

Reverse Transcriptase and Protease Sequence Evolution in Two HIV-1–Infected Couples

*Sarah Palmer, †Dominique Vuitton, ‡Matthew J. Gonzales, †Agnès Bassignot, and
‡Robert W. Shafer

**HIV Drug Resistance Program, National Cancer Institute, National Institutes of Health, Frederick, Maryland, U.S.A.;
†Université de Franche-Comté, Centre Hospitalier Universitaire, Besançon, France; and ‡Division of Infectious Diseases and
Geographic Medicine, Stanford University, Stanford, California, U.S.A.*

Summary: We analyzed the reverse transcriptase (RT) and protease sequences of HIV-1 isolates obtained over 7 years from two couples with known transmission histories. Phylogenetic trees constructed from the sequence data reflected the known transmission histories, despite the fact that the drug resistance mutations were most consistent with the drug treatment histories. However, the RT sequences from one couple diverged by 2.9% even before therapy was begun, and three (0.9%) of 339 unrelated individuals had viruses that shared a common ancestor with sequences from the recipient member of the couple but not with sequences from the transmitter. The divergence between the first two isolates from this couple is consistent with a pre-transmission interval during which the transmitter developed a heterogeneous virus population. The closeness between the three controls and the recipient's first RT sequence may indicate slower evolution on the branches of the control sequences. Although the RT and protease genes contain phylogenetic information, they are sub-optimal for reconstructing transmission history because the genetic distance between RT and protease isolates from unrelated individuals may occasionally approximate the distance between RT and protease isolates from related individuals. **Key Words:** HIV—Antiretroviral resistance—HIV transmission history—AF487122–AF487139.

The error-prone replication of HIV-1 leads to the development of a viral quasispecies within individuals and to virus diversification at the population level as the virus spreads between infected individuals (1). Because the genetic diversity between individuals exceeds the diversity within individuals, it has been possible—using *env* and parts of *gag*—to perform molecular epidemiological studies of HIV transmission (2,3). Indeed, recent epidemiological studies and legal cases concerning the transmission of HIV between individuals have relied for the most part on *env* sequences and to a lesser degree on *gag* sequences (4,5).

The reverse transcriptase (RT) and protease genes are frequently sequenced in clinical settings to assist physicians in selecting antiretroviral therapy (6), but due to their conserved nature, these genes have not been examined in a phylogenetic context. Not only is there less variability in the RT and protease genes, but also the variability that is present is often caused by drug therapy rather than epidemiological factors. Nonetheless, recent studies have shown that RT and protease display enough interindividual variation to make subtype determination possible (7–9). In addition, interindividual variation is used to exclude PCR contamination since it is unlikely for the RT and protease sequences from two patients to be <2% different (10).

Because the utility of RT and protease for tracking transmission has not been examined, we analyzed the RT and protease sequences of HIV-1 isolates from two

Address correspondence and reprint requests to Sarah Palmer, PhD, HIV Drug Resistance Program, National Cancer Institute at Frederick, 1050 Boyles Street, Building 535, Room 109, Frederick, MD 21702-1201, U.S.A.; e-mail: spalmer@ncifcrf.gov

Manuscript received May 22, 2002; accepted July 23, 2002.

TABLE 1. Antiretroviral treatments and reverse transcriptase (RT) and protease mutations of an HIV-1-infected couple (patients A and B)

	NRTI	NNRTI	PI		Reverse transcriptase												Protease																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																															
					41	49	64	67	69	70	75	83	118	177	184	210	211	214	215	218	219	228	10	24	35	36	37	46	54	57	62	63	64	82																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																														
Patient A Jan 92	—	—	—		K	R																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																										</

Beneath the headings *Reverse transcriptase* and *Protease* are the positions that differed from consensus in at least one of the sequences from patients A and B. The letter directly beneath the position is the consensus amino acid at the position. The letters beneath that are the mutations present in the patient sequences. The presence of more than one mutation indicates a mixture. NRTI, nucleoside RT inhibitors; NNRTI, nonnucleoside RT inhibitors; PI, protease inhibitors; AZT, zalcitabine; DDC, zalcitabine; D4T, stavudine; ABC, abacavir; EFV, efavirenz; RTV, ritonavir; IDV, indinavir; NFV, nelfinavir; SQV, saquinavir.

TABLE 2. Antiretroviral treatments and reverse transcriptase (RT) and protease mutations of an HIV-1-infected couple (patients D and E)

													Reverse transcriptase															
					35	39	41	43	67	69	70	75	98	104	122	174	179	181	184	190	208	210	211	215	218	219	228	248
	NRTI	NNRTI	PI		V	T	M	K	D	T	K	V	A	K	K	Q	V	Y	M	G	H	L	R	T	D	K	L	E
Patient D																												
June 91	—	—	—										S			L												D
May 95	AZT, DDI, DDC, 3TC	—	—		M		L		N	D	R		S						V				K	F	E	Q		D
Jul 96	AZT, 3TC	—	SQV		M		L		N	D	R		S						V		H		K	F	E	Q	H	D
Sep 98	D4T, DDI, AZT, 3TC	NVP, EFV	SQV, IDV, NFV, RTV		M	A	L	E	N	D		M	S		E		I	I		A	Y	L	K	F	E	Q	H	D
Patient E																												
Dec 91	—	—	—										S	N													H	D
Jun 96	AZT, DDI, DDC	—	—						N		R		S	N									K	F		Q	H	
Feb 97	D4T, 3TC	—	—				L	M		N		R	S	N					V				K	F	E	Q	L	D
Sep 98	D4T, 3TC	—	IDV				L		N				S	N			V		V				K	F	E	Q	H	D
Protease																												
	10	20	35	36	37	41	46	54	62	63	71	72	73	82	84	90	93											
	L	K	E	M	N	R	M	I	I	L	A	I	G	V	I	L	I											
Patient D																												
June 91	I		D	L						P																		I
May 95			D	L	D					P																		L
Jul 96	I		D	L	D					P																		
Sep 98	I	R	D	L	D		I	M	V	P	V	L	S		V	M												
Patient E																												
Dec 91		R	D	L		K				P	T																	L
Jun 96		R	D	L		K			I	P	T																	L
Feb 97		R	D	L		K			V	P	T																	L
Sep 98		R	D	L		K		V	V	P	T			A		L	M											L

couples with known transmission histories to ascertain the extent to which the phylogenetic relatedness of these genes was preserved in patients receiving antiretroviral therapy.

METHODS

Patients

Plasma samples were obtained from two HIV-1-infected couples attending an outpatient clinic (Department of Internal Medicine and

Clinical Immunology, University Hospital, Besançon, France) between 1991 and 1998. The HIV-1 transmission history within each couple was known. Patient A, a male intravenous drug user (intravenous drug use between 1978 and 1982), was found to be HIV-1-seropositive in November 1991 when his female sexual partner (patient B), who had no HIV-1 risk factors other than contact with patient A, was found to be HIV-1-seropositive. Patient A also was infected with HCV and had symptoms of chronic active hepatitis; patient B remained HCV-negative for the whole study period. Patient D was infected with HIV-1 in 1987; his male partner (patient E) seroconverted in August 1991. Both recipients (patients B and E) had symptoms of primary HIV

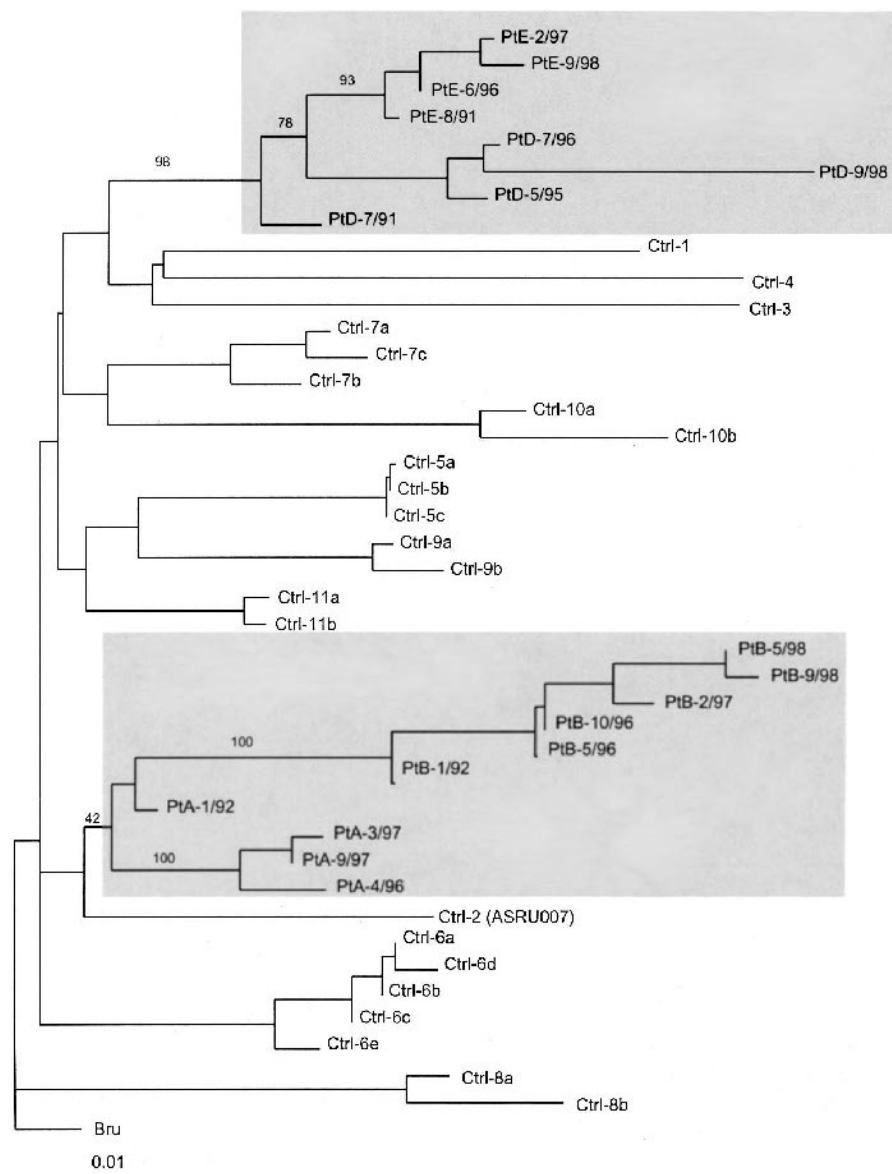


FIG. 1. Neighbor-joining tree constructed from a combination of the reverse transcriptase and protease genes using the HKY85 nucleotide substitution model with a gamma distribution (HKY85 + Γ). Gray boxes highlight sequences from the HIV-1-infected couples. Bootstrap values are provided for the most recent common ancestor of each transmission pair and of the more recent sequences from each patient within the pair.

infection, and antigen to HIV-1 was found in their sera within the 45 days preceding seroconversion. Patient B was seronegative in 1988 (volunteer screening, according to the French National Agency for AIDS Research) when she gave birth to a seronegative child; her partner (patient A) was not tested for HIV at that time.

Sequence Analysis of Patient Strains

A 1.3-kb fragment of cDNA encompassing HIV-1 protease and the first 300 codons of RT was sequenced from patient plasma as previously described (11–13). Sequences were compared with the consensus B sequence to derive a list of mutations (14). Sequences from the two HIV-1-infected couples and control patients (not described previously) were submitted to the GenBank database (AF487122–AF487139).

Phylogenetic Analysis

Phylogenetic trees were constructed from the complete protease sequence, the 5'-polymerase coding region of the RT (codons 1–250), and the concatenation of the protease and RT sequences by means of fastDNAmL (15) and Phylogenetic Analysis Using Parsimony (PAUP*) (16). Neighbor-joining maximum parsimony and maximum likelihood trees were created using a variety of nucleotide substitution models including the uncorrected, Kimura 2 parameter model (which accounts for a transition/transversion bias) and the HKY85 model with gamma distribution (HKY85 + Γ , which accounts for a transition/transversion bias, variable base frequency data, and variable substitution rates at different nucleotide positions) (15,16).

Subtype B sequences from the HIV Outpatient Clinic at Université de Franche-Comté (Besançon, France) and the Stanford University HIV RT and Protease Sequence Database were included in the trees as

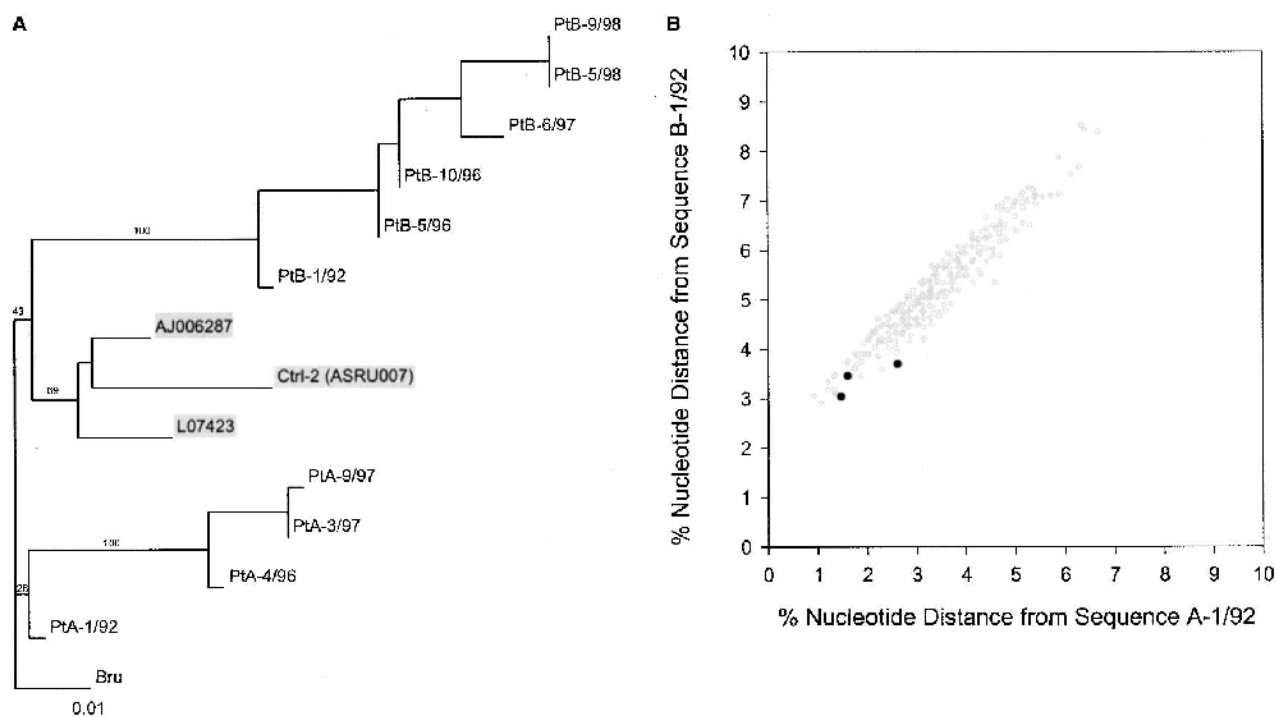


FIG. 2. (A) Maximum likelihood tree constructed from the reverse transcriptase sequences from patients A and B and three epidemiologically unrelated sequences that share a common ancestor with patient B sequences but not with patient A sequences. ASRU007 was isolated in the United Kingdom in 1997, AJ006287 was isolated in Spain in 1989, and L07243 was isolated in Germany some time prior to 1992. Bootstrap values are provided at selected nodes. (B) Distribution of nucleotide distances (%) of other previously reported subtype B GenBank sequences to the baseline sequences from patients A and B using the HKY85 nucleotide substitution model. Black circles indicate the position of the three unrelated isolates that shared a common ancestor with patient B sequences.

controls (17). To minimize bias introduced by HIV drug therapy, trees were also created following the removal of nucleotides at positions associated with drug resistance (protease codons 10, 20, 24, 30, 32, 36, 46, 47, 48, 50, 53, 54, 63, 71, 73, 74, 77, 82, 84, 88, 90, and 93; RT codons 41, 62, 65, 67, 69, 70, 74, 75, 77, 98, 100, 101, 103, 106, 108, 115, 116, 151, 179, 181, 184, 188, 190, 210, 215, 219, 225, and 236).

RESULTS

Plasma samples collected from each patient at four or five time points between 1991 and 1998 were genotyped. The antiretroviral treatment histories and RT and protease mutations are shown in Tables 1 and 2.

Figure 1 shows the neighbor-joining tree constructed from combined protease and RT sequences using the HKY85 + Γ nucleotide substitution model. Bootstrap values are provided for the most recent common ancestor (MRCA) of each transmission pair and the more recent sequences from each patient within the pair. Maximum likelihood and maximum parsimony trees constructed from complete sequences and sequences lacking positions associated with the development of drug resistance had branching patterns identical to those of the neighbor-joining tree.

The sequences obtained from patient D in July 1991 were closest to the MRCA for the sequences from patients D and E, consistent with the history that the virus was transmitted from patient D to patient E at about that time. The bootstrap value at the MRCA node of patients D and E was 98.

The sequence obtained from patient A in January 1992 was closest to the MRCA for the sequences from patients A and B, consistent with the transmission of virus from patient A to patient B at about that time. However, the bootstrap value at the MRCA node for patients A and B was only 42. The genetic distance between the sequences from patients A and B at their initial time points was also higher than expected (2.9%, RT; 0.35%, protease).

Because of the high genetic distance between the initial RT sequences from patients A and B and because of the low bootstrap value at their MRCA node, we constructed neighbor-joining, maximum likelihood, and maximum parsimony trees using RT sequences from the two patients and an additional 316 subtype B control sequences from GenBank, for a total 339 control sequences (including the 23 initial control sequences). In these trees, one of the original control sequences

(ASRU007) and two of the 316 control sequences were found to share an MRCA node with patient B but not with patient A. ASRU007 was isolated in the United Kingdom in 1997, AJ006287 was isolated in Spain in 1989, and L07243 was isolated in Germany some time prior to 1992. A maximum likelihood tree constructed with the RT sequences from patients A and B and the three control sequences is shown in Figure 2A. A plot of the distance of each control sequence to the baseline sequences from patients A and B using the HKY85 nucleotide substitution model is shown in Figure 2B.

DISCUSSION

We examined the phylogenetic relatedness of RT and protease sequences of HIV-1 isolates from two HIV-1-infected French couples with known transmission histories. For both couples, a precise date for the recipients' infections could be assessed because of clinical and biologic signs of primary HIV-1 infection. The phylogenetic relatedness of these virus isolates was examined over time even as the patients were receiving different antiretroviral treatment regimens. The phylogenetic trees reflected the known transmission histories, despite the fact that the drug resistance mutations were most consistent with the patients' drug treatment histories.

However, the initial RT sequences from one couple were relatively divergent. Indeed, three control sequences from unrelated individuals were found to share an MRCA node with recipient sequences (patient B) but not with transmitter sequences (patient A). The divergence between the first two time points for patients A and B is most consistent with a long pretransmission interval during which time patient A's virus developed into a heterogeneous virus population (18). Patient A was presumably infected between 1978 and 1982; patient B was his sexual partner between the end of 1982 and 1991. She was found to be HIV-1-seronegative in 1988 and had clinical and biologic signs of HIV-1 infection in September 1991. The closeness between the three control sequences and patient B's first sample may represent slower evolution in the RT along the branches of the control sequences possibly because of the requirements for enzymatic function.

This study suggests that although the RT and protease genes contain phylogenetic information, they are suboptimal for reconstructing transmission history because the genetic distance between RT and protease isolates obtained from unrelated individuals may occasionally approximate the distance between RT and protease isolates from related individuals. Sequencing of other more variable HIV-1 genes will continue to be required for molecular epidemiological and forensic analyses.

Acknowledgment: The authors thank Victoria Hellmann for excellent technical assistance.

REFERENCES

1. Casado C, Garcia S, Rodriguez C, et al. Different evolutionary patterns are found within human immunodeficiency virus type 1-infected patients. *J Gen Virol* 2001;82:2495–508.
2. Ou CY, Ciesielski CA, Myers G, et al. Molecular epidemiology of HIV transmission in a dental practice. *Science* 1992;256:1165–71.
3. Leitner T, Escanilla D, Franzen C, et al. Accurate reconstruction of a known HIV-1 transmission history by phylogenetic tree analysis. *Proc Natl Acad Sci USA* 1996;93:10864–9.
4. Albert J, Wahlberg J, Leitner T, et al. Analysis of a rape case by direct sequencing of the human immunodeficiency virus type 1 pol and gag genes. *J Virol* 1994;68:5918–24.
5. Leitner T, Kumar S, Albert J. Tempo and mode of nucleotide substitutions in gag and env gene fragments in human immunodeficiency virus type 1 populations with a known transmission history. *J Virol* 1997;71:4761–70.
6. Hirsch MS, Brun-Vezinet F, D'Aquila RT, et al. Antiretroviral drug resistance testing in adult HIV-1 infection: recommendations of an International AIDS Society-USA Panel. *JAMA* 2000;283:2417–26.
7. Pieniazek D, Rayfield M, Hu DJ, et al. Protease sequences from HIV-1 group M subtypes A-H reveal distinct amino acid mutation patterns associated with protease resistance in protease inhibitor-naïve individuals worldwide. HIV Variant Working Group. *AIDS* 2000;14:1489–95.
8. Yahi N, Fantini J, Tourres C, et al. Use of drug resistance sequence data for the systematic detection of non-B human immunodeficiency virus type 1 (HIV-1) subtypes: how to create a sentinel site for monitoring the genetic diversity of HIV-1 at a country scale. *J Infect Dis* 2001;183:1311–7.
9. Gonzales MJ, Machekano RN, Shafer RW. Human immunodeficiency virus type 1 reverse-transcriptase and protease subtypes: classification, amino acid mutation patterns, and prevalence in a northern California clinic-based population. *J Infect Dis* 2001;184:998–1006.
10. Shafer RW, Dupnik K, Winters MA, et al. *A guide to HIV-1 reverse transcriptase and protease sequencing for drug resistance studies. HIV sequence compendium*. Los Alamos, NM: Los Alamos National Laboratories, 2000:506–57.
11. Winters MA, Schapiro JM, Lawrence J, et al. HIV-1 protease genotypes and in vitro protease inhibitor susceptibility from long-term saquinavir-treated individuals who switched to other protease inhibitors. *J Virol* 1998;72:5303–6.
12. Nijhuis M, Boucher CAB, Schuurman R. Sensitive procedure for amplification of HIV-1 RNA using a combined reverse transcriptase and amplification reaction. *Biotechniques* 1995;19:178–82.
13. Shafer RW, Hertogs K, Zolopa AR, et al. High degree of interlaboratory reproducibility of human immunodeficiency virus type 1 protease and reverse transcriptase sequencing of plasma samples from heavily treated patients. *J Clin Microbiol* 2001;39:1522–9.
14. Kuiken CL, Foley B, Hahn BH, et al. *Human retroviruses and AIDS: a compilation and analysis of nucleic acid and amino acid sequences*. Los Alamos, NM: Los Alamos National Laboratories, 1999.
15. Olsen GJ, Matsuda H, Hagstrom R, et al. fastDNAmL: a tool for construction of phylogenetic trees of DNA sequences using maximum likelihood. *Comput Appl Biosci* 1994;10:41–8.
16. Swofford DL. *PAUP*. Phylogenetic analysis using parsimony (*and other methods)*. Version 4. Sunderland, MA: Sinauer Associates, 1998.
17. Shafer RW, Stevenson D, Chan B. Human immunodeficiency virus reverse transcriptase and protease sequence database. *Nucleic Acids Res* 1999;27:348–52.
18. Leitner T, Albert J. The molecular clock of HIV-1 unveiled through analysis of a known transmission history. *Proc Natl Acad Sci USA* 1999;96:10752–7.