JOURNAL OF VIROLOGY, Sept. 1994, p. 5918–5924
0022-538X/94/$04.00+0

Vol. 68, No. 9

Analysis of a Rape Case by Direct Sequencing of the Human Immunodeficiency Virus Type 1 *pol* and *gag* Genes

JAN ALBERT,[1]* JOHAN WAHLBERG,[2] THOMAS LEITNER,[1,2] DAVID ESCANILLA,[1] AND MATHIAS UHLÉN[2]

Department of Clinical Virology, Swedish Institute for Infectious Disease Control, Karolinska Institute, S-105 21 Stockholm,[1] and Department of Biochemistry and Biotechnology, Royal Institute of Technology, S-100 44 Stockholm,[2] Sweden

Received 14 June 1993/Accepted 24 May 1994

Transmission of human immunodeficiency virus type 1 (HIV-1) from a male accused of rape and deliberate transmission of HIV-1 was investigated by sequencing of the HIV-1 *pol* and *gag* genes from virus obtained from the male and from the female victim. Parts of the reverse transcriptase and p17^{gag} genes were amplified and directly sequenced from uncultured peripheral blood mononuclear cells. The sequences were compared with sequences from 21 unrelated HIV-1-infected controls from the same geographic area (Stockholm, Sweden). Bootstrap analysis of phylogenetic trees demonstrated that the sequences from the female were significantly more closely related to the sequences from the male than to sequences from the controls. Furthermore, we found that the male and female shared two distinct genetic variants of HIV-1. In p17^{gag} the major variant had an unusual, out-of frame deletion of 3 nucleotides which the minor variant lacked. These results indicated that the male had transmitted more than one infectious unit to the female. From this study we concluded that it was highly likely that the HIV-1 strains carried by the male and the female were closely epidemiologically linked.

The possibility of studying the molecular epidemiology of human immunodeficiency virus type 1 (HIV-1) infection has been greatly augmented by the introduction of the PCR. Most studies have concentrated on sequence variation in the V3 loop of the external envelope glycoprotein gp120, since this region contains important determinants for neutralization and cell tropism (19, 20). Thus, several authors have shown that V3 loop sequences obtained during or shortly after seroconversion are homogeneous, although the virus population later displays a high degree of sequence variation (23, 31, 33, 34, 36). Sequence analysis of the V3 loop was also used to investigate the "Florida dentist case" (18). From this analysis Ou et al. concluded that a dentist had infected several of his patients. In a similar study, Holmes et al. (10) used p17^{gag} sequences to investigate whether a HIV-1 infected surgeon was likely to have infected one of his patients. Their investigation showed that the patient was probably infected by transfusion of HIV-1-contaminated blood and not by the surgeon.

We have recently described a rapid, automated method for direct genomic sequencing (2, 12, 27, 28, 30). Here, we have used this method to determine *pol* and *gag* gene sequences in HIV-1 from a male accused of rape and from the female victim. The male was a Swedish intravenous-drug abuser who was found to be HIV-1 positive in 1986. Before the present analysis was conducted, he had been convicted of rape and deliberate transmission of HIV-1 in the Stockholm district court. The evidence the conviction rested upon did not include any forensic analysis since the female did not file a report until 1 year after the alleged rape. However, she had been shown to seroconvert for HIV-1 antibodies within a few weeks after the rape. Therefore, a genetic analysis of the HIV-1 strains carried by the male and the female was performed before the case was tried in the court of appeal. Thus, the purpose of the investigation was to determine if viral sequences from the female were significantly more closely related to sequences from the male than to sequences from relevant controls. The study demonstrated that the HIV-1 strains carried by the male and the female were genetically very closely related. In addition, we found that the female probably was infected by more than one infectious unit from the male.

MATERIALS AND METHODS

Patients. Two samples of whole blood were obtained on separate days from the male and from the female. In addition, we had access to frozen peripheral blood mononuclear cells (PBMC) from the female obtained 17 months earlier, i.e., 5 weeks after the suspected transmission. The female experienced a symptomatic primary HIV-1 infection 3 weeks after the transmission, and when the first PBMC sample was obtained after an additional 2 weeks, her serum was HIV-1 antibody positive and HIV-1 antigen negative. Frozen PBMC samples from 21 unrelated HIV-1-infected individuals (10 intravenous-drug abusers and 11 homosexual men) from the same geographic area (Stockholm, Sweden) served as controls. Care was taken in the selection of the controls, especially the intravenous-drug abusers, to ensure that they represented a cross section of virus strains present in Stockholm. Thus, these samples were obtained between 1986 and 1989 from patients attending both of the infectious disease clinics in Stockholm.

Direct DNA sequencing. A region of the *pol* gene corresponding to amino acids 8 to 222 of the reverse transcriptase was directly sequenced from uncultured PBMC from the male, the female, and the controls. Similarly, a region corresponding to amino acids 32 to 132 of p17^{gag} and the first 8 amino acids of p24^{gag} was sequenced from the male, the female, and selected controls. The PBMC were isolated from whole blood, lysed, and used for PCR amplification as previously described (1, 28). The nested primers used to amplify a region corresponding to amino acids 8 to 222 of the reverse transcriptase were as follows (their positions in the HIV-1 MN strain are in parentheses): the outer pair was JA99, 5′-GGGGGAATTG GAGGTTTTATCAAAG-3′ (2409 to 2433), and RIT137 5′-

* Corresponding author. Phone: 46-8-735-1300.

TTCTGTATGTCATTGACAGTCCAGC-3′ (3340 to 3316); the inner pair was JA100, 5′-GACCTACACCTGTCAACAT AATTGG-3′ (2500 to 2524), and RIT136, 5′-biotin-GATG GAGTTCATAACCCATCCAAAG-3′ (3270 to 3246). The nested primers for p17^{gag} were as follows: the outer primers were JA112 5′-GGAGAGAGATGGGTGCGAGAGCGTC-3′ (779 to 803), and JA115, 5′-GGCTCCTTCTGATAATGCT GAAAAC-3′ (1332 to 1308); the inner primers were JA113, 5′-CGATGGGAAAAAATTCGGTTAAGGC-3′ (829 to 853), and JA114, 5′-biotin-TCTCTTCTACTACTTTTACCCATG C-3′ (1279 to 1255). The product from the first PCR was diluted 100-fold before the second PCR with the inner primers. The temperature cycles were 94°C for 1 min, 60°C for 1 min, and 72°C for 2 min, repeated 30 times for both *pol* primer pairs, and 94°C for 30 s, 60°C for 30 s, and 72°C for 1 min, repeated 30 times for both p17^{gag} primer pairs. The PCR product was directly sequenced without prior cloning by an automated sequencing method based on a paramagnetic solid support (2, 27). The sequences of the fluorescein isothiocyanate (FITC)-labelled sequencing primers used in this study were as follows: RIT227, 5′-FITC-AATCTGTTGACTCA GATTGG-3′ (2529 to 2548); RIT136F, 5′-FITC-GATG GAGTTCATAACCCATCCAAAG-3′ (3270 to 3246); JA116, 5′-FITC-GAAAAAATTCGGTTAAGGCC (835 to 859); and JA117, 5′-FITC-GCATTTAAAGTTCTAGGTGA-3′ (1261 to 1237). Each sample (two from the male and three from the female) was analyzed on separate occasions, and both DNA strands were sequenced to yield a partial overlap in the middle of the sequenced region. Recommended laboratory routines were followed to avoid false positives in the PCR, and relevant positive and negative controls were included in each analysis (15).

Cloning. The accuracy of the direct sequencing method was analyzed by comparison with sequences from multiple clones generated from the PCR-amplified material as follows. The product from the PCR with the outer primers (JA99 and RIT137) was amplified with the modified inner primers RIT331, 5′-CCCGAATTCGATGGAGTTCATAACCCATC CAAAG-3′ (3270-3246), and RIT333, 5′-CCCGGATCCATC CATACAATACTCCAGTATTTGC-3′ (2725-2749) (underlined bases correspond to restriction sites for *Eco*RI and *Bam*HI, respectively). Following amplification and phenol extraction, the fragment was restricted by 10 U each of *Eco*RI and *Bam*HI. The restricted fragment was ligated to *Eco*RI and *Bam*HI-digested pRIT28 (13) and transformed to *Escherichia coli* RRI M15 (21). The cloned fragment was sequenced as described previously (11).

For p17^{gag}, an approach described by Simmonds et al. (24) was used to sequence individual viral clones. The HIV-1 DNA copy numbers in one of the PBMC lysates obtained at 18 months from the male and the female were determined by limiting dilution as previously described (4, 24). Then 45 PCRs were performed on PBMC lysates which had been diluted to contain on average 0.3 molecule per PCR. These amplifications yielded 13 positive PCRs from the male and 15 from the female, each likely to represent the amplification of a single HIV-1 molecule. The PCR products were directly sequenced as described above.

Phylogenetic tree analysis. Phylogenetic tree analyses of nucleotide sequences were conducted with three different programs from the PHYLIP package, version 3.41, provided by J. Felsenstein (6): DNAML (maximum likelihood), DNA PARS (maximum parsimony) (6), and NEIGHBOR (neighbor joining) (22). The distance matrix needed for the neighbor-joining method was obtained with the DNADIST program. For DNAML and DNADIST the value for the transition/transversion ratio was set at 1.5 as described by Holmes et al. (10). Multiple runs performed with the JUMBLE option with all three programs gave the same general branching pattern. Additional analyses performed using the TREEALIGN software (version November 1990) written by Jotun Hein (7) and distributed by the EMBL database also gave the same general pattern. Bootstrap analysis was performed on maximum likelihood and parsimony trees with the DNABOOT program of PHYLIP; in each case at least 100 resamplings were performed.

RESULTS

Analysis of the sequences from the male and the female. The direct DNA sequence analysis showed that the *gag* and *pol* genes of the virus populations harbored by the male and the female were very closely related (Fig. 1 and 2). In *pol*, all sequences from the male and the female differed from the consensus sequence at position 49 (Lys-49→Arg) and position 204 (Glu-204→Lys). In addition, mixed sequences were observed at positions 169, 178, 202, 207, and 211, indicating the presence of genetic polymorphism within the individual samples (Fig. 1). Similarly, the p17^{gag} sequences from both the male and the female differed from the consensus sequence at positions 91, 94, 121, 122, 124, and 138 (Fig. 2). In addition, several positions in *gag* also displayed sequence polymorphisms. Interestingly, most of the polymorphic sites in the *pol* and *gag* genes were shared between the male and the female, although the relative proportion of the individual bases sometimes differed. It should be stressed that the sample which was obtained from the female 5 weeks after infection also displayed most of these polymorphic sites.

To prove that the apparent genetic polymorphism was not an artifact of the direct sequencing method, the PCR products from the *pol* gene were cloned and sequenced individually. These experiments confirmed the existence of polymorphism in all sites identified as polymorphic by direct sequencing (Fig. 3). Note that several additional mutations were observed among individual clones. They may represent true variations, but they may also be errors introduced by the *Taq* polymerase during PCR.

We also amplified and directly sequenced the p17^{gag} genes of single HIV-1 DNA molecules. Figure 4 shows the nucleotide sequences over a region corresponding to amino acids 110 to 127 of the p17^{gag} protein, where the direct PBMC sequence indicated that the virus populations in both the male and the female harbored virus variants differing in length. This was confirmed by the analysis of single viral DNA molecules, which showed that 4 of 13 molecules from the male and 1 of 15 molecules from the female had normal lengths, whereas a majority of the molecules had an unusual out-of frame deletion of 3 nucleotides.

Phylogenetic tree analyses. In a first set of experiments we sequenced the *pol* genes from the male, the female, and all of the controls (unrelated Swedish intravenous-drug abusers and homosexual men). The sequences from the male, the female, and the controls belonging to the same risk group as the male (intravenous-drug users) are presented in Fig. 1. Phylogenetic tree analyses using several different algorithms showed that the *pol* gene sequences from the male and the female clustered together. A representative *pol* tree generated with the maximum likelihood algorithm is presented in Fig. 5A. As expected, the sequences from the male and the female were more closely related to sequences from other intravenous-drug abusers than to sequences from homosexual men.

Bootstrap analysis was used to examine whether the clade

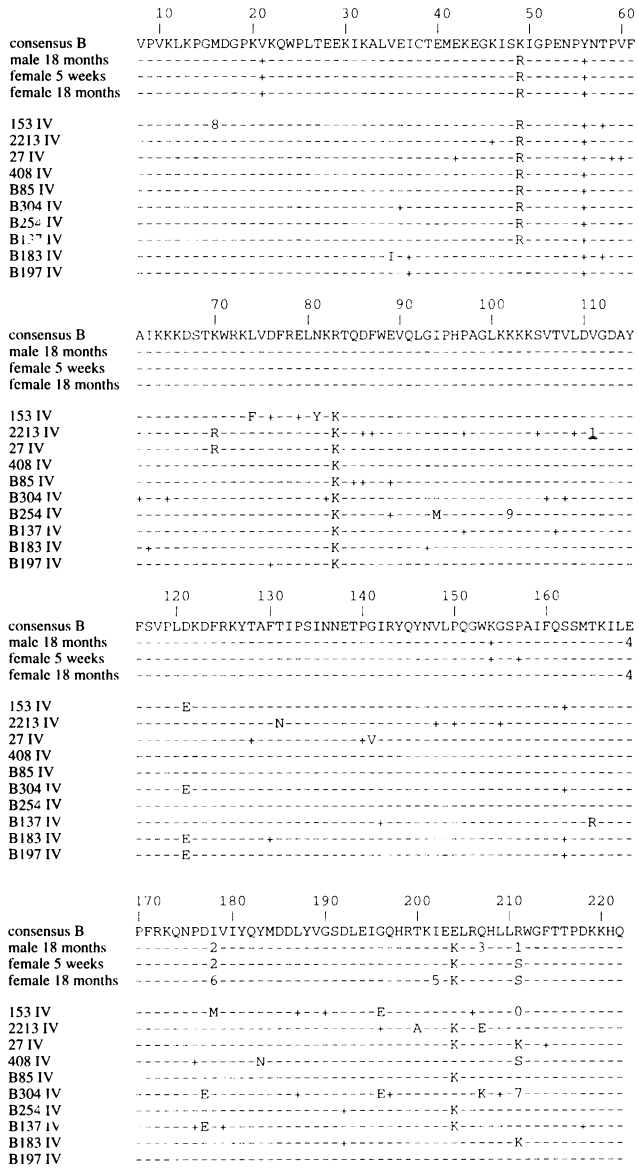

FIG. 1. Comparison of HIV-1 reverse transcriptase amino acid sequences from the male, the female, and 10 intravenous-drug abusers (IV). The amino acid sequences were deduced from direct DNA sequences of virus populations in uncultured PBMC. The sequences have been aligned to a consensus of HIV-1 subtype B sequences (Los Alamos Database). A dash indicates identity with the consensus sequence; a letter indicates an amino acid change relative to the consensus; a plus sign indicates a silent nucleotide change or polymorphism relative to the consensus. A number indicates a nonsilent polymorphism according to the following code: 1 = K, N, R, and S; 2 = I and M; 3 = Q and E; 4 = D and E; 5 = I and V; 6 = I, M, and L; 7 = R and S; 8 = K and M; 9 = K and E; 0 = R and G; and 1 = V and L. A nucleotide was scored as polymorphic only if at least 20% of the virus population was reproducibly estimated to be mutant.

consisting of the *pol* gene sequences from the male and the female was likely to represent a true monophyletic group (Fig. 5A). In different analyses the bootstrap values ranged from 58 to 63%. The exact interpretation of results from bootstrap analysis is unclear, but Hillis and Bull (8) have shown that bootstrap proportions of >70% in most situations correspond

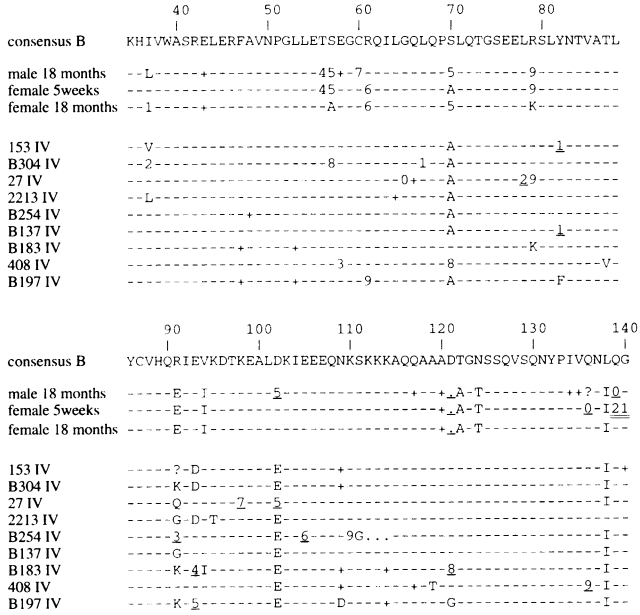

FIG. 2. Comparison of HIV-1 p17^{gag} amino acid sequences from the male, the female, and nine intravenous-drug abusers. The amino acid sequences were deduced from direct DNA sequences of virus populations in uncultured PBMC. The amino acid sequences have been aligned to a consensus of HIV-1 subtype B sequences (Los Alamos Database). A period indicates a gap introduced to align the sequences; an underlined period indicates that a major proportion of the virus population had an deletion at this position whereas a minor proportion of the virus population had an aspartic acid; other symbols are as described for Fig. 1. A number indicates a nonsilent polymorphism according to the following code: 1 = I and L; 2 = I and V; 3 = E and G; 4 = S and T; 5 = A and S; 6 = R and stop; 7 = C and F; 8 = S and stop; 9 = K and R; 0 = G and V; $\underline{1}$ = F and Y; $\underline{2}$ = F and L; $\underline{3}$ = G and R; $\underline{4}$ = A and D; $\underline{5}$ = D and E; $\underline{6}$ = E and K; $\underline{7}$ = K and Q; $\underline{8}$ = D and G; $\underline{9}$ = M and I; $\underline{0}$ = L and Q; $\underline{\underline{1}}$ = A and G; and $\underline{\underline{2}}$ = H and Q. A nucleotide was scored as polymorphic only if at least 20% of the virus population was reproducibly estimated to be mutant.

to a probability of >95% that the corresponding clade is real. Thus, we could not conclusively show that the *pol* gene sequences from the female were statistically more closely related to sequences from the male than to sequences from the some of the controls, even though the sequences from the male and the female clustered together.

To increase the statistical power of our investigation we decided to also sequence part of the *gag* gene from the male, the female, and the most closely related controls, i.e., the unrelated intravenous-drug abusers (Fig. 2). The p17^{gag} region was chosen since it appears to be especially suitable for epidemiological investigations (10). Figure 5B shows the phylogenetic tree and the bootstrap values obtained with the p17^{gag} sequences (control B85 was not included since PCR amplification from this sample was not successful). The p17^{gag} tree resembled the *pol* gene tree in that the sequences from the male and the female clustered together. The high bootstrap value (97%) strongly suggested that they represented a true monophyletic clade.

Transmission of more than one infectious unit. The analysis of the HIV-1 *gag* and *pol* gene sequences from the male revealed several positions with clear evidence of genetic polymorphism (Fig. 1 and 2). Many of these positions were also polymorphic in the sequences from the female. Furthermore,

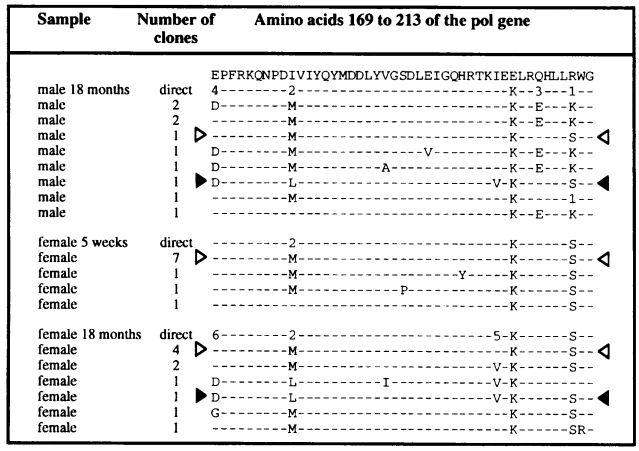

FIG. 3. Comparison of HIV-1 reverse transcriptase amino acid sequences obtained by direct DNA sequencing and by DNA sequencing of multiple clones obtained by molecular cloning of the PCR products. Two distinct genetic variants present in both the female and the male have been indicated by open and solid triangles. Other symbols are as described for Fig. 1.

the male and the female shared two distinct genetic variants in both the *pol* and the *gag* genes (Fig. 3 and 4). In $p17^{gag}$ the major virus variant in both the male and the female had an unusual out-of frame deletion of 3 nucleotides which the major variant lacked. These findings strongly suggest that more than one infectious unit was transmitted from the male to the female, since otherwise it would be highly unlikely that so many sequence polymorphisms would be shared between the male and the female.

Detection of AZT resistance-associated mutations in untreated intravenous-drug users. The HIV-1 *pol* gene sequences from the intravenous-drug users revealed that two patients (no. 27 and 2213) harbored virus with a Lys-70→Arg mutation (Fig. 1) that has been reported to confer partial zidovudine (AZT) resistance (16). This was surprising since these patients never had received AZT treatment. Patient 27 was tested in January 1987, and at this time no patients had been given AZT in Sweden. Therefore, the Arg-70 mutation in this patient represents a natural and not a drug-induced sequence variation. Patient 2213 was tested in September 1989 during an ongoing primary infection. Thus, the presence of the Arg-70 mutation in this patient could be due either to a natural variation as in patient 27 or to the transmission of an AZT-resistant HIV-1 variant.

DISCUSSION

This article describes the use of DNA sequencing of the HIV-1 *gag* and *pol* genes in the investigation of a suspect in a rape case. By comparing the sequences from the male, the female, and the controls we concluded that the HIV-1 strains carried by the two parties were very closely related and that it was highly likely that they were epidemiologically linked. On the basis of the analysis presented here and other evidence in the case, the verdict from the district court was upheld in the court of appeal. It is important to stress that even though our investigation showed that the strains carried by the male and the female were epidemiologically linked, we could not determine the direction of the transmission, nor could we formally rule out the possibility that both the male and the female were infected by a third party. Thus, it was essential that the results from our sequence investigation be used in conjunction with other epidemiological information in the case.

In an epidemiological investigation of this type the selection of controls is very important. Thus, all the controls were from the same geographical area (Stockholm, Sweden) as the male and the female. Furthermore, we took care to select samples from individuals belonging the same risk group as the male (intravenous-drug abusers) as well as from another risk group (homosexual men). The samples from the intravenous-drug abusers were obtained over several years and from both of Stockholm's two infectious disease clinics, to ensure that they were likely to represent a cross section of HIV-1 strains from Stockholm's intravenous-drug abusers. The phylogenetic tree analysis showed that the sequences from the intravenous-drug abusers, including those from the male and from the female, formed a cluster which was distinct from and less divergent than that of the homosexual men. This fitted with the epidemiological information, since the male had been infected

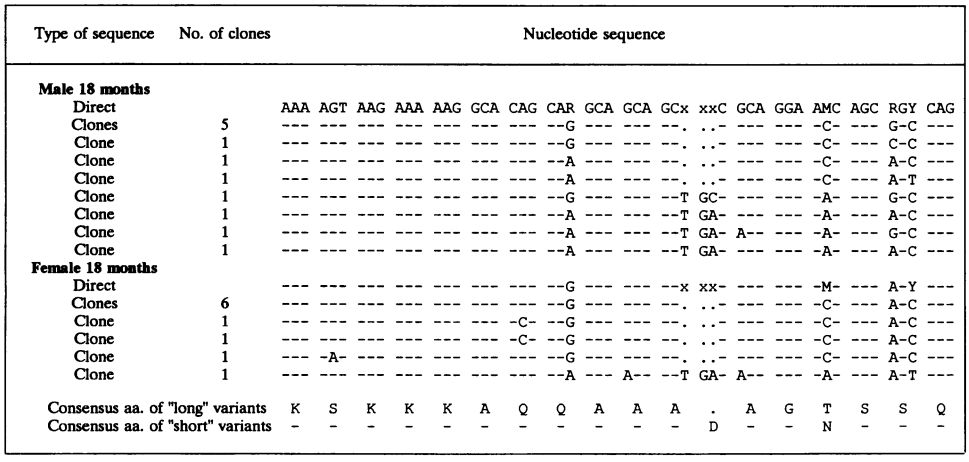

FIG. 4. Comparison of HIV-1 $p17^{gag}$ nucleotide sequences obtained by direct sequencing of undiluted PBMC lysates and by sequencing of single HIV-1 molecules (clones) obtained by limiting dilution of the PBMC lysates prior to PCR. xxx, the direct sequence indicated that a minor proportion of the virus population had an insertion at this position. Other symbols are as described for Fig. 2. aa, amino acids.

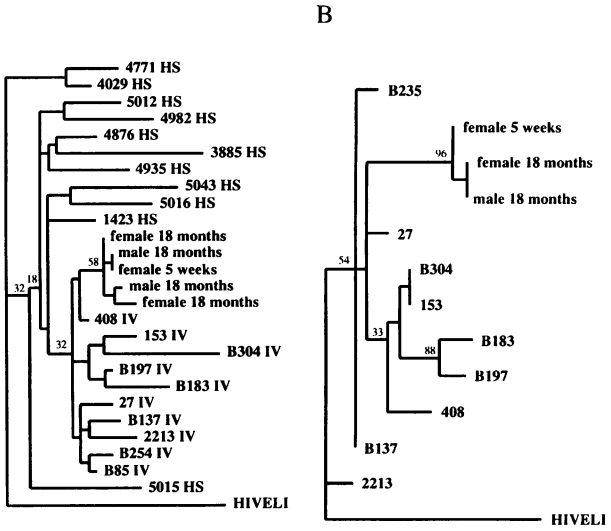

FIG. 5. (A) Phylogenetic tree analysis comparing HIV-1 reverse transcriptase nucleotide sequences obtained by direct sequencing of PBMC from the male, the female, 10 unrelated intravenous-drug abusers (IV), and 11 unrelated homosexual men (HS). The sequence of the African HIV-1 isolate ELI (HIVELI) was used as an outgroup to root the tree. The analysis was performed with DNA sequences corresponding to amino acids 8 to 222. Some controls had received AZT treatment; therefore, the nucleotides corresponding to amino acids 41, 67, 70, 215, and 219, which confer AZT resistance (16, 26), were excluded from the analysis. The tree was generated with the DNAML program from the PHYLIP package, version 3.4 (6). Selected bootstrap values are given. (B) Phylogenetic tree analysis comparing HIV-1 $p17^{gag}$ nucleotide sequences obtained by direct sequencing of PBMC from the male, the female, and nine unrelated intravenous-drug abusers. The sequence of the African HIV-1 isolate ELI (HIVELI) was used as an outgroup to root the tree. The analysis was performed with DNA sequences corresponding to amino acids 32 to 140. The tree was generated with the DNAML program from the PHYLIP package, version 3.4 (6). Selected bootstrap values are given.

through intravenous-drug use in Stockholm. Furthermore, the fact that the sequences from the male and the female clustered together with those from the intravenous-drug abusers indicated that the controls were adequately chosen.

Our investigation showed that the male and female shared two distinct genetic variants in both *gag* and *pol*; in $p17^{gag}$ the two variants even differed in length. This type of sequence similarity forms a very strong epidemiological link, which is not fully appreciated by phylogenetic tree analysis. In addition, this strongly indicated that more than one infectious unit was transmitted from the male to the female. Earlier investigations, which have shown that V3 sequences are homogeneous early after infection, have by some investigators been interpreted as evidence for the transmission of a single infectious unit (33). However, Zhang et al. as well as Zhu et al. recently showed that the homogeneity is confined to the *env* gene and is not present in $p17^{gag}$ (34, 36). Our study supports these data since the sample from the female obtained 5 weeks after transmission was heterogeneous in sequence, especially in the $p17^{gag}$ gene. Data suggesting that more than one infectious unit may be transmitted have been presented only once previously: in the Florida dentist case, the dentist and patient A were shown to share two genetic variants of HIV-1 (14). Clearly more studies are needed to further address this problem, since our study has not determined how many genomes were transmitted or whether transmission of several genomes is common or rare. These questions are important for understanding how HIV-1 establishes infection and therefore for designing strategies for reducing mother-to-child and sexual transmission of HIV-1.

Analysis by DNA sequencing has to our knowledge been used only once earlier in forensic medicine, i.e., in the Florida dentist case (18). However, this is the first time that evidence of this type has been used in court, since a settlement was reached in the Florida case before the case was taken to court. It is interesting to compare the strategies employed in these two investigations. One important difference is that we studied the *gag* and the *pol* genes instead of the more variable *env* gene. Recent data suggest that the V3 loop may be less suitable for epidemiological studies than other regions of the genome, since specific V3 sequences appear to be selected for during primary HIV-1 infection (34). Furthermore the V3 loop is under strong immunological selection pressure after seroconversion (9, 25, 32). Thus, sequence variation in the V3 loop may in some situations reflect not only evolutionary distance but also different immunological selection. In our study we first analyzed the *pol* gene, since we had observed that sequential samples from single individuals usually displayed unique signature sequence patterns in *pol* (27). However, the bootstrap analysis indicated that the number of informative sites was too low to show that the sequences from the female were significantly more closely related to those from the male than to those from some of the controls. We therefore decided to also analyze the $p17^{gag}$ region, which is more variable than the reverse transcriptase but still does not appear to be under strong selection pressure (34). Our decision was influenced by recent data which suggest that the $p17^{gag}$ region may be especially well suited for epidemiological studies (10).

In this study we have shown that consensus sequences obtained by the direct sequencing method and by sequencing of multiple clones are virtually identical, confirming our previous findings (17). This eliminates the need to analyze multiple clones from each sample for this type of epidemiological study. Another advantage of direct sequencing is that errors introduced by the *Taq* polymerase during PCR do not cause problems (24). Even in the most extreme case, in which one single DNA molecule is added to the PCR mixture and a *Taq* error occurs in the first PCR cycle, only 25% of the final PCR product would display the mutation, since the two original strands and one of the two copies generated in the first PCR would still be correct. In the present investigation we know that more than 10 DNA molecules were added to each PCR mixture, since we quantified the HIV-1 DNA load in the samples. Thus, it is highly unlikely that the direct sequences contain *Taq*-induced errors. In contrast, we cannot exclude the possibility that some of the rare mutations in our pol clones are *Taq* induced, since these clones were generated from the PCR products. Holmes et al. (10) used an alternative method to overcome the problems associated with *Taq*-induced errors, which we also used to generate our $p17^{gag}$ clones. Single HIV-1 molecules derived from limiting dilution of the DNA samples were PCR amplified and directly sequenced. However, this method does not eliminate the need to sequence several molecules to determine a consensus sequence.

We have previously shown that it is possible to detect a genetic variant that represents 10 to 15% of the virus population in a sample by direct sequencing (17). In the present study, a nucleotide was scored as polymorphic only if at least 20% of the virus population was reproducibly estimated to be mutant to avoid falsely calling a nucleotide polymorphic. This level of resolution is likely to be sufficient for epidemiological investigations but may be insufficient in some other situation.

In two of the intravenous-drug users that served as controls

we detected the Lys-70→Arg mutation that has been reported to confer AZT resistance (16). This finding is interesting since these patients had not received AZT treatment. Zhang et al. (35) recently reported a similar finding for one patient. This may explain why this mutation usually is the first that develops in AZT-treated patients (3). For one of the two patients the Arg-70 mutation could not have been drug induced, since the sample from this patient was obtained before AZT was introduced in Sweden. However, for the other patient, who was tested during primary HIV-1 infection, it cannot be excluded that an AZT-resistant variant was transmitted from an AZT-treated patient. Erice et al. (5) recently reported a similar finding, and we have also documented one such case (29).

In summary, the analysis revealed that the sequences from the male and the female were very closely related. It was even possible to identify two distinct genetic variants which were shared by the male and the female. Thus, it is highly likely that HIV strains carried by the male and the female are epidemiologically linked.

ACKNOWLEDGMENTS

This study was supported by grants from the Swedish Board for Technical Development and the Swedish Medical Research Council. David Escanilla is supported by a grant from the Swedish Agency for Research Cooperation with Developing Countries (SAREC).

We are grateful to E. Holmes for invaluable help and suggestions; to E. Norrby and E. M. Fenyö for critically reading the manuscript; to T. Hultman, J. Lundeberg, U. Gyllensten, and E. Jonsson for helpful discussions; and to L. Benthin, K. Aperia, and B. Abrahamsson for expert technical assistance.

REFERENCES

1. **Albert, J., and E. M. Fenyö.** 1990. Simple, sensitive and specific detection of HIV-1 in clinical specimens by polymerase chain reaction with nested primers. J. Clin. Microbiol. **28:**1560–1564.
2. **Albert, J., J. Wahlberg, J. Lundeberg, S. Cox, E. Sandström, B. Wahren, and M. Uhlén.** 1992. Persistence of azidothymidine-resistant human immunodeficiency virus type 1 RNA genotypes in posttreatment sera. J. Virol. **66:**5627–5630.
3. **Boucher, C. A. B., E. O'Sullivan, J. W. Mulder, C. Ramautarsing, P. Kellam, G. Darby, J. M. A. Lange, J. Goudsmit, and B. A. Larder.** 1992. Ordered appearance of zidovudine resistance mutations during treatment of 18 human immunodeficiency virus-positive subjects. J. Infect. Dis. **165:**105–110.
4. **Brinchmann, J. E., J. Albert, and F. Vartdal.** 1991. Few infected $CD4^+$ T cells but a high proportion of replication-competent provirus copies in asymptomatic human immunodeficiency virus type 1 infection. J. Virol. **65:**2019–2023.
5. **Erice, A., D. L. Mayers, D. G. Strike, K. J. Sannerud, F. E. McCutchan, K. Henry, and H. Balfour, Jr.** 1993. Primary infection with zidovudine-resistant human immunodeficiency virus type 1. N. Engl. J. Med. **328:**1163–1165.
6. **Felsenstein, J.** 1991. Phylip manual, version 3.4, Berkeley Herbarium. University of California, Berkeley.
7. **Hein, J.** 1990. A unified approach to alignment and phylogenies. Methods Enzymol. **183:**626–645.
8. **Hillis, D. M., and J. J. Bull.** 1993. An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. Syst. Biol. **42:**182–192.
9. **Holmes, E. C., L. Q. Zhang, P. Simmonds, C. A. Ludlam, and A. J. Leigh Brown.** 1992. Convergent and divergent sequence evolution in the surface envelope glycoprotein of human immunodeficiency virus type 1 within a single infected patient. Proc. Natl. Acad. Sci. USA **89:**4835–4839.
10. **Holmes, E. C., L. Q. Zhang, A. Smith Rogers, and A. J. Leigh Brown.** 1993. Molecular investigation of HIV infection in a patient of an HIV-infected surgeon. J. Infect. Dis. **167:**1411–1414.
11. **Hultman, T., S. Bergh, T. Moks, and M. Uhlén.** 1991. Bidirectional solid-phase sequencing of in vitro-amplified plasmid DNA. BioTechniques **10:**84–93.
12. **Hultman, T., S. Ståhl, E. Hornes, and M. Uhlén.** 1989. Direct solid phase sequencing of genomic and plasmid DNA using magnetic beads as solid support. Nucleic Acids Res. **17:**4937–4946.
13. **Hultman, T., S. Ståhl, T. Moks, and M. Uhlén.** 1988. Approaches to solid phase DNA sequencing. Nucleosides Nucleotides **7:**629–638.
14. **Korber, B., and G. Myers.** 1992. Signature pattern analysis: a method for assessing viral sequence relatedness. AIDS Res. Hum. Retroviruses **8:**1549–1560.
15. **Kwok, S., and R. Higuchi.** 1989. Avoiding false positives with PCR. Nature (London) **339:**237–238.
16. **Larder, B. A., and S. D. Kemp.** 1989. Multiple mutations in HIV-1 reverse transcriptase confer high-level resistance to zidovudine (AZT). Science **246:**1155–1158.
17. **Leitner, T., E. Halapi, G. Scarlatti, P. Rossi, J. Albert, E. M. Fenyö, and M. Uhlén.** 1993. Analysis of heterogeneous viral populations by direct DNA sequencing. BioTechniques **15:**120–127.
18. **Ou, C.-Y., C. A. Ciesielski, G. Myers, C. I. Bandea, C.-C. Luo, B. Korber, J. I. Mullins, G. Schochetman, R. L. Berkelman, A. N. Economou, J. J. Witte, L. J. Furman, G. A. Satten, K. A. MacInnes, J. W. Curran, and H. W. Jaffe.** 1992. Molecular epidemiology of HIV transmission in a dental practice. Science **256:**1165–1171.
19. **Palker, T. J., M. E. Clark, A. Langlois, T. J. Matthews, K. J. Weinhold, R. R. Randall, D. P. Bolognesi, and B. F. Haynes.** 1988. Type-specific neutralization of the human immunodeficiency virus with antibodies to env-encoded synthetic peptides. Proc. Natl. Acad. Sci. USA **85:**1932–1936.
20. **Rusche, J. R., K. Javaherian, C. McDanal, J. Petro, D. L. Lynn, R. Grimaila, A. Langlois, R. C. Gallo, L. O. Arthur, P. J. Fischinger, D. P. Bolognesi, S. D. Putney, and T. J. Matthews.** 1988. Antibodies that inhibit fusion of human immunodeficiency virus-infected cells bind a 24-amino acid sequence of the viral envelope, gp120. Proc. Natl. Acad. Sci. USA **85:**3198–3202.
21. **Rüther, U.** 1982. pUR 250 allows rapid chemical sequencing of both DNA strands of inserts. Nucleic Acids Res. **10:**5765–5772.
22. **Saitou, N., and M. Nei.** 1987. The neighbour-joining method: a new method for reconstructing evolutionary trees. Mol. Biol. Evol. **4:**406–425.
23. **Scarlatti, G., T. Leitner, E. Halapi, J. Wahlberg, P. Biraghi, M. A. Clerici-Schoeller, H. Wigzell, E. M. Fenyö, J. Albert, M. Uhlén, and P. Rossi.** 1993. Comparison of V3 sequences from HIV-1 infected children to the RNA and DNA virus populations of their mothers. Proc. Natl. Acad. Sci. USA **90:**1721–1725.
24. **Simmonds, P., P. Balfe, J. F. Peutherer, C. A. Ludlam, J. O. Bishop, and A. J. Leigh Brown.** 1990. Human immunodeficiency virus-infected individuals contain provirus in small numbers of peripheral mononuclear cells and at low copy numbers. J. Virol. **64:**864–872.
25. **Simmonds, P., L. Q. Zhang, F. McOmish, P. Balfe, C. A. Ludlam, and A. J. Leigh Brown.** 1991. Discontinuous sequence change in human immunodeficiency virus (HIV) type 1 env gene sequences in plasma viral and lymphocyte-associated proviral populations in vivo: implications for models of HIV pathogenesis. J. Virol. **65:**6266–6276.
26. **St. Clair, M. H., J. L. Martin, G. Tudor-Williams, M. C. Bach, C. L. Vavro, D. M. King, P. Kellam, S. D. Kemp, and B. A. Larder.** 1991. Resistance to ddI and sensitivity to AZT induced by a mutation in HIV-1 reverse transcriptase. Science **253:**1557–1559.
27. **Wahlberg, J., J. Albert, J. Lundeberg, S. Cox, B. Wahren, and M. Uhlén.** 1992. Dynamic changes in HIV-1 quasispecies from azidothymidine (AZT) treated patients. FASEB J. **6:**2843–2847.
28. **Wahlberg, J., J. Albert, J. Lundeberg, A. von Gegerfelt, K. Broliden, G. Utter, E. M. Fenyö, and M. Uhlén.** 1991. Analysis of the V3 loop in neutralization-resistant human immunodeficiency virus type 1 variants by direct solid phase DNA sequencing. AIDS Res. Hum. Retroviruses **7:**983–989.
29. **Wahlberg, J., J. Fioré, G. Angarano, M. Uhlén, and J. Albert.** 1994. Apparent selection against transmission of zidovudine resistant HIV-1 variants. J. Infect. Dis. **169:**611–614.
30. **Wahlberg, J., J. Lundeberg, T. Hultman, and M. Uhlén.** 1990. General colorimetric method for DNA diagnostics allowing direct solid-phase genomic sequencing of the positive samples. Proc.

Natl. Acad. Sci. USA **87:**6569–6573.

31. **Wolfs, T. F. W., G. Zwart, M. Bakker, and J. Goudsmit.** 1992. HIV-1 genomic diversification following sexual and parental virus transmission. Virology **189:**103–110.
32. **Wolfs, T. F. W., G. Zwart, M. Bakker, M. Valk, C. L. Kuiken, and J. Goudsmit.** 1991. Naturally occurring mutations within HIV-1 V3 genomic RNA lead to antigenic variation dependent on a single amino acid substitution. Virology **185:**195–205.
33. **Wolinsky, S. M., C. M. Wike, B. T. M. Korber, C. Hutto, W. P. Parks, L. L. Rosenblum, K. J. Kunstman, M. R. Furtado, and J. L. Munoz.** 1992. Selective transmission of human immunodeficiency virus type-1 variants from mothers to infants. Science **255:**1134–1137.
34. **Zhang, L. Q., P. MacKenzie, A. Cleland, E. C. Holmes, A. J. Leigh Brown, and P. Simmonds.** 1993. Selection for specific sequences in the external envelope protein of human immunodeficiency virus type 1 upon primary infection. J. Virol. **67:**3345–3356.
35. **Zhang, L. Q., P. Simmonds, C. A. Ludlam, and A. J. Leigh Brown.** 1991. Detection, quantification and sequencing of HIV-1 from the plasma of seropositive individuals and from factor VIII concentrates. AIDS **5:**675–681.
36. **Zhu, T., H. Mo, N. Wang, D. S. Nam, Y. Cao, R. A. Koup, and D. D. Ho.** 1993. Genotypic and phenotypic characterization of HIV-1 in patients with primary infection. Science **261:**1179–1181.