Molecular Epidemiology of HIV Transmission in a Dental Practice

Author(s): Chin-Yih Ou, Carol A. Ciesielski, Gerald Myers, Claudiu I. Bandea, Chi-Cheng Luo, Bette T. M. Korber, James I. Mullins, Gerald Schochetman, Ruth L. Berkelman, A. Nikki Economou, John J. Witte, Lawrence J. Furman, Glen A. Satten, Kersti A. MacInnes, James W. Curran, Harold W. Jaffe, Laboratory Investigation Group and Epidemiologic Investigation Group

Geophys. Res. Lett. **16**, 703 (1989).

71. D. Kley *et al.*, in *Global and Regional Environmental Atmospheric Chemistry: Proceedings of the International Conference on Global and Regional Environmental Atmospheric Chemistry* (Department of Energy Conference 890525, 1989, available from the National Technical Information Service, Springfield, VA), p. 9.

72. D. Rind, R. Suozzo, N. K. Balanchandran, M. J. Prather, *J. Atmos. Sci.* **47**, 475 (1990).

73. K. B. Hogan, J. S. Hoffman, A. M. Thompson, *Nature* **354**, 181 (1991); A. M. Thompson, K. B. Hogan, J. S. Hoffman, *Atmos. Environ.*, in press.

74. C. S. Atherton and J. E. Penner, *J. Geophys. Res.* **95**, 14027 (1990).

75. I. S. A. Isaksen, T. Berntsen, S. Solberg, in *Ozone in the Atmosphere*, R. D. Bojkov and P. Fabian, Eds. (A. Deepak, Hampton, VA, 1989), pp. 576–579.

76. M. J. Prather, Ed., "An assessment model for atmospheric composition" (*NASA Conf. Publ. 3023*, NASA, Washington, DC, 1989).

77. S. Madronich and C. Granier, *Geophys. Res. Lett.* **19**, 465 (1992).

78. A. M. Thompson, *J. Geophys. Res.* **89**, 1341 (1984); S. Madronich, *ibid.* **92**, 9740 (1987).

79. S. A. Penkett and K. A. Brice, *ibid.* **319**, 655 (1986).

80. J. A. Logan, *J. Geophys. Res.* **90**, 10463 (1985). An update for the North American data is in J. A. Logan, in *Tropospheric Ozone*, I. S. A. Isaksen, Ed. (North Atlantic Treaty Organization Advanced Study Institutes, Reidel, Dordrecht, Holland, 1988), pp. 327–344.

81. Locations of sites: Barrow (71°N, 157°W); Mauna Loa (19.5°N, 156°W); Samoa (14°S, 171°W); South Pole (90°S, 0). See (*36*); S. J. Oltmans and W. D. Komhyr, *J. Geophys. Res.* **91**, 5229 (1986).

82. K. M. Valentin, thesis, Johannes-Gutenburg University, Mainz (1990).

83. P. J. Crutzen and P. H. Zimmermann, *Tellus Ser. B* **43**, 136 (1991).

84. The following people have been generous with reports of recent work: C. Bruhl, J. Fishman, I. S. A. Isaksen, M. Kanakidou, M. A. K. Khalil, K. Law, Y. Lu, A. Neftel, J. Pinto, M. Prather, and R. Prinn. M. Kanakidou, J. Logan, and W. Stockwell made helpful comments on the manuscript. Some of the model comparisons (Tables 1 and 2) grew out of the United Nations Environmental Programme–World Meteorological Organization Ozone Assessment Group Workshop (London, June 1991) and a North Atlantic Treaty Organization Advanced Research Workshop on Methane (Mt. Hood, Oregon, October 1991). My research on oxidizing changes is supported by NASA Programs in Earth Observing Systems and Tropospheric Chemistry and by the U.S. Environmental Protection Agency.

Molecular Epidemiology of HIV Transmission in a Dental Practice

Chin-Yih Ou, Carol A. Ciesielski, Gerald Myers, Claudiu I. Bandea, Chi-Cheng Luo, Bette T. M. Korber, James I. Mullins, Gerald Schochetman, Ruth L. Berkelman, A. Nikki Economou, John J. Witte, Lawrence J. Furman, Glen A. Satten, Kersti A. MacInnes, James W. Curran, Harold W. Jaffe, Laboratory Investigation Group,* Epidemiologic Investigation Group†

Human immunodeficiency virus type 1 (HIV-1) transmission from infected patients to health-care workers has been well documented, but transmission from an infected health-care worker to a patient has not been reported. After identification of an acquired immunodeficiency syndrome (AIDS) patient who had no known risk factors for HIV infection but who had undergone an invasive procedure performed by a dentist with AIDS, six other patients of this dentist were found to be HIV-infected. Molecular biologic studies were conducted to complement the epidemiologic investigation. Portions of the HIV proviral envelope gene from each of the seven patients, the dentist, and 35 HIV-infected persons from the local geographic area were amplified by polymerase chain reaction and sequenced. Three separate comparative genetic analyses—genetic distance measurements, phylogenetic tree analysis, and amino acid signature pattern analysis—showed that the viruses from the dentist and five dental patients were closely related. These data, together with the epidemiologic investigation, indicated that these patients became infected with HIV while receiving care from a dentist with AIDS.

Increasingly, molecular biologic techniques have been used to study the epidemiology of infectious diseases. For viral infections of humans, techniques to analyze viral genetic sequence information, such as oligonucleotide fingerprinting of RNA genomes with ribonuclease, mapping of DNA genomes with restriction endonucleases, and genomic sequencing, have been used to study viral transmissions from person to person, within communities, and between countries (*1*). Requisite to such studies is the existence of viral genetic variation; the greater the variation, the greater the power of the methods to distinguish strains of the virus. For a virus with substantial genomic variation, identification of strains with a high degree of genetic relatedness may imply an epidemiologic linkage between persons infected with these strains.

The human immunodeficiency virus (HIV) has a high mutation rate (*2, 3*), such that HIVs from different individuals are found to be genetically distinct (*3*). In this article, we describe the use of genomic sequencing to investigate a cluster of HIV infections in a Florida dental practice. The high degree of genetic relatedness observed among the HIV strains from a dentist with acquired immunodeficiency syndrome (AIDS) and five of his infected patients supports the epidemiologic investigation that indicated that these patients became infected with HIV while receiving dental care.

Epidemiologic Investigation

In July 1990, we reported that a young woman with AIDS (patient A) had most likely acquired her HIV-1 infection while undergoing invasive dental procedures by a Florida dentist with AIDS (*4*). Following publication of the report, the dentist publicly requested that his former patients be tested for HIV infection. Among approximately 1100 persons whose blood was tested by the Florida Department of Health and Rehabilitative Services (HRS), two patients (patients B and C) were found to be HIV-positive. An additional infected patient (patient D) was ascertained by HRS through cross matching a list of the dentist's former patients with the Florida AIDS case registry. Two other patients of the dentist (patients E and G) contacted the Centers for Disease Control (CDC) to report that they were HIV-infected. A former sex partner named by patient E was found to be HIV-infected and had also been a patient of the dentist (patient F). Characteristics of these seven infected patients and the dentist are included in Table 1.

Patient D had previously been reported

1165

to HRS as having behavioral risks for HIV infection. To establish any risk factors for exposure to HIV for the remaining six patients, follow-up investigations were conducted (5). These investigations included interviews with the patients and their families and friends; review of medical, dental, and health department records; interviews with health-care providers; and testing of the patients' sex partners for HIV infection. Interviews and medical records established that patient F also had behavioral risk factors for HIV infection. The other five patients (patients A, B, C, E, and G) denied injecting drug use since 1978 and had no history of transfusion or receipt of blood products; the two male patients (patients C and G) denied having had sex with men. The possibility that patient C had engaged in high-risk behaviors was raised during the epidemiologic investigation; however, behavioral exposures to HIV could not be documented.

Only one of the known sex partners of these five patients tested positive for HIV infection. This person (patient F) had been an infrequent sex partner of patient E from 1987 until the fall of 1988. Patient E was first tested for HIV infection in October 1988 and was seropositive; patient F had been tested in October and December 1988 and was seronegative. In September 1989, approximately 1 year after his last sexual contact with patient E and his last reported dental visit, patient F had an illness compatible with an acute retroviral infection; he first tested positive for HIV in December 1990. Thus, patients E and F appeared to have contracted their HIV infections from different sources.

The dentist was diagnosed with symptomatic HIV infection in late 1986 and with AIDS in September 1987, when a biopsy of his palate showed Kaposi's sarcoma and his CD4$^+$ lymphocyte count was less than 200 per microliter. He continued to practice general dentistry for approximately 2 years after he was diagnosed with AIDS. During this 2-year period, each of the five patients without confirmed exposures to HIV made multiple visits to the dental office for invasive procedures, including extractions or root canal therapy.

HIV Genetic Variation

Because the epidemiologic findings suggested that transmission may have occurred in the dental practice, we conducted a laboratory investigation to determine whether there was evidence of genetic similarity among the viruses infecting the dentist and his patients. HIV is known to undergo genetic change or variation during its life cycle resulting in a myriad of related viral progenies (commonly called quasi-species)

(6). Several factors, including duration of infection (7), host immune pressure (8), disease stage (9), and therapy (10), may contribute to the degree and rate of HIV genetic variation. HIV strains from persons with a known common infection link, for example, sex partners (11), mothers and their infants (11, 12), blood donors and recipients (13), and hemophilic men receiving common lots of contaminated blood products (2, 14), have been shown to exhibit a closer genetic relatedness than HIV strains from persons without a direct transmission link. Thus, we used the degree of genetic similarity, in combination with epidemiologic information, to evaluate possible HIV transmission linkage.

Analysis of HIV-1 genetic variation involves comparison of multiple sequences covering a region of the viral genome. Among the structural genes of HIV, the envelope gene (env) exhibits the highest degree of genetic diversity. The distribution of genetic diversity in env is not uniform as evidenced by the presence of interspersed conserved (C) and variable (V) domains in the external glycoprotein gp120. We focused most of our analyses on the C2 and V3 domains because (i) these domains contain nucleotide sequences with sufficient variability to distinguish between strains, a feature essential for establishing HIV transmission linkage, (ii) C2-V3 sequences have been used in previous studies of epidemiologically linked infections (2, 11, 12, 14), and (iii) the HIV Sequence Database (3) contains a relative abundance of C2-V3 sequences for comparative purposes.

Blood specimens from the dentist, patients A through G, and 35 local HIV-seropositive persons (local controls or LCs)

were collected between March 1990 and April 1991. Local control specimens were collected from two HIV clinics located within 90 miles of the dental practice. Of the 35 LCs, 7 did not have symptoms related to their HIV infection, 11 were symptomatic but had not developed AIDS, and 17 had AIDS.

DNA fragments of approximately 680 base pairs containing the C2, V3, V4, C3, and V5 domains of the gp120 gene were amplified by a two-step polymerase chain reaction (PCR) procedure (15, 16) on DNA from peripheral blood mononuclear cells (PBMCs) of the dentist and the seven patients. The resulting amplified DNA products were sequenced directly or cloned into an M13 vector and sequenced (15). Care was taken to prevent DNA carryover during the PCR procedure (15, 17), a problem that has been previously reported (18, 19). We also took measures to facilitate the management and identification of the source of specimens (15). The initial results were verified by repeating the PCR amplification procedure with a second vial of PBMCs from either the first blood collection (the dentist and patients D and E) or a second blood collection (patients A, B, C, F, and G); products from the second round of amplification were sequenced directly (15). In addition, HIV sequences from patients A and B and one LC were independently verified by another laboratory (20).

The Dental Group

To assess the genetic relatedness between the HIV strains infecting the dentist and his seven patients, multiple C2-V3 sequences (Table 1) as well as V4-C3-V5

Table 1. Dental cohort clinical information and HIV nucleotide variation in the C2-V3 domain of the envelope gene.

Person	Sex	Known risk factor	Clinical status*	M13 clones (no.)	Intraperson variation† (%)	Interperson variation† (%)		
						To dentist		To 30 LCs‡
Dentist	M	Yes	AIDS	6	3.3 (0.8–5.4)			11.0 (5.8–16.0)
Patient A	F	No	AIDS	6	2.0 (0.0–4.5)	3.4	(0.8–6.2)	10.9 (5.4–14.8)
Patient B	F	No	Asymptomatic (CD4 = 222/μl)	12	1.9 (0.4–3.7)	4.4	(2.1–7.0)	11.2 (6.2–16.5)
Patient C	M	No§	Asymptomatic (CD4 = <50/μl)	5	1.2 (0.4–1.6)	3.4	(2.1–4.9)	11.1 (7.0–15.6)
Patient E	F	No	Asymptomatic (CD4 = 567/μl)	6	2.1 (0.4–3.7)	3.4	(1.2–6.6)	10.8 (5.8–14.8)
Patient G	M	No	Asymptomatic (CD4 = 400/μl)	5	2.8 (1.6–3.7)	4.9	(2.9–7.0)	11.8 (6.2–16.9)
Patient D	M	Yes	AIDS	5	7.5 (0.0–9.9)	13.6	(11.5–15.6)	13.1 (7.8–17.3)
Patient F	M	Yes	Asymptomatic (CD4 = 253/μl)	6	3.0 (0.8–5.8)	10.7	(8.2–13.6)	11.9 (7.0–17.3)

*Clinical status at the time of specimen collection. CD4 is the CD4$^+$ lymphocyte count   †Sequence alignment over 243 nucleotide positions in C2-V3 was obtained with the program PIMA (37) followed by manual refinement with MASE (3). Subsequently, all M13 clone sequences were compared to each other and to direct sequences (15) of PCR-amplified products using the SIMILARITY function of MASE. Only single nucleotide differences are scored; gaps were not scored. The average percentages of differences are shown with the range of values in parentheses.   ‡Sequences from five of the LCs were shorter than 243 nucleotides and were not included in this analysis.   §Possible risk factors for HIV infection were suggested but not documented.

sequences from the dentist and the seven dental patients were compared. Five to 12 C2-V3 sequences (containing a portion of the C2 and all of the V3 region) and several V4-C3-V5 sequences from independent M13 clones were obtained from each of the seven dental patients. We defined the average or genetic distance between the viruses of one individual and those of another individual or set of individuals as the average percentage of sequence divergence (excluding positions with gaps) of all available pairs of C2-V3 nucleotide sequences. Genetic distances between the viruses of the dentist and the patients (21), as well as the ranges of differences, are shown in Table 1. For the patients A, B, C, E, and G, the average distance to the dentist's viruses clustered in the range of 3.4 to 4.9%, distances comparable to those reported for known epidemiologically linked infections (2, 11, 12). In contrast, viruses present in patients D and F were more distantly related to those of the dentist, 13.6% and 10.7%, respectively. These latter distances are typical of equivalently measured distances for viruses taken from epidemiologically unlinked infections and are consistent with the epidemiologic finding that patients D and F engaged in high-risk behaviors that could have resulted in their HIV infections. These and other laboratory results (Fig. 1 and Table 2), in conjunction with the epidemiologic findings, led to the decision to exclude patients D and F from the cluster of linked infections in the statistical analyses presented below.

When the percentages of sequence divergence for viral sequences from the same person were calculated, the averages found in the viruses of the dentist and each of the patients A, B, C, E, and G were low (0.0 to 5.4%). These data are consistent with previous reports of intraperson genetic variation (2, 3, 6), indicating that none of the five patients was infected with HIV from more than one source.

Although no case of a stable form of HIV has been reported (3), the existence of a predominant HIV-1 variant in this part of Florida could not be presumptively ruled out. The existence of such a variant could potentially explain the finding of highly related HIV sequences among patients A, B, C, E, and G, independent of their contact with the dentist. Therefore, we determined the HIV C2-V3 sequences of the 35 HIV-seropositive LCs to examine the extent of genetic variation in the local geographic area. The sequences for 33 of the LCs were determined directly from the amplified products (15), which permitted the determination of the most common residue at each position (11, 22). In addition, the C2-V3 sequences from 7 of the 35 LCs, including the two for which direct

sequences were not available, were determined from their respective M13 clones. Thirty of the 35 C2-V3 sequences (both direct and clone) from the LCs were longer than 243 nucleotides and were included in the genetic distance comparison (Table 1). In calculating the genetic distances in Table 1, we used direct sequences only when clone sequences were not available. The five other LCs were not included because of their slightly shorter nucleotide sequence lengths.

As shown in Table 1, average distances between the viruses of dental patients A, B, C, E, and G and those of the LCs ranged from 10.8 to 11.8%. These distances are in accord with the average interpatient distance found among the 30 LCs, namely 12.0%. In addition, the average distances between the dentist's viruses and those of the 30 LCs is 11.0%, whereas the average distance between the dentist's viruses and those of patients A, B, C, E, and G is 4.0%.

The possibility that the HIV sequences in patients A, B, C, E, and G were closer in their genetic distance to the dentist's viruses than to the LC's viruses as a result of chance alone was tested using the Wilcoxon rank-sum statistic. The test statistic was significant ($P = 6 \times 10^{-6}$), suggesting

that this cluster of related viruses in the dentist and patients A, B, C, E, and G is not due to chance (23). This cluster cannot be explained by the existence of a stable viral variant within the geographic area. Other data from the V4-C3-V5 region are consistent with our analysis of the C2-V3 region sequence relationships.

Phylogenetic Tree Analysis

The pairwise distance measurements reported in Table 1 do not provide information about either the relationship of the viruses from the infected dental patients to each other or the viral genetic distances from the patients to individual LCs. Furthermore, these measurements do not delineate the proximal source of the viruses infecting the individual dental patients. To more closely examine the full range of relationships between the viruses of the dentist, the patients, and the LCs, we subjected the C2-V3 nucleotide sequence data set to phylogenetic tree analyses. These cluster analyses were intended to be both "exploratory" and, within the limitations of the sequence lengths, "statistical" (24).

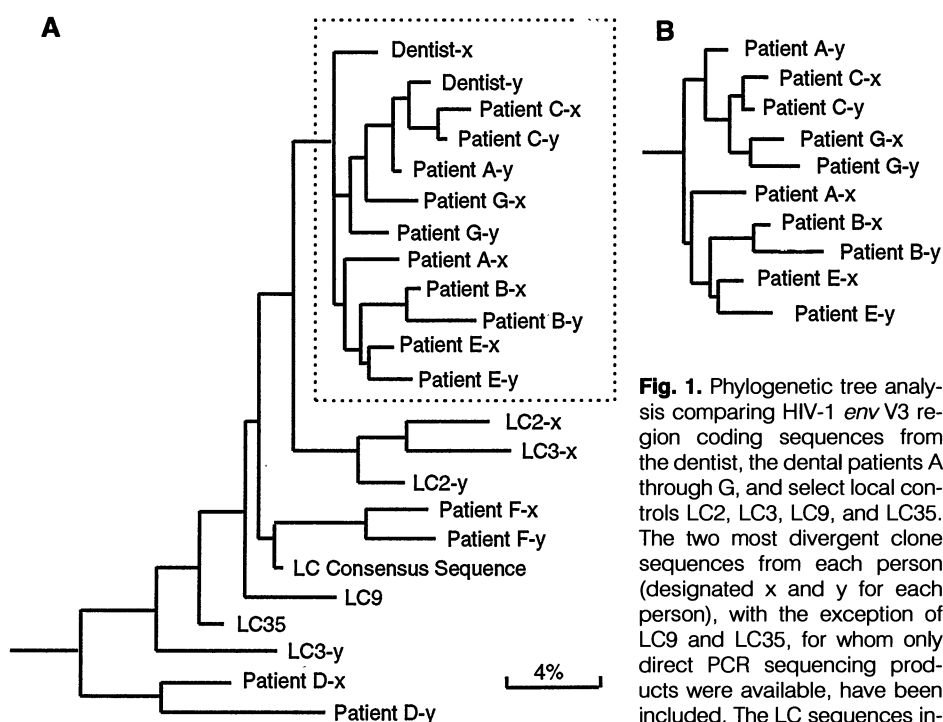A representative tree comprising the dental group and select LC sequences is

Fig. 1. Phylogenetic tree analysis comparing HIV-1 *env* V3 region coding sequences from the dentist, the dental patients A through G, and select local controls LC2, LC3, LC9, and LC35. The two most divergent clone sequences from each person (designated x and y for each person), with the exception of LC9 and LC35, for whom only direct PCR sequencing products were available, have been included. The LC sequences included were those found by pairwise distance measurement (Table 1) and signature pattern analysis (Table 2) to be the closest control sequences to the dental group sequences, which are enclosed by a box in (**A**). An LC consensus sequence has also been included, and the tree was rooted upon the African sample ELI. The PAUP parsimony algorithm was used to analyze 279 aligned sites (25), of which 146 sites were varied. When the dentist's viral sequences were withdrawn from the analysis, or required to cluster with the LC consensus sequence (25), the dental clade remained otherwise unaffected, as shown in (**B**). Vertical distances are for clarity only; the lengths of the horizontal branches are proportional to the single base changes and can be read as percentage differences with the scale bar.

Table 2. Frequencies of amino acids that define a signature pattern in the dentist's viruses.

Group	Frequencies							
	*Reference set**							
	E	T	E	S	T	A	I	Q
Reference set	0.63	0.81	0.69	0.59	0.69	0.66	0.72	0.56
Florida LCs	0.67	0.67	0.63	0.75	0.42	0.60	0.84	0.45
Dentist's viruses	0	0	0	0	0	0	0	0
A, B, C, E, G viruses	0	0	0	0	0	0.20	0	0
D and F viruses	0.55	1.0	0.73	0.91	0	0.45	0.82	0.82
	Dentist's signature†							
	A	I	A	G	A	E	V	H
Reference set	0.06	0.13	0.06	0.16	0.25	0.06	0.25	0.16
Florida LCs	0.15	0.23	0.04	0.15	0.58	0.06	0.14	0.28
Dentist's viruses	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
A, B, C, E, G viruses	1.0	0.94	0.97	1.0	1.0	0.69	1.0	1.0
D and F viruses	0.36	0	0	0.09	1.0	0	0	0

*Thirty-two distinct V3 region sequences available from the HIV Sequence Database (*3*). Those positions at which the dentist's viruses differed from that amino acid found in 50% or more of the reference set sequences constitute the dentist's signature pattern. Eight noncontiguous amino acids were objectively identified (see Fig. 2 for positions involved).   †The dentist's signature pattern is found by definition in all six cloned sequences of the dentist's viruses. Signature patterns were also determined for patients A, B, C, E, and G. Signature similarity: dentist, 6 clones with 8/8 amino acids identical with signature pattern; patient A, 6 clones with 8/8; patient B, 11 clones with 8/8 and 1 clone with 7/8; patient C, 4 clones with 8/8 and 1 clone with 7/8; patient E, 4 clones with 8/8 and 2 clones with 7/8; patient G, 5 clones with 8/8; patient D, 4 clones with 2/8 and 1 clone with 1/8; and patient F, 1 clone with 2/8 and 5 clones with 1/8.

shown in Fig. 1. A parsimony algorithm was used to generate this tree (*25*). Due to the large number of LCs, not all clone sequences available in this study could be satisfactorily represented in a single tree. Therefore, we focused the analysis on trees constituted from the two most divergent sequences (when available) from each of the samples and included only those LC sequences that by distance measurement (Table 1) or signature pattern analysis (Table 2) were found to be most closely related to the dental group sequences. In addition, a consensus sequence derived from all of the LC sequences (LC consensus sequence) was included. Many equally parsimonious trees were found; however, the monophyletic nest containing the sequences of the dentist and of patients A, B, C, E, and G was common to all of the most parsimonious trees. This nest, indicated by the boxed area in Fig. 1A, will be referred to as the dental clade.

Other trees were constructed by (i) using consensus sequences; (ii) rearranging the order in which sequences were presented to the computer program; (iii) including all LC sequences, other U.S.–derived sequences, and V4-C3-V5 sequences; (iv) using only nucleotides representing the first and second base positions in codons; (v) removing 64 noninformative sites (of the total 146 varied sites) (*26*); and (vi) successively removing and reintroducing the patient and LC sequences. All of the most parsimonious trees generated in these different ways retained the dental clade shown in Fig. 1; the viral sequences from patients A, B, C, E, and G invariably nested with the dentist's viral sequences. When the dentist's viral sequences were removed from the analysis, or required to cluster with the

LC consensus sequence in a "user-defined" tree (*25*), the remaining portion of the dental clade remained intact (Fig. 1B). Thus, tree analyses were consistent with the genetic distance analysis summarized in Table 1, and the tree analyses defined a strong linkage among the sequences from patients A, B, C, E, and G, independent of the presence or absence of the dentist's viral sequences. However, tree analyses could not specify the order, that is, the direction, of transmission within the dental group.

It is prudent with cluster analyses involving a relatively small number of phylogenetically informative sites to subject the cluster analyses to statistical bootstrap analysis (*27*). In each of 100 iterations of the bootstrap analysis, nucleotide sites were randomly selected (with replacement) from the 146 sites that had been used as the basis for construction of the phylogenetic tree shown in Fig. 1A. A new tree was then constructed for each iteration. By this procedure, the monophyletic grouping of the dental sequences was observed in 79 of 100 replicates (*26*). The value 79 of 100 should not be construed as a *P* value for a statistical test of the hypothesis that the dental clade forms a monophyletic nest; results of such tests (for instance, the genetic distance analyses that use rank-sum statistics and the signature pattern analysis) are presented below. Instead, the results should be interpreted as an indication of how often we would conclude, from randomly selected nucleotide sites, that the dental clade was a distinct cluster. This procedure provides some measure of the power of the tree analysis, although exact interpretation of results from bootstrap analysis of parsimony trees is a subject of controversy (*28*).

Signature Pattern Analyses

Early in the course of this investigation, we noted the shared occurrence of a unique amino acid pattern in some clones of the dentist's and patient A's viruses (*4*). Although this particular pattern was not observed in all of the clones from these two individuals, nor in sequences from the other dental patients, it encouraged us to undertake a third measure of genetic similarity based upon phylogenetically informative characters. This approach will be referred to as amino acid signature pattern analysis.

A unique signature pattern of eight residues was detected within the 120–amino acid C2-V3 region of the dentist's viruses by comparing an alignment of his viral amino acid sequences to an alignment of a reference set consisting of 26 North American, 5 Haitian, and 1 European HIV-1 sequences (*29*). The resulting pattern consisted of all sites in which all six clones from the dentist differed from what is found in over 50% of sequences in the reference set (Table 2 and marked with asterisks in Fig. 2; although consensus sequences are shown for illustrative purposes, every available clone sequence was used in this analysis). The dentist's signature pattern consisted of the following eight noncontiguous amino acids: A, I, A, G, A, E, V, and H. In contrast, the most common amino acids at these sites in the reference set were E, T, E, S, T, A, I, and Q, which occurred with frequencies ranging from 0.56 to 0.81. No sequence in the reference set contained more than four of the dentist's signature amino acids; most contained two or fewer.

We then inspected the viral sequences of the dental patients and the LCs for the presence of the dentist's signature pattern. No clone sequence from patients A, B, C, E, and G possessed fewer than seven of the eight signature amino acids (Table 2). As with the reference set of 32 sequences, most LC sequences showed two or fewer of the signature amino acids. In computing the statistical significance of these findings, we used the sequences from those individual clones for each of the patients A, B, C, E, and G that agreed least with the dentist's signature. For the seven LC cases in which there were multiple clones, we used the clone sequence that agreed best with the dentist's signature. This biased the outcome toward the hypothesis that the viruses of the dental patients and the LCs were equally similar to the dentist's viruses. Even so, the Wilcoxon rank-sum test ($P = 8 \times 10^{-6}$) shows the five patients' viruses have significantly better agreement with the dentist's signature pattern than do the 28 LC viruses for which agreement at all eight signature amino acid sites could be assessed

(30). This finding corroborated the genetic distance analysis (Table 1) and phylogenetic tree analyses (Fig. 1). Furthermore, the viruses from patients D and F are again distinct from the viruses infecting the other five patients; they possessed no more than two signature amino acids.

Both the genetic distance and signature pattern analyses identified a few LCs that were close to, but distinct from, the dental cluster. LC35 and LC9 had C2-V3 nucleotide sequences that were closest to those of the dentist by genetic distance analysis, averaging 6.7% (range 5.8 to 7.8%) and 6.9% (range 5.8 to 8.2%), respectively. The sequences from these two LCs possessed four of the eight amino acids in the dentist's signature pattern (Fig. 1). The maximum number of the dentist's signature amino acids, five, found in any of the LC sequences occurred in one clone sequence from LC3. However, this clone had an average distance of 9.8% from the dentist's clone sequences, and the other five clones from LC3 all contained only two signature amino acids. By phylogenetic analyses, these LC sequences did not appear to be part of the dental clade (Fig. 1).

We next addressed the question of whether two of the measures of genetic similarity (genetic distance and amino acid signature pattern analysis) were truly different measures or were merely restatements of the same measures. If the information contained in these two measures overlapped, we could expect to see a high degree of correlation between them. However, Kendall's τ_b (31), a measure of correlation computed from the 28 LCs for which both measures were available, was −0.07 (95% confidence interval from −0.37 to 0.24; a value of 0 indicates a lack of correlation). Therefore, we conclude that the analyses of average distance and signature pattern are separate assessments of genetic similarity.

With the discovery that phylogenetic information is found in amino acid signature patterns, we also examined the signature patterns belonging to the dental patients' viruses. With the same reference set and criteria employed in the definition of the dentist's signature, unique signatures for each of the dental patients' sets of viruses were determined with the following results: (i) patients A, B, and E have identical or nearly identical signatures to that of the dentist and (ii) patients C and G have larger signatures, 15 and 14 residues, respectively (A, I, G, V, A, D, R, E, V, V, I, T, H, P, and V and A, I, A, G, A, D, V, Y, E, V, V, I, H, and V). Nevertheless, the signature from patient C's viruses is closer to the dentist's viral sequences than to any other sequences. The signature from patient G agrees best with a viral sequence from patient C but next best to a dentist

viral sequence. No LC sequences or reference set sequences were found to match as many as 50% of the amino acids in the signatures from patients A, B, C, E, and G.

Although the dentist was the most likely source of the infection for the five patients, all the patients were not necessarily infected with each of the dentist's HIV variants. This is evident upon inspection of the deduced amino acid sequences, which demonstrate two C2-V3 subtypes for the dentist (Dent-I and Dent-II in Fig. 2). Both the dentist's subtypes are also present in patient A's viral sequences (Pt-A-I and Pt-A-II). However, sequences closer to Dent-I were found in patients B and E, and sequences closer to Dent-II were found in patients C and G.

Conclusion

In summary, seven HIV-infected patients were identified in the practice of a dentist with AIDS. Of these seven patients, five had no identified risk for HIV infection and had invasive procedures performed by the dentist. These five patients were infected with HIV strains that had nucleotide sequences and amino acid signature patterns that were closely related to those of the dentist's viruses. In addition, the HIV sequences of the dentist and these five patients were distinct from those found in the two dental patients with known behavioral risks for HIV infection and in 35 other HIV-infected persons residing in the same
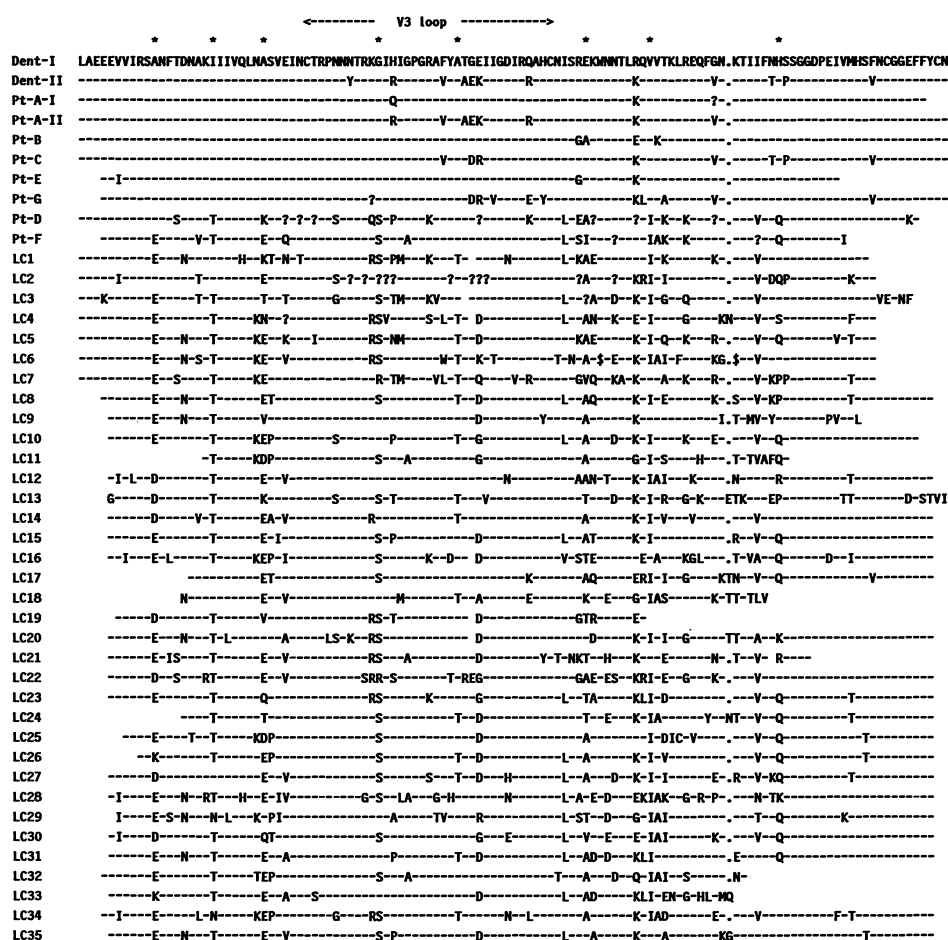

Fig. 2. Comparison among V3 region amino acid sequences from 43 HIV-infected persons from Florida. For the purpose of this figure only, consensus sequences are shown. Because two distinct sequence subtypes were present in the dentist and patient A, two consensus sequences are shown for the dentist (Dent-I and Dent-II) and patient A (Pt-A-I and Pt-A-II). Sequences from the C2-V3 clone ($n = 72$) and direct ($n = 41$) sequences have been deposited with GenBank (accession numbers M90847 to M90906; M90914 to M90966). Sequences from the V4-C3-V5 region have also been deposited with GenBank (accession numbers M91084 to M91115; M91121 to M91149; and M91153 to M91156). The reference sequence is Dent-I from the dentist. Residues identical to Dent-I are shown as a dash. Regions not sequenced are left blank. The sequence stretch corresponding to the V3 loop is indicated. Other symbols are ?, no consensus amino acid at the position indicated; *, noncontiguous amino acids of the dentist's signature pattern (Table 2); ·, a gap to maintain alignment of the amino acids; and $, a stop codon. Single-letter abbreviations for the amino acid residues are: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.

geographic area. Therefore, both the epidemiologic and laboratory data indicate that these five patients were infected during their dental care.

The precise mode of HIV transmission to these patients could not, however, be identified. Patients could have been directly exposed to the dentist's blood as a result of percutaneous injuries sustained by the dentist while performing invasive procedures on these patients. Although none of the patients reported that the dentist injured himself while caring for them, they would not necessarily have been aware if such injuries had occurred. Transmission might also have occurred by contamination of instruments or other dental equipment with blood from the dentist or, less likely, a patient or patients already infected by the dentist. All five patients had invasive procedures performed by the dentist after he was diagnosed with AIDS. At this late stage of disease, higher virus titers may have been present in the dentist's blood and he may have been more likely to transmit virus than earlier in the course of his HIV disease (32).

The theoretical possibility of HIV transmission from infected health-care workers to patients during invasive procedures has been previously acknowledged (33), and there are well-documented reports of HIV transmission from patients to health-care workers after percutaneous exposure to HIV-infected blood (34). In addition, transmission of hepatitis B virus, which has epidemiologic transmission patterns similar to HIV, from health-care workers to patients during invasive medical and dental procedures has been reported (35). For most of these outbreaks, the precise mode of hepatitis B virus transmission from the worker to patient could not be identified, as was true for the HIV transmissions reported here.

Analysis of viral nucleotide sequence distances, inferred phylogenetic relationships, and signature patterns of deduced amino acid sequences played a central role in the investigation of this cluster of HIV-infected patients. Similar analyses may prove to be useful in other investigations of HIV transmission, such as those seeking to determine the source of HIV infection in persons with multiple risk factors (for example, a health-care worker who has sustained an occupational exposure to HIV and reports male-to-male sexual contact). Genetic analysis is also proving to be valuable in tracking the spread of HIV in countries, such as Thailand, where the epidemic is of recent onset (36).

In the current investigation, the divergence of HIV sequences within the Florida background population was sufficient to identify strain variation. However, in countries where the virus has been recently introduced, there may be limited HIV genetic variation and analysis of greater sequence lengths may be necessary to identify strain variation and to determine possible epidemiologic linkage. Nonetheless, this investigation demonstrates that detailed analysis of HIV genetic variation is a new and powerful tool for understanding the epidemiology of HIV transmission.

Note added in proof: Subsequent to the preparation of this report, the Centers for Disease Control became aware of another HIV-infected dental patient (patient H). This patient has acknowledged behavioral risk factors for HIV infection; his HIV clone sequences (GenBank accession numbers M90907 through M90912) are not closely related to the dentist's viruses, differing by more than 10.7% in the C2-V3 region.

REFERENCES AND NOTES

1. J. S. Smith *et al.*, *N. Engl. J. Med.* **324**, 205 (1991); H. J. Lin *et al.*, *J. Infect. Dis.* **164**, 284 (1991); G. P. Holmes *et al.*, *Ann. Intern. Med.* **112**, 833 (1990); R. Rico-Hesse, M. A. Pallansch, B. K. Nottay, O. M. Kew, *Virology* **160**, 311 (1987).
2. P. Balfe *et al.*, *J. Virol.* **64**, 6221 (1990).
3. G. Myers *et al.*, Eds., "Human retroviruses and AIDS 1991: A compilation and analysis of nucleic acid and amino acid sequences," (Theoretical Division, T10, Los Alamos National Laboratory, Los Alamos, NM, 1991).
4. *Morbid. Mortal. Wkly. Rep.* **39**, 489 (1990).
5. *Ibid.* **40**, 21 (1991); *ibid.*, p. 377; C. A. Ciesielski *et al.*, *Ann. Int. Med.* **116**, 798 (1992).
6. J. P. Vartanian, A. Meyerhans, B. Asjo, S. Wain-Hobson, *J. Virol.* **65**, 1779 (1991); S. Delassus, R. Cheynier, S. Wain-Hobson, *ibid.*, p. 225; A. Meyerhans *et al.*, *Cell* **58**, 901 (1989); M. Goodenow *et al.*, *J. Acquired Immune Defic. Syndrome* **2**, 344 (1989); M. Alizon *et al.*, *Cell* **46**, 63 (1986); B. R. Starcich *et al.*, *ibid.* **45**, 637 (1986).
7. B. H. Hahn *et al.*, *Science* **232**, 1548 (1986); M. S. Saag *et al.*, *Nature* **334**, 440 (1988).
8. J. A. McKeating *et al.*, *AIDS* **3**, 777 (1989); P. Nara, in *Retroviruses of Human AIDS and Related Animal Diseases*, M. Girard and L. Valette, Eds. (Pasteur Vaccins, Paris, 1989), pp. 203–215.
9. T. Shioda, J. A. Levy, C. Cheng-Mayer, *Nature* **349**, 167 (1991).
10. C. A. B. Boucher *et al.*, *Lancet* **336**, 585 (1990); B. A. Larder, G. Darby, D. D. Richman, *Science* **243**, 1731 (1989); D. D. Richman, J. M. Grimes, S. W. Lagakos, *J. Acquired Immune Defic. Syndrome* **3**, 743 (1990).
11. H. Burger *et al.*, *Lancet* **336**, 134 (1990); H. Burger *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 11236 (1991).
12. S. M. Wolinsky *et al.*, *Science* **255**, 1134 (1992).
13. A. Srinivasan *et al.*, *Blood* **232**, 1548 (1986).
14. K. Cichutek *et al.*, *AIDS* **5**, 1185 (1991).
15. Peripheral blood mononuclear cells (PBMCs) were isolated, divided, and processed in a laboratory where no PCR work had been performed. To avoid potential cross-contamination or mix-up of specimens, only one frozen PBMC aliquot of the dentist or his patients was processed at a time. Amplification of HIV *env* DNA was performed by a two-step PCR procedure. The primer pair used in the first amplification step was CL207 (5'-GTATGAATTCAACTGCTGTTAAATGGCAGT-3') and CO72 (5'-TATAGAATTCACTTCTCCAAT-TGTCCCTCAT-3'). Approximately 1 µg of PBMC DNA was used in a standard 100-µl reaction (16) and the PCR profile was 96°C, 1 min, 65°C, 2.5 min, temperature ramping time 30 s and the cycle was repeated 40 times. Five microliters of the amplified product was diluted 50-fold and 5 µl of the diluted DNA was reamplified 30 cycles with a new primer pair C072 and derivatives of CL207 containing a 9-base extension (5'-CTAGCAGAA-3') at the 3' end of CL207. Both the CL207 and its extended derivatives have an Eco RI restriction endonuclease recognition sequence (GAATTC). Amplified DNA from the secondary PCR was purified by NACS columns (Bethesda Research Laboratory, Gaithersburg, MD), sequenced directly or digested with Eco RI, and cloned into M13 vectors. To assist the tracking of recombinant phages generated from the dentist and the patients A through G, a specific dinucleotide sequence was added 3' to the Eco RI site of each of the CL207 derivatives to serve as person-specific dinucleotides or "bar codes." When possible, a different M13 vector was used to generate clones from different persons in the dental cohort. These two additional measures facilitated the management and identification of the source of HIV-containing phage. M13 clones containing HIV inserts were screened by plaque hybridization using ^{32}P-labeled CO249 (5'-CAAATATTACA-GGGCTGCATTAACAAGAGATGGTGGTAA-3') or its complementary sequence CO250 (5'-TTAC-CACCATCTCTTGTTAATAGCAGCCCTGTAA-TATTTG-3') as probes derived from the conserved C3 domain. The Taq Dye Primer Sequencing Kit and the 373A DNA sequencer (Applied Biosystem, Foster City, CA) were used. Fluorescence-tagged M13 universal primers were used for sequencing M13-derived clones, whereas a fluorescence-tagged derivative of CL207 primer was used for direct sequencing of amplified DNA.
16. M. Rogers *et al.*, *N. Engl. J. Med.* **320**, 1649 (1989).
17. Precautions included (i) ultraviolet irradiation of vials and laboratory benches (18), (ii) apportioning reagents for single use, (iii) use of positive displacement pipetting devices, and (iv) use of a room dedicated specifically for PCR (19).
18. G. Sarkar and S. S. Sommer, *Nature* **343**, 27 (1990); C.-Y. Ou, J. L. Moore, G. Schochetman, *BioTechniques* **10**, 442 (1991).
19. S. Kwok, *Amplifications* **2**, 4 (1989); _____ and R. Higuchi, *Nature* **339**, 237 (1989).
20. The V3, V4, and V5 sequences of patients A and B and a local control (GenBank accession numbers M92100 to M92150) were also determined by A. J. L. Brown and his colleagues at the University of Edinburgh. PBMC DNA from patient A and an LC were prepared in a CDC laboratory in which no work had been done with HIV. PBMCs from patient B were sent directly from CDC's Epidemic Response Laboratory to Edinburgh. None of the three specimens were handled by CDC personnel involved in the PCR and sequencing of the Florida specimens.
21. For the nucleotide sequence comparisons shown in Table 1, a total of 94 sequences, derived from both M13 clones and from direct sequencing of amplified products, were analyzed over 243 shared positions in the C2-V3 region, for a total of 4371 pairwise comparisons. In keeping with current practice, averages and ranges of sequence differences are reported as percentages rather than as number of nucleotide differences. The robustness of this analysis, although involving small differences in nucleotide counts, is supported by (i) the outcome of the Wilcoxon rank-sum analysis; (ii) the consistency observed for the ranges, which do not show marked variation (Table 1); and (iii) the agreement of the intrapatient differences with what has been reported (2, 3, 6).
22. H. Steuler, B. Storch-Hagenlocher, B. Wildemann, *AIDS Res. Hum. Retroviruses* **8**, 53 (1992); R. Gibbs *et al.*, in *PCR Technology: Principles and Applications for DNA Amplification* (Stockton, New York, 1989), pp. 171–191.
23. The Wilcoxon rank-sum statistic was calculated from the average distance between the dentist and each of patients A, B, C, E, and G and the 30 LCs for whom genetic distance measurements were available. When clone sequences were available, these were used in preference to direct sequences. In all Wilcoxon statistics reported here, average ranks were used in case of ties. All *P* values for the Wilcoxon statistics given in this

paper were obtained from the exact distribution of the rank-sum statistic [for example, see T. P. Hettmansperger *Statistical Inference Based on Ranks* (Wiley, New York, 1984)]. Since the P value for the Wilcoxon statistic may depend on the mix of direct and clone sequences, we also conducted a bootstrap hypothesis test [P. Hall and S. R. Wilson, *Biometrics* **47**, 757 (1991)]. This analysis also took into account the varying number of clone sequences available per individual. We found a 95% confidence interval for the P value to be between 2.7×10^{-6} and 4.4×10^{-6}. We also computed a Wilcoxon rank-sum statistic from average distances between the cloned viral sequences of the dentist and the direct sequences of the five dental patients and those 28 LCs for whom direct sequences of sufficient length were available ($P = 8 \times 10^{-6}$). Finally, we computed a Wilcoxon rank-sum statistic from distances from the direct sequence of the dentist to the direct sequences from the patient group and the 28 LCs ($P = 2 \times 10^{-5}$).

24. J. Felsenstein, *Evolution* **39**, 783 (1985); *Annu. Rev. Genet.* **22**, 521 (1988).

25. Tree analyses were principally conducted with the use of the MULPARS and branch and bound programs of PAUP (Phylogenetic Analysis Using Parsimony, version 2.4.2, written and distributed by D. L. Swofford, Illinois Natural History Survey, Urbana). The LC consensus sequence was the designated hypothetical ancestor (Hypanc). Trees were rooted upon the African sample sequence ELI or upon the midpoint of the greatest patrisitic distance with the same outcome. In a "user-defined" tree, the dentist's sequences were required to cluster with the LC consensus sequence. To examine the consequences of rearranging the order of input of sequences, trees were constructed using the JUMBLE option of DNAPARS program of PHYLIP (version 3.2, written and distributed by J. Felsenstein, University of Washington, Seattle).

26. Of the 146 varied sites, 64 sites manifested nucleotide changes in no more than one of the sequences included in the analysis; for the purposes of some bootstrapping evaluations, these sites were deemed noninformative [W.-M. Li and D. Graur, in *Fundamentals of Molecular Evolution* (Sinauer, Sunderland, MA, 1991), pp. 111–113] and were accordingly excluded. The effect upon

the bootstrapping turned out to be negligible (80% versus 79%).

27. Bootstrap analysis utilized the DNABOOT program of PHYLIP (*24*). Increasing the number of iterations, or replicates, from 100 to 400 yielded the same proportion of trees in which the dental clade could be identified. There is no consensus about the interpretation of bootstrap results; on the one hand, confidence intervals may be larger than the set of equally parsimonious trees (*24*) but, on the other hand, bootstrap intervals can be underestimates of confidence intervals (*28*).

28. D. Hillis and J. Bull, University of Texas, Austin, unpublished observations.

29. The choice of 32 reference sequences was dictated solely by what was available for analysis in the HIV Sequence Database, with the following qualifications: (i) sibling sequences (sets of viral sequences from the same patient) were not included; (ii) only non-African, "MN-like" sequences [G. J. LaRosa *et al.*, *Science* **251**, 811 (1991)] were included in light of the fact that none of the dental or Florida control sequences manifested "African-like" V3 region sequences; and (iii) whereas over 1000 "V3 loop" sequences (33 to 35 amino acids) are present in the HIV database, the current analysis required complete V3 region sequences. The HIV-1 reference set comprises MN, LAI, JRCSF, ALA1, ADA, RF, WMJ2, JH3, SC, BRVA, BAL1, JFL, OYI, NY5, SF162, SF2, SF33, HAN, CDC451, SBA, MA221, JB02, FJS, WM, ACH9, JM, ACP1, 6008ar, 5986ar, 6000ar, 5998ar, and 6002ar (*3*). Signature amino acids in the dentist's viruses are those that differ from what is found in 50% or more of these sequences. Reduction of the reference set by exclusion of the last five sequences in the set (the Haitian samples, kindly provided by N. Halsey and his colleagues and sequenced at CDC) made no difference in the determination of this signature. Signature analysis utilizing nucleotide sequences rather than amino acid sequences confirmed without exception the results reported here. A fuller description of the computer application (VESPA, or viral epidemiology signature pattern analysis) and the results with nucleotide signature patterns is being prepared (B. T. M. Korber and G. Myers, unpublished observations).

30. The Wilcoxon rank-sum statistic was calculated from a mix of clone and direct sequences. Hence

the P value, which here is simply twice the reciprocal of the binomial coefficient $\binom{33}{5}$, may be questioned. Thus, we also conducted an analysis with only direct sequences from the dental patients and the 33 LCs for whom direct sequences were available. Because the direct sequences of two of the dental patients were too short to assess agreement at all eight sites, we tested the hypothesis that the probability of agreement at i sites, given that j were available, was equal among the dental patients and the LC by Fisher's exact test ($P = 1.1 \times 10^{-4}$) [G. H. Freeman and J. H. Halton, *Biometrika* **38**, 141 (1951)].

31. Because agreement in the signature pattern analysis is measured on a scale from 0 to 8, we have used a form of Kendall's τ, denoted τ_b, which is adjusted for ties [A. Stuart, in *Encyclopedia of Statistical Sciences* (Wiley, New York, 1983), vol. 4, p. 367]. A less familiar measure, which uses only the untied data and hence may be a better estimator when data have extensive ties, is Kruskal and Goodman's γ. We find that $\gamma = 0.08$, with asymptotic confidence interval (-0.42, 0.27).

32. D. D. Ho, T. Moudgil, M. Alam, *N. Engl. J. Med.* **321**, 1625 (1989); R. W. Coombs *et al.*, *ibid.*, p 1626.

33. *Morbid. Mortal. Wkly. Rep.* **36**, 2S (1987); *ibid.* **35**, 221 (1986).

34. R. Marcus and the CDC Cooperative Needlestick Surveillance Group, *N. Engl. J. Med.* **319**, 1118 (1988).

35. J. Welch *et al.*, *Lancet* i, 205 (1989); F. E. Shaw, Jr. *et al.*, *J. Am. Med. Assoc.* **255**, 3260 (1988); M. A. Kane and L. A. Lettau, *J. Am. Dent. Assoc.* **110**, 834 (1985); J. Ahtone and R. A. Goodman, *ibid.* **106**, 219 (1983).

36. C-Y. Ou *et al.*, *AIDS Res. Hum. Retroviruses*, in press.

37. R. F. Smith and T. F. Smith, *Protein Eng.* **5**, 35 (1992).

38. We thank A. Gifford and R. Lenroot for providing excellent assistance in computer analysis; A. L. Brown and his colleagues for independently confirming the HIV sequences in patients A and B and an LC; D. K. Hillis for his assistance with the phylogenetic tree analysis; B. Holloway, E. George, and N. de la Torre for synthesis and purification of DNA oligomers; D. Hammerlein for handling blood specimens; and R. Moseley for excellent editorial assistance.