralization, cellular tropism, and cytopathogenicity. Direct DNA sequencing data revealed highly heterogeneous viral sequences for both the patient and the victim samples, particularly within the V4 and V5 domains (Fig. 3, which is published as supporting information on the PNAS web site, www.pnas.org). The ranges of intrahost divergence among the 50 molecular clones for the patient and victim were 0.00–5.57% and 0.13–2.77%, respectively. The ranges of intrahost divergence for the two molecular clone sequences characterized by the Michigan group compared with the BCM set were 0.13–5.18% and 0.64–2.14%, respectively, which were obtained from separate blood draws each for the patient and victim ≈18 months after the BCM draw. Despite the large number of viral strains present in blood and the advanced progression of disease for the latter blood draw, sampling of viral isolates by the different laboratories revealed highly similar results with the highest identity between the BCM and MIC sequences being 99.87% and 99.36% for the patient and victim samples, respectively. These data confirmed and validated the identities of the suspected transmission pair, because they fall within the range of variation of the BCM data sets.

We investigated a second genetic locus by direct DNA sequencing (BCM) and cloning and sequencing (MIC) analyses of the amino-terminal half of the RT gene (Fig. 4, which is published as supporting information on the PNAS web site). The range of intrahost divergence was 0.00–2.03% and 0.58–2.32% for the patient BCM and MIC sequences, respectively, and they differed from each other by 0.15–2.47%. The victim RT samples also showed less diversity between sequences: 0.07% (BCM), 0.15% (MIC), and 0.07–0.65% (intergroup divergence). Analyses of the patient sequences revealed two distinct populations, with one group showing amino acid mutations known to confer resistance to AZT (39–41).

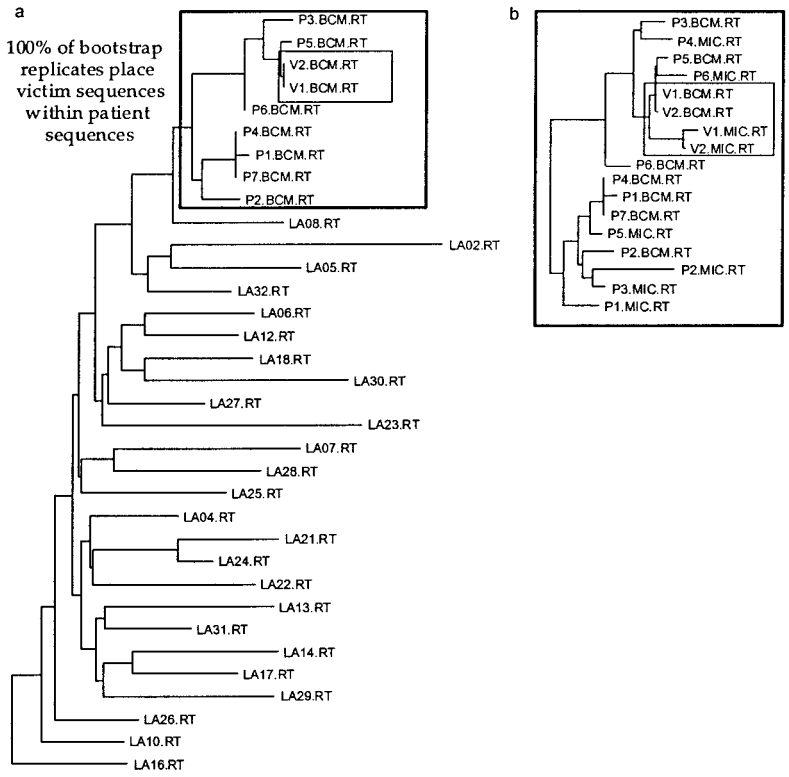

Fig. 2. Phylogenetic analysis of the RT region; details of the analysis are the same as for Fig. 1. Nucleotide alignment was based on the protein alignment in Fig. 4. (*a*) Tree based on sequences from BCM. (*b*) Subtree of patient and victim sequences, including those added by MIC. In both *a* and *b*, the smaller set of boxed sequences represents the sequences from the victim, and the larger set of boxed sequences represents the patient plus victim sequences. The victim sequences were found to be embedded within the patient sequences in all analyses and for all models of evolution examined. In addition to the 100% bootstrap support of this relationship for the minimum evolution analyses, the parsimony bootstrap support was 96% and the Bayesian posterior support was 100%.

The timing and accumulation of the different drug resistance mutations have been shown to be variable between HIV-1-infected individuals (42, 43). Stratifying the patient sequences based on AZT resistance (AZT^R), the range of divergence for all AZT^R sequences was 0.00–1.45% and for all AZT sensitive (AZT^S) sequences was 0.44–1.23%. The victim's HIV-1 DNA sequences were more similar to the patient's AZT^R sequences than to the patient's AZT^S sequences (the ranges of divergence were 0.15–1.52% compared with 1.60–2.61%, respectively).

Sequence analyses of 20 HIV-1 DNA public sequences that showed the most significant BLAST scores revealed that molecular clone sequences for the patient and victim differed by 8.05–13.54% and 8.54–15.37%, respectively, from the GenBank sequences. The gp120 LA control sequences differed from those of the patient by 7.41–14.22% and from those of the victim by 8.02–15.43%. The random LA controls and the 20 closest HIV-1 DNA sequences selected from GenBank exhibited similar divergence, though the most similar sequences to the victim and patient were found among the LA controls. These data suggest that the selection of control sequences from the local geographic area were appropriate for this study. Comparing RT sequences, 57% of the LA controls showed various amino acid substitutions known to confer resistance to AZT. These LA control sequences were more divergent from the stratified patient AZT^R, patient AZT^S, and the victim sequences and differed by 2.61–6.75%, 3.19–6.53%, and 3.34–6.53%, respectively.

Phylogenetic Results. In the parsimony analyses, all 100,000 bootstrap replicates of the gp120 gene data supported the victim and patient sequences as the most closely related within the analysis ($P < 0.00001$), and 95,826 bootstrap replicates of the RT gene data supported the victim sequences as embedded within a group of patient sequences ($P < 0.04174$). In the maximum-likelihood distance analyses, all 1,000 bootstrap replicates of the gp120 gene data ($P < 0.001$; Fig. 1) supported the closer relationship between the patient and victim viral sequences compared with any of the LA controls, and all 10,000 bootstrap replicates of the RT gene data ($P < 0.0001$; Fig. 2*a*) supported the victim sequences as embedded within a group of patient sequences. All 25,000 sampled trees from the MCMC analyses also supported these relationships ($P < 0.00004$). The relationships of the patient and victim RT sequences were virtually identical based on both the originally sampled sequences (sequenced at BCM) and those subsequently sequenced at MIC (Fig. 2*b*). The close relationship between the victim and patient samples was thus supported by both of the genes that we examined, using all major methods of phylogenetic analysis (parsimony, minimum evolution, and likelihood), and a broad range of evolutionary models.

Discussion

Direction of Transmission. Although the inferred sister relationship between patient and victim viruses is consistent with the alleged transmission event, this finding by itself does not establish the direction of the transmission nor does it prove that additional individuals could not have been involved in a series of intermediate transmissions. However, if the sequences are sampled close enough in time to the transmission event, the direction between a suspected pair can often be established (44). Typically only a single or a few viral isolate(s) have been shown to be transmitted during primary infection (38, 45, 46), and if samples are obtained shortly after this event, a subset of source sequences will be found to be more closely related to the recipient sequences than all source sequences are to each other. Thus, source sequences that are paraphyletic with respect to the recipient sequences provide evidence for the direction of transmission. This paraphyletic relationship will be lost through time as a result of lineage extinction, but can be observed between transmission pairs that are sampled within a short period of the transmission event. The window of opportunity for observing this paraphyletic relationship is expected to vary as a function of rate of evolution of the various parts of the genome and degree of immunoselection and/or drug selection for the different gene proteins.

All of the parsimony, minimum evolution, and Bayesian analyses of the RT sequences showed that patient sequences were paraphyletic with respect to the victim sequences. That is, virus sequences from the victim were nested within the larger clade of sequences from the patient. This finding was supported in 95,826 parsimony bootstrap replicates, in all 10,000 bootstrap replicates in the maximum-likelihood distance analysis (Fig. 2*a*), and in all 25,000 sampled trees from the MCMC Bayesian analysis. The gp120 sequences, however, showed a weak monophyletic grouping of the patient sequences, which was supported by 51–53% of bootstrap replicates in parsimony and minimum evolution analyses (the remaining 47–49% of the replicates support paraphyly of the patient sequences). This observed discrepancy is likely due to differences in evolutionary rates of the RT and gp120 genes, and the strong immunoselection on the latter. Faster rates of sequence evolution and subsequent shorter coalescence times for gp120 gene trees, which results in apparent monophyly for the patient sequences, are expected. The paraphyletic relationship of the suspected pair for the RT gene was reproduced by the Michigan group (Fig. 2*b*). Thus, the data presented here, which were obtained from two independent blood draw samplings from the suspected transmission pair and analyzed by independent laboratories, provide strong evidence that the direction of transmission was from the patient to the victim and that transmitted HIV-1 lineage(s) were of AZT-resistant genotypes.

The introduction of phylogenetic analysis of HIV-1 DNA sequences into the United States court system inevitably will result in greater numbers of forensic investigations of suspected HIV-1 transmission cases. Moreover, these studies have broad applications for the identification of putative sources of existing and new pathogens that can cause food-borne infections and hazardous agents that can be used for the purpose of biological warfare.

The verdict of State of Louisiana vs. Richard J. Schmidt was appealed to the Louisiana State Supreme Court, where it was upheld in 2000. On March 4, 2002, the United States Supreme Court also rejected an appeal of the case, thus ending the judicial proceedings. Precedent for the use of phylogenetic analysis to support or reject criminal viral transmission cases has thus been established in United States courts of law. It is ironic that this case originated in Louisiana, which enacted the Balanced Treatment for Creation-Science and Evolution-Science Act in 1982. In 1987, the United States Supreme Court found this act unconstitutional. The increasing role of scientific methods and hypothesis testing within the legal system challenges scientists to uphold the highest possible levels of rigor and objectivity. Guidelines similar to those established for forensic testing of human DNA polymorphisms (14) should be established. Although fallible, the self-correcting nature of scientific observation and interpretation and the amenability of scientific methods to repeated testing is well suited for not only understanding the world in which we live, but also for the illumination of historical events, including acts against society.

We thank Arthur Young, Detective Jim Craft, District Attorney Michael Harson, and Assistant District Attorney Keith Stutes for their work on this case. We also thank Dorothy Lewis for providing the $CD4^+$ cell count measurements and John Drach for assistance in the sequencing work conducted at the University of Michigan. During the course of the research, M.L.M. was supported in part by National Institute of Allergy and Infectious Diseases Basic Research Grant AI07483 on HIV and AIDS, and D.P.M. (DEB-9726427 and DBI-9974525) and D.M.H. (EIA-0121680) were supported by grants from the National Science Foundation.

1. Wei, X., Ghosh, S. K., Taylor, M. E., Johnson, V. A., Emini, E. A., Deutsch, P., Lifson, J. D., Bonhoeffer, S., Nowak, M. A., Hahn, B. H., *et al.* (1995) *Nature* **373,** 117–122.
2. Ho, D. D., Neumann, A. U., Perelson, A. S., Chen, W., Leonard, J. M. & Markowitz, M. (1995) *Nature* **373,** 123–126.
3. Coffin, J. M. (1995) *Science* **267,** 483–489.
4. Perelson, A. S., Neumann, A. U., Markowitz, M., Leonard, J. M. & Ho, D. D. (1996) *Science* **271,** 1582–1586.
5. Ou, C.-Y., Ciesielski, C. A., Myers, G., Bandea, C. I., Luo, C. C., Korber, B. T., Mullins, J. I., Schochetman, G., Berkelman, R. L., Economou, A. N., *et al.* (1992) *Science* **256,** 1165–1171.
6. Smith, T. F. & Waterman, M. S. (1992) *Science* **256,** 1155–1156.
7. DeBry, R. W., Abele, L. G., Weiss, S. H., Hill, M. D., Bouzas, M., Lorenzo, E., Graebnitz, F. & Resnick, L. (1993) *Nature* **361,** 691.
8. Holmes, E. C., Leigh Brown, A. J. & Simmond, P. (1993) *Nature* **364,** 766.
9. Hillis, D. M. & Huelsenbeck, J. P. (1994) *Nature* **9,** 24–25.
10. Albert, J., Wahlberg, J., Leitner, T., Escanilla, D. & Uhlén, M. (1994) *J. Virol.* **68,** 5918–5924.
11. Blanchard, A., Ferris, S., Chamaret, S., Guétard, D. & Montagnier, L. (1998) *J. Virol.* **72,** 4537–4540.
12. Holmes, E. C., Zhang, L. Q., Simmonds, P., Rogers, A. S. & Leigh Brown, A. J. (1993) *J. Infect. Dis.* **167,** 1411–1414.
13. State of Louisiana vs. Richard J. Schmidt, 15th Judicial District Court, Lafayette Parish, LA, Criminal Docket No. 73313, *Reasons For Ruling* of Louisiana State 15th Judicial District Court Judge Durwood Conque (1997); State of Louisiana vs. Richard J. Schmidt, 699 So. 2d 488, K97–249 LA Court of Appeal, 3rd Circuit (1997); writ denied 706 So. 2d 451, 97–2220 LA (1997).
14. Crow, J. F, Berger, M. A., Diamond, S. S., Kaye, D. H., Kazazian, H. H., Motulsky, A. G., Nakylaki, T. A., Nei, M., Sensbaugh, G. F., Siegmund, D. O. & Stigler, S. M. (1996) *The Evaluation of Forensic DNA Evidence* (Natl. Acad. Press, Washington, DC).
15. Gibbs, R. A. & Chamberlain, J. S. (1989) *Genes Dev.* **3,** 1095–1098.
16. Kwok, S. & Higuchi, R. (1989) *Nature* **339,** 237–238.
17. Metzker, M. L., Ansari-Lari, M. A., Liu, X.-L., Holder, D. J. & Gibbs, R. A. (1998) *BioTechniques* **25,** 446–462.
18. Metzker, M. L., Lu, J. & Gibbs, R. A. (1996) *Science* **271,** 1420–1422.
19. Gao, F., Morrison, S. G., Robertson, D. L., Thornton, C. L., Craig, S., Karlsson, G., Sodroski, J., Morgado, M., Galvao-Castro, B., von Briesen, H., *et al.* (1996) *J. Virol.* **70,** 1651–1667.
20. Korber, B. T. M., Learn, G., Mullins, J. I., Hahn, B. H. & Wolinsky, S. (1995) *Nature* **378,** 242–243.
21. Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997) *Nucleic Acids Res.* **25,** 3389–3402.
22. Swofford, D. L., Olsen, G. J., Waddell, P. J. & Hillis, D. M. (1996) in *Molecular Systematics*, eds. Hillis, D. M., Moritz, C. & Mable, B. K. (Sinauer, Sunderland, MA), 2nd Ed., pp. 407–514.
23. Rannala, B. & Yang, Z. H. (1996) *J. Mol. Evol.* **43,** 304–311.
24. Larget, B. & Simon, D. (1999) *Mol. Biol. Evol.* **16,** 750–759.
25. Huelsenbeck, J. P. & Ronquist, F. R. (2001) *Bioinformatics* **17,** 754–755.
26. Posada, D. & Crandall, K. A. (1998) *Bioinformatics* **14,** 817–818.
27. Swofford, D. L. (1998) PAUP: Phylogenetic Analysis Using Parsimony and Other Methods (Sinauer, Sunderland, MA).
28. Jukes, T. H. & Cantor, C. R. (1969) in *Mammalian Protein Metabolism*, ed. Munro, H. N. (Academic, New York), pp. 21–132.
29. Kimura, M. (1980) *J. Mol. Evol.* **16,** 111–120.
30. Felsenstein, J. (1981) *J. Mol. Evol.* **17,** 368–376.
31. Hasegawa, M., Kishino, H. & Yano, T. (1985) *J. Mol. Evol.* **22,** 160–174.
32. Felsenstein, J. (1985) *Evolution (Lawrence, Kans.)* **39,** 783–791.
33. Zharkikh, A. & Li, W. H. (1992) *Mol. Biol. Evol.* **9,** 1119–1147.
34. Zharkikh, A. & Li, W. H. (1992) *J. Mol. Evol.* **35,** 356–366.
35. Hillis, D. M. & Bull, J. J. (1993) *Syst. Biol.* **42,** 182–192.
36. Felsenstein, J. & Kishino, H. (1993) *Syst. Biol.* **42,** 193–200.
37. Wilcox, T. P., Zwickl, D. J., Heath, T. & Hillis, D. M. (2002) *Mol. Phylogenet. Evol.* **25**.
38. Scarlatti, G. Leitner, T., Halapi, E., Wahlberg, J., Marchisio, P., Clerici-Schoeller, M. A., Wigzell, H., Fenyo, E. M., Albert, J., Uhlen, M. & Rossi, P. (1993) *Proc. Natl. Acad. Sci. USA* **90,** 1721–1725.
39. Larder, B. A., Darby, G. & Richman, D. D. (1989) *Science* **243,** 1731–1734.
40. Larder, B. A. & Kemp, S. D. (1989) *Science* **246,** 1155–1158.
41. Kellam, P., Boucher, C. A. B. & Larder, B. A. (1992) *Proc. Natl. Acad. Sci. USA* **89,** 1934–1938.
42. Boucher, C. A. B., O'Sullivan, E., Mulder, J. W., Ramautarsing, C., Kellam, P., Darby, G., Lange, J. M., Goudsmit, J. & Larder, B. A. (1992) *J. Infect. Dis.* **165,** 105–110.
43. Cleland, A., Watson, H. G., Robertson, P., Ludlam, C. A. & Leigh Brown, A. J. (1996) *J. Acquired Immune Defic. Syndr.* **12,** 6–18.
44. Bowman, B. H. & White, T. J. (1994) in *AIDS Testing*, eds. Schochetman, G. & George, J. R. (Springer, New York), pp. 284–301.
45. Wolinsky, S. M., Wike, C. M., Korber, B. T., Hutto, C., Parks, W. P., Rosenblum, L. L., Kunstman, K. J., Furtado, M. R. & Munoz, J. L. (1992) *Science* **255,** 1134–1137.
46. Wolfs, T. F. W., Zwart, G., Bakker, M. & Goudsmit, J. (1992) *Virology* **189,** 103–110.