VIROLOGY 211, 198–203 (1995)

Sequence Distances between *env* Genes of HIV-1 from Individuals Infected from the Same Source: Implications for the Investigation of Possible Transmission Events

CATHERINE ARNOLD,*,[1] PETER BALFE,† and JONATHAN P. CLEWLEY*

*Molecular Biology Unit, Virus Reference Division, Central Public Health Laboratory, 61 Colindale Avenue, London NW9 5HT, United Kingdom; and †Division of Virology, University College London Medical School, Windeyer Building, 46 Cleveland Street, London W1P 6DB, United Kingdom

Received March 14, 1995; accepted May 30, 1995

Previously described transmission studies have shown that HIV strains isolated from individuals infected from a common source are more homogeneous than HIV strains isolated from individuals with unrelated infections. This has been the basis, in at least four instances, for deciding whether apparently epidemiologically related cases represent actual transmissions. To date, HIV transmission studies have usually included sequence data from the most likely source of infection, and the probability of transmission from the donor to the recipient has been assessed by measuring sequence similarity against control data using likelihood analysis. We have recently studied a putative transmission involving a UK health care worker (CPHL1), a patient of CPHL1 (CPHL2), and CPHL3, a member of the same "sex circle" as CPHL2. We have used sequence distance and neighbour joining methods as well as likelihood analysis as means of determining genetic relatedness. Though no other source of infection was available our findings did not support the possibility that CPHL1 had infected CPHL2. Strain CPHL3 was closer to CPHL2 than to CPHL1. It is shown that control data from documented transmission events can be used to establish the source of infection in the absence of an index case. It is also shown that the C2–V3 region analysed in previous transmission studies is unreliable for accurate phylogenetic analysis. The results indicated that gp120 is a more informative region than C2–V3 for molecular transmission studies. Sequence distances between the *env* genes of related and unrelated infections have been derived in this work. © 1995 Academic Press, Inc.

INTRODUCTION

In May 1992 Ou *et al.* described the use of sequence data to investigate the transmission of HIV-1 from a Florida dentist to several of his patients. They directly sequenced and cloned PCR products from the C2–V5 region of *env* (680 bp) from the dentist, patients, and local controls. They chose this region as it had been used previously in transmission investigations (Balfe *et al.*, 1990; Burger *et al.*, 1991; Wolinsky *et al.*, 1992) and because the Los Alamos HIV sequence database contained a relative abundance of C2–V3 sequences which could be used for comparative purposes. However, the choice of this region of the genome for transmission investigations has been criticised (DeBry, 1993; Holmes *et al.*, 1994; Palca, 1993; Smith, 1992) on the basis that it is too small and too variable for reliable conclusions to be drawn. Also, evolutionary convergence of unrelated sequences can occur in V3 (Holmes *et al.*, 1992) which may confound the identification of linked infections. There has thus been continuing controversy about whether the dentist actually infected his patients, though the most thorough analysis of the data (Hillis and Huelsenback, 1994) has upheld the conclusion that the dentist transmitted HIV-1 to several of his patients.

In 1993 Holmes *et al.* described the use of p17 PCR products (316 bp), produced by a limiting dilution technique, as an alternative to C2–V5 sequencing for determining genetic relatedness between genomes from an HIV-1-infected surgeon and a patient. Unlike the Florida dentist case, material from an alternative source of infection for the patient (an HIV-1-infected blood donor) was available for this investigation. The basis for choosing p17 for analysis was that, unlike V3, this region does not seem to show extensive evolutionary convergence. A possible disadvantage of using a small fragment such as p17 may be that it is not variable enough to be a reliable marker for distinguishing between related and unrelated infections.

Another HIV-1 transmission, from a rapist to his victim, was investigated in 1993 by Albert and colleagues. They chose part of the *pol* gene (640 bp) and used the relatively simple method of direct DNA sequence analysis to show that virus populations harboured by the male and female were highly similar. Although the *pol* gene is relatively conserved they described an unusual amino acid signature pattern shared by the male and female and concluded that the two infections were related. This type of amino acid signature pattern analysis was also used in the Florida dentist investigation (Ou *et al.*, 1992).

The three HIV transmission investigations discussed above all had access to specimens from the likely source

[1] To whom reprint requests should be addressed.

0042-6822/95 $12.00

TABLE 1

Patients Investigated in This Study

Patient	Details	Subtype	Time specimen obtained (years)[a]	Intrapatient distance (%)
CPHL1	Male health care worker	B	6–7?	2.88
CPHL2	Female patient of CPHL1	B	6–7?	1.84
CPHL3	Female member of same sex circle as CPHL2	B	n/k[b]	1.45
CPHL4	Female partner of CPHL5	D	8	3.92
CPHL5	Male partner of CPHL4	D	8	3.15
CPHL6	Male sexual partner of three females (CPHL7–9)	B	2–3	3.76
CPHL7	Female recipient of virus from CPHL6	B	2–3	2.70
CPHL8	Female recipient of virus from CPHL6	B	2–3	6.52
CPHL9	Female recipient of virus from CPHL6	B	2–3	1.83

[a] Approximate time after infection or transmission that the specimen was received in the laboratory.
[b] n/k, not known.

of infection, and the probability of transmission from the donor to the recipient was assessed by measuring sequence similarity using maximum likelihood methods. Recently Jaffe *et al.* (1994) have reported an absence of HIV-1 transmission in a second Florida dental practice. They compared V3 sequence distances, obtained by cloning or direct sequencing of bulk PCR products, from the dentist, the HIV-1 positive patients, and two of the patients' sexual partners. Unlike the three studies described previously, Jaffe *et al.* did not have specimens from the most probable source of infection for the patients. Therefore, their study also involved an analysis of the genetic relatedness among HIV strains together with identification of potential risks for acquisition of HIV infection and an investigation of control measures in the dental practice.

In this paper we describe an investigation of possible transmission involving an HIV-1-infected health care worker (HCW; CPHL1) and a female patient (CPHL2) which does not include any alternative source of infection. We also describe how data from control sequences used in this investigation (transmission from a male to three females) indicate that it may be possible to establish a common source of an infection in a transmission investigation in the absence of the index case.

MATERIALS AND METHODS

Patients studied

The epidemiological investigation of the infected HCW in this case was described by Hochuli and colleagues (Hochuli *et al.*, 1995). Table 1 gives details of the patients studied in this paper. In 1986 patient CPHL2 had undergone three procedures performed by HCW CPHL1 that carried a possible exposure risk. Thus transmission, if it occurred between CPHL1 and CPHL2, took place 6–7 years before specimens were collected from them both. Subsequent to the initial investigation, material became available from a second female (CPHL3) who was a member of the same sex circle as CPHL2 (a sex circle being here defined as a group of people who regularly exchange sexual partners within the group). No alternative source of infection was identified epidemiologically. Controls for this investigation include material from an established male to female transmission, CPHL4 and 5, and a male to three females transmission, CPHL6, 7, 8, and 9.

Preparation of DNA for sequencing[2]

Proviral DNA was prepared from EDTA blood samples by differential lysis or buffy coat lymphocyte purification followed by dilution to single molecules by limiting dilution (Simmonds *et al.*, 1991). Nested PCR was carried out with one PCR primer tagged with biotin in the second round. PCR conditions were 10 m*M* Tris–HCl, pH 8.3, 50 m*M* KCl, 200 μM of each dNTP, 1.5 m*M* $MgCl_2$, 0.5 U *Taq* polymerase, and 5 pmol of each primer. First round primers 989, TCATCAGAACAGTCAGACTCATCAAGC, or 988, GTAGCAATAATAATAGCAATAG, and 633, TCCCACTCCATCCAGGTC, or 631, CCAGACTGTGAGTTGCAACAGATGC (all 5′–3′), were used under the following cycling conditions: 94° for 40 sec, 47° for 35 sec, 72° for 4 min for 30 cycles. Nested PCR was carried out using the same reaction conditions, using inner primers 944, AGAAAGAGCAGAAGACAGTGGCAATG, and 609, CCCATAGTGCTTCCGGCCGCTCCCAAG (biotinylated), under the following cycling conditions: 95° for 40 sec, 55° for

[2] The nucleotide sequence data reported in this article have been deposited with the GenBank database under Accession Nos. U23112–U23138.

35 sec, 72° for 4 min for 35 cycles. The PCR products were purified on streptavidin-coated magnetic beads or by using Centricon-100 columns (Amicon) to remove excess dNTPs and primers (Sambrook *et al.,* 1989). After crude quantitation of amplicons visualised by ethidium bromide staining by comparison with a known standard run on 1% SeaKem agarose gels, both strands were sequenced with a total of 11–15 sequencing primers per PCR product on an Applied Biosystems 373A sequencer, using DyeDeoxy Terminator sequencing kits, according to the manufacturer's instructions. At least three gp120 molecules were sequenced for each individual, with an average sequencing redundancy of 2.4 (i.e., each base was sequenced at least 2.4 times). After assimilation using SeqEd v. 1.03 (Applied Biosystems, Inc.), the data were transferred to Clustal V (Higgins *et al.,* 1991) for multiple sequence alignment, together with other envelope sequences from the Los Alamos database representative of all the currently recognised subtypes (Louwagie *et al.,* 1993). After alignment in Clustal V, the sequences were transferred to the PHYLIP suite of programs for the creation of distance matrix files and phylogenetic analysis (Felsenstein, 1989) and subsequent generation of phylogenetic trees. Distance matrices composed of percentage pairwise differences (or distances) between individual molecules generated by DNADIST from PHYLIP were used to calculate standard deviations from the mean for different groups of sequences.

Analysis

Alignment of the envelope sequences by Clustal V required the insertion of a large number of gaps due to length variation between individual samples. This complicated the interpretation of the phylogenetic relationships between individual species because there is, at present, no consensus on the correct interpretation of such gaps. However, we found that the results of our analyses were broadly similar whether or not gaps were accounted for. Simulation studies suggest that phylogenetic analysis using maximum likelihood inference is an effective way to discover evolutionary relationships (Saitou and Imanishi, 1989) when material from the alternative possible source of infection is available, as it allows a statistical comparison of the relative likelihoods of the different transmission pathways. However, all methods of phylogenetic reconstruction assume that most substitutions are neutral (Holmes, 1993) and it is therefore difficult to assess the performance of the different methods in the presence of natural selection, convergent evolution (Holmes *et al.,* 1992), and the extreme bias in base substitution seen in HIV-1, in particular the strong tendency for G to A changes (Moriyama *et al.,* 1991). The PHYLIP suite of programs was used to construct phylogenetic trees using both maximum likelihood and neighbour joining methods, following sequence distance matrix construction in DNADIST (Felsenstein, 1989). The phylogenetic trees obtained were optimised by global branch swapping. In order to assess the robustness of the tree, bootstrap resampling was performed using 1000 random subsets. The same analyses were repeated using just the C2–V3 region of the envelope gene, previously used in several studies to assess phylogenetic relationships. MegAlign (Lasergene) from DNASTAR, Inc., was also used to construct phylogenetic trees, using Clustal V to align the sequences followed by the neighbour joining method.

RESULTS

Initially our investigation concerned the possible transmission of HIV-1 from a UK health care worker (CPHL1) to a female patient undergoing invasive surgery (Table 1). We sequenced several single gp120 molecules from each individual and compared the pairwise nucleotide distances between sequences. This was found to be 13.84% (±0.59, $n = 24$) (Table 2). The mean pairwise distance of gp120 between unlinked subtype B virus from material received and sequenced in this laboratory was 12.22% (±1.21, $n = 130$) (Table 2). Mean pairwise distance for the gp120 molecules from a subtype D-linked infection (CPHL4 and 5), where blood samples from both individuals were taken as much 8 years post-transmission, was 8.8% (±1.25, $n = 9$) (Table 2). Intrapatient gp120 sequence distances, indicating the degree of quasispecies complexity, are shown in Table 1. The sequence distances between CPHL1 and 2 were also compared with control sequences obtained from the Los Alamos database. As specific controls for the methods used here we sequenced several gp120 molecules from probable heterosexual transmission cases (established by sequence similarity and epidemiological evidence). The first set was a male to female HIV-1 subtype D transmission (CPHL5 to 4) and the second a subtype B transmission from a male (CPHL6) to three females (CPHL7, 8, and 9). The mean pairwise gp120 distance between molecules from the donor and each of the recipients (CPHL6, 7, 8, and 9) was 7.38% (±0.98, $n = 21$) (Table 2). For these cases, blood samples from the donor and the recipients were taken approximately 2–3 years postinfection. The pairwise gp120 distance between the recipients themselves (CPHL7, 8, and 9) was 7.94% (±0.79, $n = 16$).

The mean pairwise distance between gp120 molecules from CPHL2 and CPHL3 was 10.99% (±0.36, $n = 12$) (Table 2). Figure 1 shows the phylogenetic tree obtained from neighbour joining analysis with the CPHL sequences and subtype representatives from the Los Alamos database. Analysis of the same data set using other phylogenetic methods (see Materials and Methods) gave essentially the same results. A thousand bootstrap

TABLE 2

Comparison of Percentage Pairwise Nucleotide and Amino Acid Distances between the *env* Data Sets

	CPHL1/2	CPHL2/3	Unlinked B[a]	Linked B (CPHL6, 7, 8, 9)	Linked D (CPHL4, 5)
Mean	13.84 (22.28)[b]	10.99 (19.55)	12.22 (19.85)	7.38 (14.84)	8.8 (16.67)
SD[c]	0.59 (1.63)	0.36 (0.34)	1.21 (2.34)	0.98 (1.43)	1.25 (1.42)
N	24	12	130	21	9

[a] Unlinked subtype B gp120 sequence comparisons; sequence data from HIV database.
[b] Amino acid sequence distances are in parentheses.
[c] SD, ±standard deviation.
[d] *N*, number of comparisons between molecules.

resamplings were performed for the trees presented here and it was found that the separate branches containing the CPHL1, CPHL2, and CPHL3 gp120 sequences remained in all resamplings. Phylogenetic analysis of amino acid sequences rather than nucleotides gave trees with the same branching order. All *env* nucleotide sequences obtained in this study gave open reading frames (ORFs).

In summary, all methods failed to group the HCW (CPHL1) and patient's (CPHL2) sequences together. HIV sequences from these subjects are more closely related to other sequences than they are to each other. The CPHL2 sequences were found to be more closely related to the other sex circle member (CPHL3) at 10.99% (±0.36, $n = 12$) divergence than they are to any other of the sequences analysed. The analyses were also performed using just the C2–V3 region of the gp120 molecule, which has previously been used in other analyses of transmission events (Ou *et al.*, 1992; Wolfs *et al.*, 1991). The resultant branching order for these reduced data was congruent, though the branch lengths were somewhat smaller overall. Bootstrap resampling of the subset did lead to several alternative topologies which were not seen when the whole data set was analysed. For example, the CPHL3 set of three sequences, which when analysed as full-length genes are separated from the rest of the sequence set (1000/1000 bootstrap resamplings), are only separated as a distinct group in two-thirds of resamplings of the subset (634/1000), joining the CPHL8 data set in the remaining samplings. This result may be due either to the small number of nucleotide differences between the two data sets (10 bases) or to convergent evolution of these regions (Holmes *et al.*, 1992). These data led us to conclude, unsurprisingly, that the 312-bp C2–V3 subregion of gp120 was a less reliable data set than a full gp120 sequence.

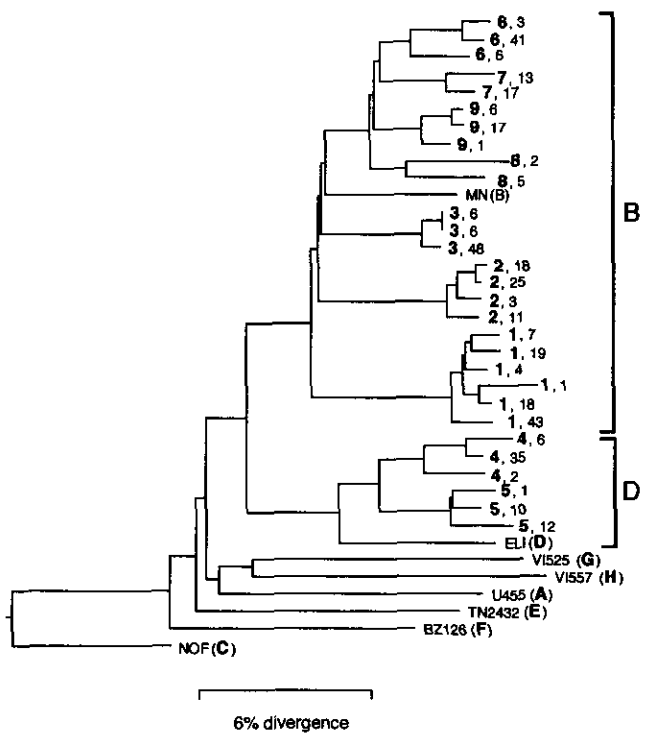

FIG. 1. Phylogenetic tree of CPHL and database gp120 sequences constructed using the neighbour joining method as implemented in Lasergene (DNAStar). The HIV-1 subtype is indicated by square brackets or in parentheses next to the strain name. Numbers in bold correspond to the CPHL patient numbers (see Table 1 and the text) followed by the individual molecule number. Trees of the same topology were obtained using distance and likelihood methods in the PHYLIP suite of programs (see Materials and Methods).

DISCUSSION

There are four important questions to answer before embarking on an investigation of HIV-1 transmission. These are: (i) which region of the genome to sequence? (ii) how should the phylogenetic analysis be carried out? (iii) what is the appropriate background population? and (iv) should the HIV variants be separated prior to amplification (i.e., should one sequence cloned/single molecules rather than bulk PCR products)? Different regions of the genome have been used to determine sequence relatedness between HIV-1 variants from proposed linked infections, including *env, gag,* and *pol* (Albert *et al.*, 1994; Holmes *et al.*, 1993; Jaffe *et al.*, 1994; Ou *et al.*, 1992). Most is known about the V3 loop of *env* as it has

been extensively studied and consequently has been the region of choice for several investigations. However, the *V3* loop is under considerable immune pressure and evolves rapidly, and this may confuse the determination of epidemiological relationships. If, on the other hand, *gag* or *pol* is used instead of V3 for determining sequence relatedness, it is possible that they may not show sufficient variation when viewed against an extensive background sequence population. Estimates of the rate of nucleotide sequence change per year are 1% in *env* and 0.5% in *gag* (Myers and Korber, 1993; Myers and Pavlakis, 1992). The importance of not obtaining a single direct consensus from a bulk PCR was demonstrated in a recent study showing that in some cases a minor HIV-1 variant was transmitted between sexual partners (Zhu *et al.*, 1993). If a direct consensus or bulk sequence from PCR is obtained and not single molecules, it is possible that minor variants could be missed.

In light of the above considerations, and in an attempt to avoid the controversies surrounding the investigation of the Florida dentist, we chose to sequence several single molecules of gp120 from each individual in our investigation of a possible transmission case involving a HCW and patient. The gp120 area of the HIV-1 genome provides both variable and conserved regions. By sequencing gp120 we could extract the 312-bp C2–V3 subregion and, using the same analysis, determine whether it was a less reliable data set than a full gp120 sequence. Theoretically, if transmission had occurred between CPHL1 and CPHL2, pairwise gp120 sequence variation from both might be expected to be around 8%. Mean pairwise differences between gp120 sequences from CPHL1 and CPHL2 were found to be 13.84% (±0.59) (Table 2), as great or greater than differences between unrelated subtype B sequences. All of the three phylogenetic methods we used (distance methods, neighbour joining, and maximum likelihood) placed unrelated sequences closer to CPHL2 than CPHL1. This was true for both nucleotide and amino acid sequence analysis.

As CPHL2 and CPHL3 were members of the same sex circle, it is likely that their infections were linked, though there may not have been direct transmission. However, the routes and possible times of transmission of HIV-1 within this sex circle are unknown, and as sequences from other members of the circle are unavailable, it cannot be definitively established whether the CPHL2 and 3 infections are linked. The mean pairwise sequence distance between single molecules of gp120 from CPHL2 and CPHL3 was 10.99% (±0.36) (Table 2). This may indicate that the sequences are related and is consistent with the virus being transmitted through more than one individual.

Results from some of the control data used in this study (CPHL6 transmission to CPHL7, 8, and 9) show that mean pairwise sequence distances of individual gp120 molecules between recipients, at 7.94% (±0.79), is only marginally higher than that between donor and recipients (7.38% ± 0.98) (Table 2). Material obtained for analysis from this group was acquired approximately 2–3 years after infection in all three cases. All had developed advanced disease with $CD4^+$ counts below 200 within this time. These data show that it may be possible to investigate transmission events when material is not available from the index case if there is more than one presumed recipient. Therefore, if transmission to the recipients takes place at a similar time, and material for analysis is obtained relatively soon after infection (i.e., 2–3 years), it is possible that the HIV variants will be similar enough to establish a common source.

The results from the subtype D-linked infection (CPHL5 to CPHL4), studied as a non-subtype B transmission control, show that although material for analysis was obtained from the two individuals at an estimated 8 years after transmission, it was still possible to establish some similarity between their gp120 sequences. However, at 8.8% (±1.25), the mean pairwise sequence distance is approaching that between unrelated infections. Analysis is more complicated than for subtype B infections as there are fewer subtype D sequence data available to provide a background population for comparative purposes.

The results from the analysis of the C2–V3 region (studied by Ou *et al.*, 1992, in the Florida dentist investigation) compared with the complete gp120 of our data set showed that the branches of the two trees generated were identical when gaps in the alignment were omitted, but were discordant when gaps were included. Using our data, the 312-bp amplicon analysed by Ou and colleagues has too few nucleotides for accurate phylogenetic analysis to be carried out as we observed that one base change in a sequence reordered the tree. These results indicated that gp120 is a more informative region to study than C2–V3 for molecular transmission studies.

ACKNOWLEDGMENTS

We thank Dr. Jane McKeating (Microbiology Department, Reading University) and Dr. John Parry (CPHL), for advice and constructive criticism, and Dr. Eddie Holmes (Oxford University) for access to computing facilities. We also thank the physicians who supplied the patient specimens used in this study.

REFERENCES

Albert, J., Wahlberg, J., Leitner, T., Escanilla, D., and Uhlen, M. (1994). Analysis of a rape case by direct sequencing of the human immunodeficiency virus type 1 *pol* and *gag* genes. *J. Virol.* **68,** 5918–5924.

Balfe, P., Simmonds, P., Ludlam, C. A., Bishop, J. O., and Brown, A. J. (1990). Concurrent evolution of human immunodeficiency virus type 1 in patients infected from the same source: Rate of sequence change and low frequency of inactivating mutations. *J. Virol.* **64,** 6221–6233.

Burger, H., Weiser, B., Flaherty, K., Gulla, J., Nguyen, P. N., and Gibbs,

R. A. (1991). Evolution of human immunodeficiency virus type 1 nucleotide sequence diversity among close contacts. *Proc. Natl. Acad. Sci. USA* **88,** 11236–11240.

DeBry, R. W. (1993). Dental HIV transmission? *Nature* **361,** 691.

Felsenstein, J. (1989). PHYLIP — Phylogeny inference package. *Cladistics* **5,** 164–166.

Higgins, D. G., Bleasby, A. J., and Fuchs, R. (1991). Clustal V: Improved software for multiple sequence alignment. *Comput. Appl. Biosci.* **8,** 189–191.

Hillis, D. M., and Huelsenback, J. P. (1994). Support for dental HIV transmission. *Nature* **369,** 24–25.

Hochuli, V., Hyndman, O., and Porter, K. (1995). Response to news that an obstetrician/gynaecologist has AIDS. *Communicable Dis. Rep.* **5,** 7–11.

Holmes, E. C. (1993). The pattern and process of base substitution in immunodeficiency viruses. *Binary* **5,** 189–190.

Holmes, E. C., Zhang, L. Q., Robertson, P., Cleland, A., Harvey, E., Simmonds, P., and Leigh-Brown, A. J. (1994). The molecular epidemiology of human immunodeficiency virus type 1 in Edinburgh. *J. Infect. Dis.* **171,** 45–53.

Holmes, E. C., Zhang, L. Q., Simmonds, P., Ludlam, C. A., and Brown, A. J. (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.

Holmes, E. C., Zhang, L. Q., Simmonds, P., Rogers, A. S., and Brown, A. J. (1993). Molecular investigation of human immunodeficiency virus (HIV) infection in a patient of an HIV-infected surgeon. *J. Infect. Dis.* **167,** 1411–1444.

Jaffe, H. W., McCurdy, J. M., Kalish, M. L., Liberti, T., Metellus, G., Bowman, B. H., Richards, S. B., Neasman, A. R., and Witte, J. J. (1994). Lack of HIV transmission in the practice of a dentist with AIDS. *Ann. Intern. Med.* **121,** 855–859.

Louwagie, J., McCutchan, F. E., Peeters, M., Brennan, T. P., Sanders-Buell, E., Eddy, G. A., van der Groen, G., Fransen, K., Gershy-Damet, G. M., Deleys, R., and Burke, D. S. (1993). Phylogenetic analysis of gag genes from 70 international HIV-1 isolates provides evidence for multiple genotypes. *AIDS* **7,** 769–780.

Moriyama, E. N., Ina, Y., Ikeo, K., Shimizu, N., and Gojobori, T. (1991). Mutation pattern of human immunodeficiency virus genes. *J. Mol. Evol.* **32,** 360–363.

Myers, G., and Korber, B. (1993). The future of HIV. *In* "Evolutionary Biology of Viruses." Raven Press, New York.

Myers, G., and Pavlakis, G. N. (1992). Evolutionary potential of complex retroviruses. *In* "The Retroviridae," pp. 51–104. Plenum, New York.

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., Witte, J. J., Furman, L. J., Satten, G. A., Macinnes, K. A., Curran, J. W., and Jaffe, H. W. (1992). Molecular epidemiology of HIV transmission in a dental practice. *Science* **256,** 1165–1171.

Palca, J. (1993). CDC closes the case of the Florida dentist. *Science* **255,** 766.

Saitou, N., and Imanishi, T. (1989). Relative efficiencies of the Fitch-Margoliash, maximum parsimony, maximum likelihood, minimum evolution and neighbor-joining methods of phylogenetic tree construction in obtaining the correct tree. *Mol. Biol. Evol.* **6,** 514–525.

Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989). "Molecular Cloning: A Laboratory Manual." Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Simmonds, P., Zhang, L. Q., McOmish, F., Balfe, P., Ludlam, C. A., and Brown, A. J. (1991). Discontinuous sequence change of human immunodeficiency virus (HIV) type 1 env sequences in plasma viral and lymphocyte-associated proviral populations in vivo: Implications for models of HIV pathogenesis. *J. Virol.* **65,** 6266–6276.

Smith, T. F., and Waterman, M. S. (1992). The continuing case of the Florida dentist. *Science* **256,** 1155–1156.

Wolfs, T. F., Zwart, G., Bakker, M., Valk, M., Kuiken, C. L., and Goudsmit, J. (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.

Wolinsky, S. M., Wike, C. M., Korber, B. T., Hutto, C., Parks, W. P., Rosenblum, L. L., Kunstman, K. J., Furtado, M. R., and Munoz, J. L. (1992). Selective transmission of human immunodeficiency virus type-1 variants from mothers to infants. *Science* **255,** 1134–1137.

Zhu, T., Mo, H., Wang, N., Nam, D. S., Cao, Y., Koup, R. A., and Ho, D. D. (1993). Genotypic and phenotypic characterization of HIV-1 patients with primary infection. *Science* **261,** 1179–1181.