JOURNAL OF VIROLOGY, Apr. 1997, p. 2647–2655 Vol. 71, No. 4
0022-538X/97/$04.00+0

Evolution and Probable Transmission of Intersubtype Recombinant Human Immunodeficiency Virus Type 1 in a Zambian Couple

MIKA O. SALMINEN,[1]* JEAN K. CARR,[1] DAVID L. ROBERTSON,[2] PATRICIA HEGERICH,[1] DEANNA GOTTE,[1] CHRISTINE KOCH,[1] ERIC SANDERS-BUELL,[1] FENG GAO,[3] PAUL M. SHARP,[2] BEATRICE H. HAHN,[3] DONALD S. BURKE,[1] AND FRANCINE E. McCUTCHAN[1]

Henry M. Jackson Foundation and Division of Retrovirology, Walter Reed Army Institute of Research, Rockville, Maryland 20850[1]; Department of Genetics, University of Nottingham, Queens Medical Centre, Nottingham NG7 2UH, United Kingdom[2]; and Departments of Medicine and Microbiology, University of Alabama at Birmingham, Birmingham, Alabama 35294[3]

Received 8 July 1996/Accepted 12 November 1996

The extraordinary genetic diversity of human immunodeficiency virus type 1 (HIV-1) results from the introduction of mutations by an error-prone reverse transcriptase and from recombination of the two RNA genomes packaged in the virion during the synthesis of proviral DNA. The occurrence of multiple, genetically distant HIV-1 subtypes and their geographic intermixing set up conditions for dramatic, rather than gradual, changes in genotype whenever genomes from different subtypes are copackaged in virions. Here we describe, for the first time, the sequential generation of multiple different, but related, intersubtype HIV-1 recombinants within an infected individual. Full-length *gag* and *env* genes were recovered directly from peripheral blood mononuclear cells or from primary virus cultures, using serial blood samples from a Zambian woman and a sample from her spouse. DNA sequencing and phylogenetic analysis established that two different A/C recombinant forms of HIV-1 predominated at two time points in the woman. A related but distinct recombinant HIV-1 was recovered from her spouse. Intersubtype recombination apparently played a central role in the evolution of HIV-1 in this couple and may contribute substantially to the rapid emergence of HIV-1 variants whenever mixed-subtype HIV-1 infections occur.

The human immunodeficiency virus type 1 (HIV-1) packages two complete single-stranded RNA genomes into the virion. With each replication cycle, the reverse transcriptase enzyme introduces point mutations (30, 35) and mediates recombination between the two RNA genomes by frequent template switching (6, 16, 19, 21, 44). Retrovirus recombination can generate viruses with altered biological properties. Experiments with feline and murine retroviruses have demonstrated that if appropriate selection pressures are applied, mixed infections can generate recombinant viruses with altered tissue tropism, pathogenicity, or host range or with changes in antigenic epitopes (15, 46).

The consequences of recombination could be particularly significant in the case of HIV-1, given the wide genetic diversity of forms that constitute the global pandemic. At least eight major HIV-1 clades, designated subtypes A through H, and a rare outlier group (O [17, 47]) have been described. The HIV-1 subtypes are genetically quite distinct, with nucleotide distances of up to 30% in the outer *env* and 15% in *gag* (20, 27, 28, 33, 47). Recently, it has been shown that 5 to 10% of the HIV-1 strains that have been at least partially sequenced are mosaic genomes with genetic material from two different subtypes (37, 38). The circumstances fostering genetic exchange between HIV-1 subtypes are largely undetermined, but it appears that recombinant viruses are already contributing substantially to the global pandemic. An analysis of the viral strains that emerged from Africa to spark the rapidly growing HIV-1 epidemic in Southeast Asia has revealed that these strains are recombinants (4, 13). Recombination has also been implicated in the acquisition of high levels of zidovudine resistance by HIV-1 in vitro (22).

Evidence that would more clearly define the role of intersubtype recombination in the diversification of HIV-1 is accumulating but still fragmentary. Most of the recombinant sequences that have been described were obtained by amplification of portions of the HIV-1 genome from primary virus cultures, rather than directly from peripheral blood mononuclear cells (PBMC). Few full-length recombinant genomes have been characterized. Serial samples are almost uniformly lacking, precluding an analysis of the evolution of recombinant forms in vivo, and few recombinants have been recovered from epidemiologically linked individuals (7, 14, 24, 39, 50).

Here we provide a comprehensive genetic analysis of HIV-1 intersubtype recombinant forms arising in vivo. We found that different recombinant genomes predominated at different times in the blood cells of a single individual, suggesting an ongoing process of generation and/or selection. At least some of the recombinants were viable in culture and apparently transmissible; related recombinant forms were recovered from an epidemiologically linked individual.

MATERIALS AND METHODS

PCR-VNTR fingerprinting. PCR amplification of VNTR (variable number of tandem repeat) loci in the human genome (24–26) was performed in a duplex format, with one reaction for the amplification of loci HUMFES and HUMvWA and a second for the amplification of loci D3S1359 and HUMTH01. The primers 5′-GGGATTTCCCTATGGATTGGG and 5′-GCGAAAGAATGAGACTACAT were used at 1 mM together with primers 5′-CCCTAGTGGATGATAAGAATAATC and 5′-GGACAGATGATAAATACATAGGATGGATGG at 2

* Corresponding author. Present address: National Public Health Institute, HIV Laboratory, Mannerheimintie 166, FIN-00300 Helsinki, Finland. Phone: 358-9-4744454. Fax: 358-9-4744461. E-mail: mika.salminen@ktl.fi.

TABLE 1. Molecular cloning of HIV-1 DNA sequences

Patient	Date (mo/day/yr)	Sample[a]	Molecular clone[b]		
			gag	*env*	Full length
Index case	2/20/1989	Index case P		2 (174-2, -3)	
		Index case C	1 (174-2.1)	3 (174-2, -7, -16)	
	6/20/1990	Index case P		1 (184-1)	
		Index case C	3[c] (184-5.6, -5.2, -5.4)	3[c] (184-5.6, -5.2, -5.4)	1 (184-5.6)
Spouse	2/20/1989	Spouse P	1 (716-17)	1 (716-3)	

[a] Source of DNA template for PCR amplification and cloning is indicated by P (primary PBMC) or C (cells from virus culture).
[b] Regions of the HIV-1 genome that were cloned and sequenced included *gag* (1.5 kb), gp160 (2.5 kb), or the virtually full length genome (9.0 kb). The number of clones sequenced (identification codes in parentheses) from each region is shown.
[c] Sequenced from full-length clones.

mM to amplify HUMFES and HUMvWA, respectively, with denaturation for 30 s at 94°C and annealing for 30 s at 54°C. Primers 5′-GTGGGCTGAAAAG CTCCCCATTAT and 5′-GTGATTCCCATTGGCCTGTTCCTC were used at 1.4 mM together with primers 5′-ATGCTAAGTGCTAAGTCAACT and 5′-G TTGCCTCTGACATGGCTTT at 0.4 mM to amplify HUMTH01 and D3S1359, respectively, with denaturation for 1 min at 94°C and annealing for 45 s at 55°C. For both reactions, a 1-min extension at 72°C and 32 cycles were used, and the reactions were in 10 mM Tris-HCl (pH 8.8)–50 mM KCl–200 mM each deoxynucleoside triphosphate (Pharmacia Biotech, Piscataway, N.J.)–1 U of *Taq* polymerase (Perkin-Elmer, Foster City, Calif.) in 100 μl with 100 ng of DNA template. Analysis of product (10 μl per reaction) was on 10% polyacrylamide gels with silver staining (Bio-Rad, Hercules, Calif.) of bands. Positive control DNA was from an individual unrelated to this study, and the negative control lane contained no DNA template.

The probability of carrying a combination of the alleles with the highest frequencies was calculated by using previously reported frequencies at each locus (25, 32, 36). When possible, allele frequencies reported for the U.S. African-American population were used as surrogates for local African populations, for which data are not available. The probability (P_1) of carrying the most frequent alleles (f) at each locus is $P_1 = (f_{\mathrm{HUMFES}} \times f_{\mathrm{HUMvWA}} \times f_{\mathrm{HUMTH01}} \times f_{\mathrm{HUMD3S1359}})^{-1}$. The probability of two randomly drawn persons carrying the same set of alleles (P_2) is obtained by $P_2 = (P_1)^2$.

PCR and cloning of viral genome segments. Full-length clones were amplified by using primers MSF12 and MSR5 and rTth and Vent polymerases as previously described (43). From virus cultures, complete *gag* and *env* genes were PCR amplified and cloned as reported previously (27, 28) but using the Expand Long Template kit (Boehringer Mannheim, Indianapolis, Ind.). Complete *env* and *gag* genes from primary PBMC were recovered by using nested PCR as follows. First-round outer primers for *env* were JL86 (5′-CCGTCTAGATGCTGTTTA TTCATTTCAGAATTGG-3′) and JL89 (5′-TCCAGTCCCCCCTTTTCTTTTA AAAA-3′), and nested primers were ED3 (5′-TTAGGCATCTCCTATGGCA GGAAGAAGCGG-3′) and JL88 (5′-TAAGTCATTGGTCTTAAAGGTACCT G-3′). For *gag* primers, MSF12 (5′-AAATCTCTAGCAGTGGCGCCCGAAC AG-3′) and BJPOL3 (5′-GTTGACAGGTGTAGGTCCTAC-3′) were used for the first-round PCR and primers G37 (5′-TCGAGAATTCAGGGGTCGTTGC CAAAGA-3′) and G40 (5′-GATCGAATTCGACTAGCGGAGGCTAGAAG-3′) were used for the second. Five percent of the first-round product was used as the template for the second round. PCR was performed according to the manufacturer's recommendations, using thermocycle conditions of 94°C for 10 s, 55°C for 30 s, and 68°C for 4 min for 30 cycles. At cycle 11, 20 s was added to the extension time for each cycle. The PCRs amplify the complete *env* gene in a fragment of 3.0 kb and the almost complete *gag* gene in a fragment of 1.4 kb. Amplification products were TA cloned by using the pCRII vector (Invitrogen, San Diego, Calif.), and plasmid DNA for sequencing was prepared with Qiagen (Chatsworth, Calif.) columns. Clones were sequenced by using dye-deoxy terminator chemistry on ABI 373A automatic sequencers (Applied Biosystems, Foster City, Calif.).

Genetic distances and phylogenetic analysis. Distance and phylogenetic analyses were performed by using programs of the PHYLIP package, version 3.4 (11), as well as CLUSTALW (45) and fastDNAml (34). DNA sequences of the *env* gene or of the *gag* gene from the cloned samples (Table 1) were aligned, and pairwise DNA distances were calculated by DNADIST, using the Felsenstein maximum-likelihood substitution model (8), with base frequencies and transition/transversion ratio estimated from the data. Phylogenetic analyses using maximum parsimony (DNAPARS and CONSENSE), neighbor-joining (DNADIST and NEIGHBOR or CLUSTALW), and maximum-likelihood (fastDNAml) methods resulted in trees of essentially identical topology throughout this study. Bootstrap values (9) were determined by the maximum-parsimony method, using SEQBOOT, DNAPARS, and CONSENSE.

Bootscanning. For the bootscanning procedure (41), the query sequences were first aligned with a set of reference sequences representing all established genetic subtypes of HIV-1, including unpublished sequences from our laboratory. The alignment was broken into sequential, overlapping segments, and bootstrapped phylogenetic analyses (maximum parsimony and neighbor joining) were performed on each segment. Parental genotypes were identified as those exhibiting bootstrap values of greater than 70%. Breakpoints were fine mapped by using a four-sequence alignment consisting of the putative recombinant sequence, both parental genotypes as 50% majority consensus sequences (all available C-type and four unpublished A_{CY} sequences [see Results section for explanation of A_{CY}], each used to construct the consensus *gag* and *env* sequences), and an outgroup (isolate MN, subtype B). Breakpoints were assigned to the midpoint of the transitions between segments from different subtypes. Using a consensus based on all A-type sequences excluding the A_{CY} sequences resulted in similar results, except that some breakpoints were left undetected (not shown, available on request). The *pol* region of each of the full-length sequences was bootscanned by using sequences from subtypes A (U455; accession number M62320), C (C2220 [42]; accession number U46016), and B (MN).

Informative site distribution analysis. To localize intragenic crossover points between regions of DNA sequences, the distribution of phylogenetically informative sites supporting alternative tree topologies was inspected. This was done by surveying the informative sites in a four-sequence alignment including the putative recombinant sequence, consensus sequences derived from the two parental subtype lineages (A_{CY} and C), and an outgroup (B isolate MN). There are three possible configurations of the informative sites, two of which support the clustering of the putative recombinant with one parental lineage or the other. The distribution of these two types of sites can be tested by determining whether a break placed at any point along the alignment produces a significant difference in the ratio of the two types of sites on each side of the cut, as assessed by a chi-square value; the optimum position of the breakpoint can be found by maximizing this value (31). Probability values were determined by carrying out 10,000 simulations.

Nucleotide sequence accession numbers. The sequences obtained in this study have been assigned GenBank accession numbers U86768 to U86781.

RESULTS

The index case for this study was an HIV-1-positive woman who participated in a clinical research study in Lusaka, Zambia, in 1989 and 1990. The *env* gene (clone ZAM184*) of a virus cultured from her PBMC in 1990 behaved as an outlier to all of the recognized HIV-1 subtypes in the initial phylogenetic analyses (27). Samples from other individuals in the same study clustered within HIV-1 genetic subtype C. When the recognition of the first mosaic HIV-1 genomes sparked a reevaluation of all available gene sequences of HIV-1 (37, 38), *env* gene ZAM184* emerged as the first example of a recombinant between HIV-1 subtypes A and C.

The availability of additional samples from the index case and from her spouse provided an opportunity to investigate the circumstances surrounding the generation of HIV-1 recombinants. Our repository contained additional cryopreserved primary PBMC from the 1990 blood sample and primary PBMC obtained 16 months earlier from the same individual and her spouse. Both of the blood samples from the index case had yielded positive primary virus cultures, from which cryopreserved cells were also available. No isolate was available from the husband's sample.

Verification of sample identity. Before initiating molecular cloning and genetic analysis of HIV-1 sequences, we verified

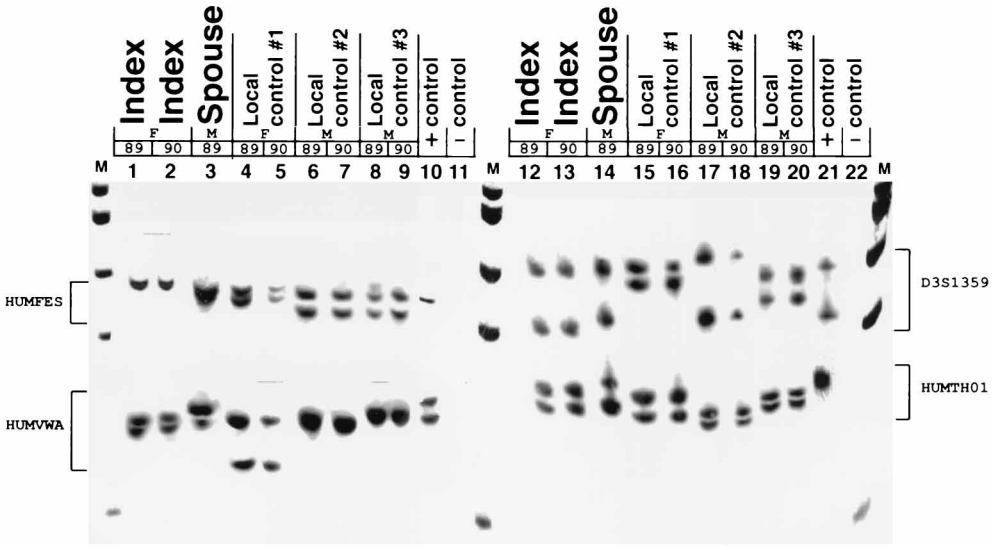

FIG. 1. Genotyping of patient primary PBMC DNAs. Shown are silver-stained gels of four PCR-amplified VNTR repeat loci from the 1989 and 1990 primary PBMC DNAs of the index case (lanes 1, 2, 12, and 13), the 1989 sample of her spouse (lanes 3 and 14), three local control individuals (lanes 4 to 9 and 15 to 20; year of collection indicated), and positive and negative controls. Lanes M are DNA size markers. The original silver-stained gel was scanned at 400 dots/inch and annotated by using Adobe Photoshop version 3.0.5 (Adobe Systems, Inc.).

that the cryopreserved primary PBMC obtained in 1989 and in 1990 from the index case were, in fact, derived from the same individual. The results of a PCR analysis using VNTR loci HUMFES, HUMvWA, D3S1359, and HUMTH01, which are highly polymorphic in the human genome (25, 32, 36), are shown in Fig. 1. Serial samples from the index case (lanes 1, 2, 12, and 13) showed similar allele patterns in all loci but differed from those of the spouse's sample (lanes 3 and 14) and from those of blood samples from three other individuals in Zambia provided at the same time points (local controls; lanes 4 to 9 and 15 to 20). Even if all of the alleles found in the index case were the most prevalent in the population, the chance for a random genotype match in unrelated samples would be less than 1 in 35,000. These data verify that the 1989 and 1990 samples in our repository were derived from the index case and establish that the spouse's sample was derived from a different individual.

Because the virus isolation procedure involved mixing of the patient's cells with donor PBMC from other individuals (2), we were unable to use this genotyping procedure to establish that the 1989 and the 1990 virus cultures corresponded to their respective primary PBMC samples, but virologic evidence establishing this link is presented.

Clones and sequences of HIV genes. HIV-1 genomic sequences were recovered by PCR amplification and molecular cloning. More than 40 kb of new sequence information was generated from the five samples as described in Table 1. Four nearly full length clones were obtained from the 1990 virus culture of the index case. Of these, one was completely sequenced; from two others, the *gag* and *env* genes were sequenced. The full-length clone which was completely sequenced was determined to be highly similar to the other three clones by restriction mapping. The complete *env* gene was sequenced from both primary PBMC and virus culture at both time points from the index case (*env* genes were separately amplified from the cultured 1989 sample) and from the primary PBMC of her spouse; from some samples, multiple clones were analyzed. Complete *gag* gene sequences were obtained from the two virus cultures of the index case and from the primary PBMC of her spouse.

Genetic distances between HIV sequences from the index case and her spouse. First, the relationships of the viral sequences from the primary PBMC of the index case to the sequences obtained from the corresponding virus cultures were examined. Table 2 shows that the groups of viral sequences recovered from primary PBMC and from culture differed by

TABLE 2. Pairwise distances among proviral HIV-1 sequences[a]

Patient	Date(s) of sample	Comparison category[b]	% Difference					
			No. of comparisons	*env*		No. of comparisons	*gag*	
				Mean	Range		Mean	Range
Index case	1989	A	6	2.3	1.2–3.1			
	1990	A	3	2.0	1.9–2.0			
	1989 and 1990	B	20	13.8	13.1–14.5	3	9.8	9.7–9.8
Index case and spouse	1989	C	5	7.4	6.7–8.0	1	5.4	

[a] Distances between DNA sequences of the *env* gene or of the *gag* gene from the samples described in Table 1 are shown.
[b] Comparisons of cultured and uncultured sequences obtained from the index case patient (A), comparisons of sequences between the two time points from the index case (B), and comparisons of sequences between the index case and her spouse (C).

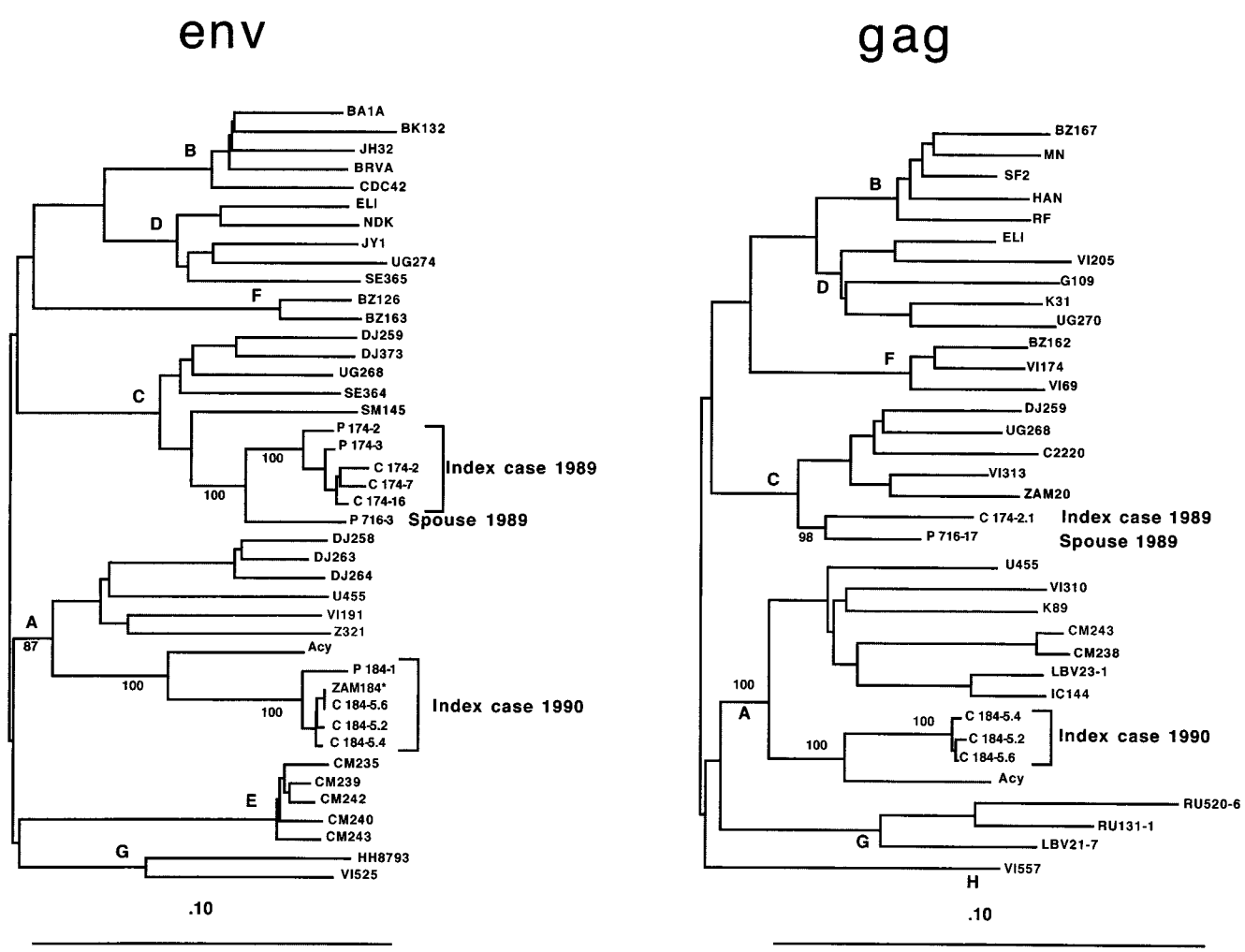

FIG. 2. Phylogenetic classification based on complete *gag* and *env* sequences. Five full-length *gag* (1.5 kb) and ten full-length *env* (2.5-kb) gene sequences from the index case and her spouse (C, cultured PBMC sample; P, primary uncultured sample) were aligned with reference sequences (27, 28, 33) of HIV-1 subtypes A through H, and phylogenetic analyses were performed (34). A_{CY} is a consensus sequence of four separate unpublished sequences which form a subgroup of the A-subtype. Sequence ZAM184* (*env* tree) has been described previously (33). Maximum-likelihood trees with maximum parsimony bootstrap values for the nodes classifying samples from the index case and from her spouse are shown, but other methods produced essentially identical results (11, 45). The shift from inclusion in subtype C for the 1989 samples to an unclassified position for the 1990 samples is evident for all clones analyzed and is observed in both the *env* and *gag* trees. Bootstrap values of key nodes are shown (bootstrap values of all subtype defining nodes were 100).

averages of 2.3% at the 1989 time point and 2.0% in the 1990 sample. These results establish that the viruses recovered in vitro were very closely related to those found in vivo at both time points and that the primary PBMC of the index case were almost certainly the source of the viral sequences recovered from culture.

In contrast, the HIV-1 sequences obtained in 1989 differed significantly from those observed in 1990. The inter-time point variation averaged 13.8% in *env* and 9.8% in *gag* (Table 2). Thus, a relatively homogenous virus population found in the index case in 1989 appeared to be essentially replaced by an equally homogeneous but genetically distant population in 1990. Finally, the DNA distances between HIV-1 sequences from the index case and from her spouse obtained in 1989 differed by averages of 7.4% in *env* and 5.4% in *gag.*

Evolutionary relationships between HIV sequences from the index case and her spouse. The phylogenetic relationships among the HIV-1 sequences from the index case and her spouse were then determined (8–10), in conjunction with representative sequences from HIV-1 subtypes A through G for *env* and A through H for *gag*. References included published sequences which have previously been shown to be nonrecombinant (33, 37, 38) as well as unpublished sequences of isolates from Cyprus (to be presented elsewhere) which, based on a shorter previously published sequence, seem to cluster with subtype A and form a subgroup that has been called A_{CY} (23). These were included since they have been shown to be the closest relative of ZAM184* within subtype A (Fig. 2). For both genes, the 1989 sequences from both the index case and her spouse formed a monophyletic cluster in subtype C. The close relationships of the sequences derived from the virus culture to those found in vivo is reinforced by this analysis, as is the close linkage of the sequences from the index case and her spouse. Sequences from the 1990 samples of the index case formed an equally tight cluster but appeared to be only distantly related (based on estimated genetic distance) to the A_{CY} subgroup of the A subtype, behaving as outliers like the initial *env* clone (ZAM184*) from the index case. Sequences found in uncultured and cultured PBMC were virtually identical.

Evidence of intersubtype recombination. The genetic basis for the dramatic shift in the phylogenetic position of viral sequences found in the PBMC of the index case over a 16-

month interval was investigated. Because the prototype *env* clone (ZAM184*) from the 1990 sample of the index case had been previously shown to be an A/C recombinant (37, 38), we examined the possibility that different recombinants were present at different times. A procedure for resolution of the parentage of recombinant HIV-1 genomes termed bootscanning (41) was used. Using a multiple sequence alignment with reference sequences, we analyzed overlapping segments of the *gag* and *env* genes separately for their phylogenetic positions. The bootstrap value establishing the consistency of grouping with subtype A or subtype C was examined for each segment. Figure 3A depicts these results graphically. All of the *gag* and *env* genes recovered from the index case and her spouse were chimeras of subtypes A and C. The sequences from the 1989 samples were largely of subtype C, with regions from subtype A in the middle of the *gag* genes and two subtype A segments close to the 3′ end of *env*. Additionally, the 1989 sequences from the index case and her spouse differed in that the spouse had an additional small segment of subtype C in the central portion of the *gag* gene (Fig. 3A). In contrast, the 1990 sample from the female index case yielded sequences that were mostly subtype A, with only small segments of subtype C genetic material in both *gag* and *env*. The disparate phylogenetic positions of the sequences derived from the serial samples are thus due to recovery of different recombinant genotypes. Apparently, recombination played a significant role in the evolution of viral forms in this individual.

We next determined whether the genetic segments of geno-

TABLE 3. Fine mapping of selected recombination breakpoints[a]

Sample[c]	Gene	Subtype	Segment	Region	No. of informative sites			*P*
					A	C	Out group	
Spouse P (1989)	*gag*	C	I	1–691	4	24	1	
								0.0014
		A	II	777–918	5	0	0	
								0.0157
		C	III	945–969	0	4	0	
								0.0469
		A	IV	1092–1290	6	1	1	
								0.0158[b]
		C	V	1302–1465	1	9	0	
Index case C (1989)	*gag*	C	1	1–705	2	29	2	
								0.0000
		A	II–IV	777–1290	15	1	4	
								0.0000[b]
		C	V	1302–1465	1	9	0	
Index case C (1990)	*gag*	A	I–IV	1–1290	43	3	8	
								0.0000
		C	V	1302–1465	1	9	0	
Spouse P (1989)	*env*	C	VI–VII	1–1718	8	60	9	
								0.0001
		A	VIII	1762–1865	5	0	0	
								0.0131
		C	IX	1873–2171	3	12	0	
								0.0002
		A	X	2186–2611	22	2	3	
								0.0000
		C	XI	2614–2749	1	12	1	
Index case C/P (1989)	*env*	C	VI, VII	1–1718	9	68	6	
								0.0001
		A	VIII	1762–1948	8	1	0	
								0.0133
		C	IX	1985–2171	2	9	0	
								0.0005
		A	X	2221–2593	19	2	3	
								0.0000
		C	XI	2630–2755	1	12	0	
Index case C/P (1990)	*env*	C	VI	1–303	1	14	1	
								0.0000
		A	VII–XI	378–2755	94	18	13	

[a] Analysis was performed as described by Robertson et al. (37, 38). The distribution of informative sites supporting classification as subtype A or C or the outgroup is shown. The last column shows the *P* value for finding a higher chi-square value for the distribution of phylogenetically informative (parsimonious) sites among 10,000 simulations of randomly shuffled data.

[b] The breakpoint was chosen on the basis of a lower than maximum chi-square value based on visual examination of the sequence alignment. Choosing the highest value would move the boundaries of the breakpoint to 1305–1371, which is caused by a single mutation at position 1305.

[c] P, analyzed from primary, uncultured PBMC; C, analyzed from cocultured PBMC; C/P, analyzed from both cocultured and uncultured PBMC.

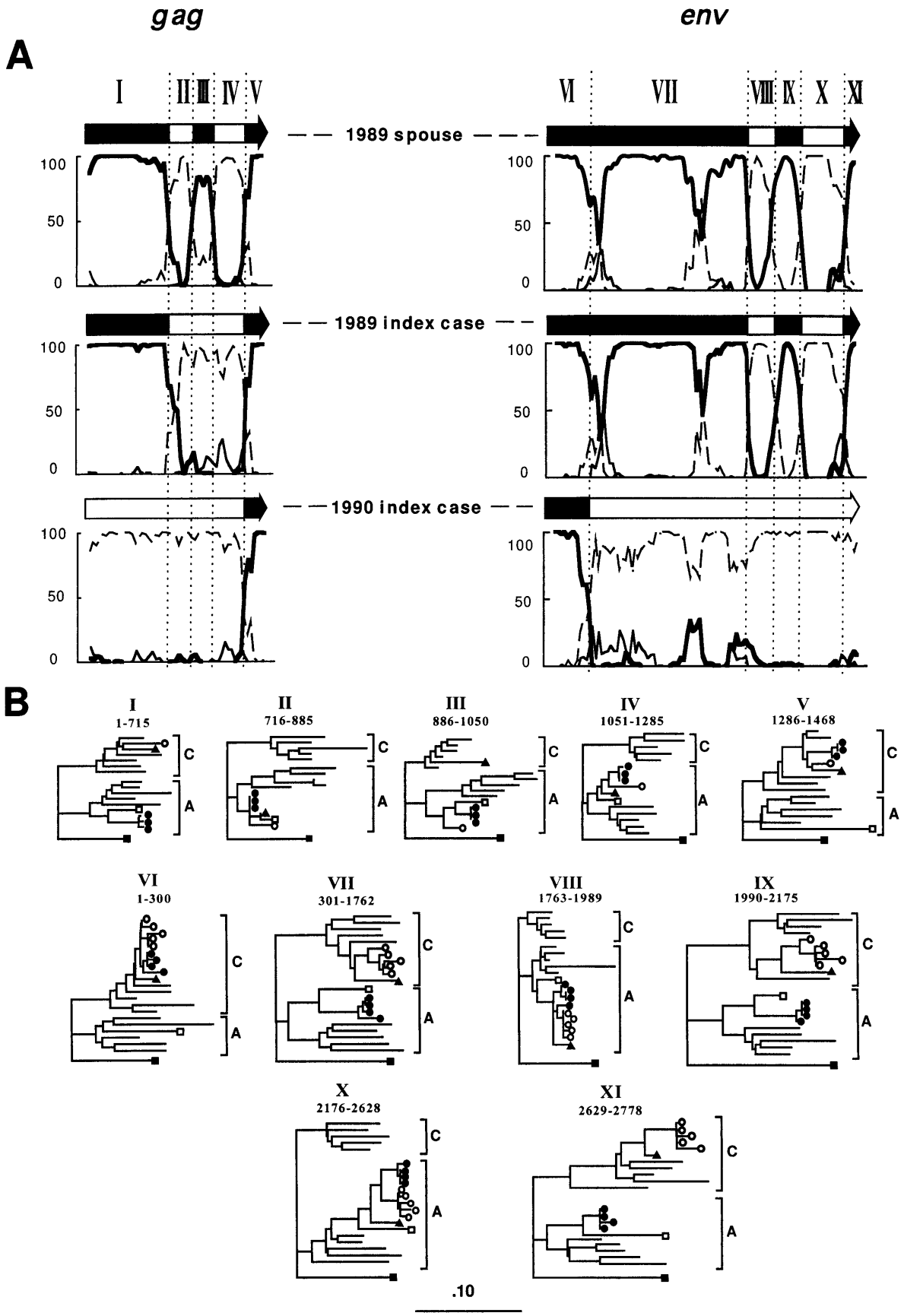
A
gag
env
I
II
III
IV
V
VI
VII
VIII
IX
X
XI
1989 spouse
1989 index case
1990 index case
100
50
0
B
I
1-715
II
716-885
III
886-1050
IV
1051-1285
V
1286-1468
VI
1-300
VII
301-1762
VIII
1763-1989
IX
1990-2175
X
2176-2628
XI
2629-2778
C
A
.10

FIG. 3. Mosaic structures of *gag* and *env* genes. Segments of genetic material from HIV-1 subtype A (open regions) or subtype C (solid regions) were mapped by parsimony bootscanning as described elsewhere (41). (A) Phylogenetic trees were constructed from sequential 200-bp segments overlapping by 20 bp. The bootstrap value grouping each segment with subtype A (thick solid line), subtype C (thin dashed line), or an outgroup (subtype B isolate MN; thin solid line) is plotted (results from multiple clones from both primary and cultured materials were similar at each time point). Recombination breakpoints were identified as the intersection of the subtype A and subtype C lines of the bootscan plots. All of the clones from this study were mosaics of subtype A and subtype C. Additional analyses on the breakpoints were performed (Table 3). (B) DNA sequences were subdivided into segments I to XI according to apparent breakpoints between subtypes A and C (positions in alignment indicated below segment number). Outgroup-rooted (B isolate MN; black filled square) maximum-likelihood phylogenetic trees (34) from these segments are shown. The positions of six reference isolates of subtypes A (A_{CY} [black open square]), five subtype C reference isolates, and all sequences from the index case (1990 [filled circles] and 1989 [open circles]) and her spouse (1989 [filled triangle]) are indicated (an analysis using all known *gag* and *env* sequences gave similar results [not shown]). Bootstrap values of the nodes joining the different combinations of index and spouse sequences were in the range of 54 to 95%. The scale bars indicates 10% nucleotide distance. In reference 38, the 1990 *env* sequence (ZAM184*) was reported to carry an additional segment of subtype C-related material between positions 1068 and 1263. This region could not be detected in our analyses and was probably a result of the use of different reference sequences in the previous study (subtype A consensus [without A_{CY}], subtype C consensus, and simian immunodeficiency virus SIV_{CPZ} as an outgroup).

type A and genotype C from the different samples were sufficiently related to establish direct, ancestral relationships among them. For this analysis, the *gag* and *env* genes were divided into segments bounded by the recombination breakpoints identified by the bootscanning process (segments I through XI [Fig. 3B]). Figure 3B shows the phylogenetic trees constructed from these segments by using reference isolates of subtypes A and C. Segments II, IV, VIII, and X appeared to derive from subtype A in all of the samples, and segments V and VI were consistently of subtype C. Phylogenetic trees from these segments support a very close association of the sequences from the index case in 1989 and 1990 and from her spouse (Fig. 3B). Segment III appeared to derive from subtype C in the 1989 sample from the spouse but was of subtype A in both the 1989 and 1990 samples from the index case. The tree from this segment (tree III [Fig. 3B]) confirms derivation from different subtypes in the index case and her spouse. Similarly, the trees constructed from segments VII, IX, and XI group the 1989 samples from the index case and her spouse within subtype C and show that the 1990 sample from the index case was of subtype A in these regions. These results are consistent with the derivation of all of the recombinant HIV-1 strains from this couple from a single subtype A parent (related to the A_{CY} isolates) and a single subtype C parent via multiple rounds of recombination. They are also equally consistent with the derivation of the various forms by recombination and/or selection from an existing pool of recombinants.

Additional evidence was gathered in direct support of the derivation of one recombinant from another by further analysis of the corresponding breakpoints between genetic segments from subtypes A and C in different recombinants. Table 3 shows the results of an investigation of the distribution of phylogenetically informative sites (37) around the observed breakpoints. Within the precision attainable by this technique, common breakpoints (asterisks in Fig. 3A) were identified between segments IV and V in all *gag* genes obtained and between regions VI to XI of *env* in the 1989 samples (Table 3). Thus, it is unlikely that the recombinant forms in the serial samples of the index case and in her spouse were generated by independent recombination events between the parental A and C viruses; rather, these analyses strongly suggest that earlier recombinant forms served as templates for the generation of later ones.

Four nearly full length molecular clones were also recovered from the 1990 PBMC coculture isolate of the index case. Sequencing of *gag* and *env* genes (Table 1) and restriction endonuclease mapping (data not shown) showed the four clones to be almost identical. One full-length clone was completely sequenced and analyzed for chimeric structure by bootscanning (41). Figure 4 shows that multiple crossover points between subtype A and C are found in the genome of this isolate. Mosaicism was found in both structural and regulatory genes as well as in genes encoding proteins of very diverse functions in the viral life cycle, including the NCp7/p6, protease, reverse transcriptase, gp120, and gp41 genes and *vpu*. The potential of intersubtype recombination to generate novel variants, possibly with altered structural and functional properties, is underscored by this analysis.

Since recombination during PCR has been described to occur under some experimental conditions (49), we performed an experiment where two largely similar cloned infectious proviruses (pNL4-3 and pHXB2; GenBank accession numbers

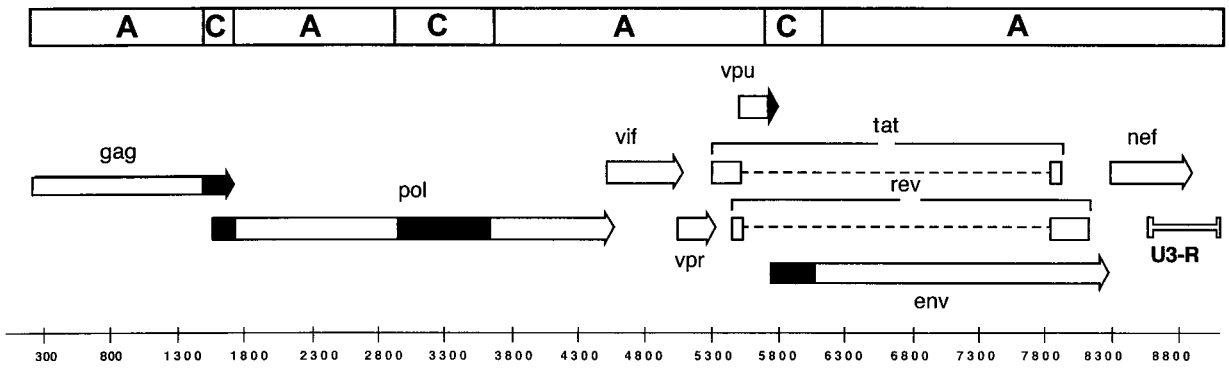

FIG. 4. Virtually full length genome of isolate ZAM184. A 9.0-kb segment containing all but 72 bases of the HIV-1 genome was PCR amplified using DNA from the PBMC coculture of the 1990 sample of the index case (Table 1) as described previously (43), cloned, and sequenced. Open reading frames and long terminal repeat U3-R regions are shown as a gene map. Segments derived from subtypes A and C (shading as in Fig. 3) were mapped by bootscanning as described for Fig. 3 except that 500-bp segments were used in the more conserved *pol* gene. Evidence of mosaicism was found throughout the genome, depicted by shading (as in Fig. 3) and above the gene map. The *gag*, *pol*, and *env* genes are A/C mosaics; the regulatory genes appear to derive from subtype A, with the exception of *vpu*, which is recombinant.

M19921 and K03455) were mixed in equal proportions and amplified by using the procedures used elsewhere in the study. This experiment showed no evidence of recombination (within the sensitivity of a Southern blot) between two differentiating restriction endonuclease-cut sites spaced approximately 3 kb apart. Furthermore, since we recovered multiple clones of similar recombined forms both from primary and cultured materials, on occasions separated several years apart (the *env* gene of the initial index case was first amplified and sequenced in 1992), we rule out PCR-mediated recombination as the source of the chimeric forms.

DISCUSSION

From the data described above, we conclude (i) that the index case harbored at least two A/C recombinant HIV-1 forms that are plausibly related by successive rounds of recombination; (ii) that different recombinant forms predominated in the PBMC of the index case at different times; (iii) that both recombinant forms were viable, as they were recovered in culture; and (iv) that at least some of the recombinant forms were transmitted between the couple.

Did these HIV-1 recombinants arise in the individuals studied here or earlier in the chain of transmission through which they became infected? We found no direct evidence for the continued presence of the parental (nonrecombinant) subtype A and subtype C viruses in either the index case or her spouse. Our data are equally consistent with infection of a mixture of recombinant forms or with a mixed infection with a subtype A and a subtype C virus. However, the recovery of a single recombinant form, rather than a mixture of forms, in each of three samples, the accumulation of subtype A genetic material in both the *gag* and *env* genes in the 1990 compared to the 1989 sample from the female, and the apparent commonality of breakpoints in independent samples and multiple clones all suggest that recombinant forms arose sequentially in one or both of these individuals. An initial analysis of cryopreserved samples from two additional couples who participated in the same 1989–1990 study in Zambia shows only HIV-1 sequences of subtype C (our unpublished data); it seems unlikely that A/C recombinant strains were prevalent in the local population from which the index case and her spouse were drawn.

The prospective impact of intersubtype recombination on the global AIDS epidemic is important to consider. Increasingly, HIV-1 subtypes that were thought to be geographically separated are coming in contact (1, 3, 5, 12, 18, 24, 26, 29, 40, 48). The data presented here illustrate the potential for a single dual infection to set up conditions for multiple segmental exchanges between HIV-1 subtypes, yielding a variety of mosaic forms that can be perpetuated, not only during the course of HIV-1 infection in a single individual but also in individuals in the ensuing chain of transmission. Rapid evolution of HIV-1, already evidenced by the occurrence of multiple subtypes, may take on an added dimension if intersubtype recombinants are afforded the opportunity to arise and become established in human populations.

ACKNOWLEDGMENTS

This work was supported in part by Cooperative Agreement DAMD17-93-V-3004 between the U.S. Army Medical Research and Materiel Command and the Henry M. Jackson Foundation for the Advancement of Military Medicine and by grant 19191 from the Finnish Academy of Science (M. O. Salminen). Sequencing of complete A_{CY} *gag* and *env* genes was supported by grants N01 AI35170 and R01AI25291 (B. H. Hahn and F. Gao).

We are thankful to Matti Lukka (National Public Health Institute, Helsinki, Finland) for expert advice considering the genetic identification of human DNA samples.

REFERENCES

1. **Brodine, S. K., J. R. Mascola, P. J. Weiss, S. I. Ito, K. R. Porter, A. W. Artenstein, F. C. Garland, F. E. McCutchan, and D. S. Burke.** 1995. Detection of diverse HIV-1 genetic subtypes in the USA. Lancet **346:**1198–1199.
2. **Burke, D. S., R. R. Redfield, D. C. Bjornson, A. K. Fowler, and C. N. Oster.** 1989. Frequent isolation of HIV-1 from the blood of patients receiving zidovudine (AZT) therapy. N. Engl. J. Med. **321:**1682.
3. **Campodonico, M., W. Janssens, L. Heyndrickx, K. Fransen, A. Leonaers, F. F. Fay, M. Taborda, G. van der Groen, and O. H. Fay.** 1996. HIV type 1 subtypes in Argentina and genetic heterogeneity of the V3 region. AIDS Res. Hum. Retroviruses **12:**79–81.
4. **Carr, J. K., M. O. Salminen, C. Koch, D. Gotte, A. W. Artenstein, P. A. Hegerich, D. St. Louis, D. S. Burke, and F. E. McCutchan.** 1996. Full-length sequence and mosaic structure of a human immunodeficiency virus type 1 isolate from Thailand. J. Virol. **70:**5935–5943.
5. **Clewley, J. P., C. Arnold, K. L. Barlow, P. R. Grant, and J. V. Parry.** 1996. Diverse HIV-1 genetic subtypes in UK. Lancet **347:**1487.
6. **Coffin, J. M.** 1979. Structure, replication, and recombination of retrovirus genomes: some unifying hypotheses. J. Gen. Virol. **42:**1–26.
7. **Diaz, R. S., E. C. Sabino, A. Mayer, J. W. Mosley, M. P. Busch, and Transfusion Safety Study Group.** 1995. Dual human immunodeficiency virus type 1 infection and recombination in a dually exposed transfusion recipient. J. Virol. **69:**3273–3281.
8. **Felsenstein, J.** 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. J. Mol. Evol. **17:**368–376.
9. **Felsenstein, J.** 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution **39:**783–791.
10. **Felsenstein, J.** 1988. Phylogenies from molecular sequences: inference and reliability. Annu. Rev. Genet. **22:**521–565.
11. **Felsenstein, J.** 1996. PHYLIP: phylogeny inference package, v.3.572c. University of Washington, Seattle, Wash.
12. **Fransen, K., A. Buvé, J. N. Nkengasong, M. Laga, and G. van der Groen.** 1996. Longstanding presence in Belgians of multiple non-B HIV-1 subtypes. Lancet **347:**1403.
13. **Gao, F., D. L. Robertson, S. G. Morrison, H. Hui, C. Stevenson, J. Decker, P. N. Fultz, M. Girard, G. M. Shaw, B. H. Hahn, and P. M. Sharp.** 1996. The heterosexual human immunodeficiency virus type 1 epidemic in Thailand is caused by an intersubtype (A/E) recombinant of African origin. J. Virol. **70:**7013–7029.
14. **Gao, F., S. G. Morrison, D. L. Robertson, C. L. Thornton, S. Craig, G. Karlsson, J. Sodroski, M. Morgado, B. Galvao-Castro, H. von Briesen, S. Beddows, J. Weber, P. M. Sharp, G. M. Shaw, B. H. Hahn, and the WHO and NIAID Networks for HIV Isolation and Characterization.** 1996. Molecular cloning and analysis of functional envelope genes from human immunodeficiency virus type 1 sequence subtypes A through G. J. Virol. **70:**1651–1667.
15. **Golovkina, T. V., A. B. Jaffe, and S. R. Ross.** 1994. Coexpression of exogenous and endogenous mouse mammary tumor virus RNA in vivo results in viral recombination and broadens the virus host range. J. Virol. **68:**5019–5026.
16. **Goodrich, D. W., and P. H. Duesberg.** 1990. Retroviral recombination during reverse transcription. Proc. Natl. Acad. Sci. USA **87:**2052–2056.
17. **Gurtler, L., P. H. Hauser, J. Eberle, A. Von Brunn, S. Knapp, L. Zekeng, J. M. Tsague, and L. Kaptue.** 1994. A new subtype of human immunodeficiency virus type 1 (mvp-5180) from Cameroon. J. Virol. **68:**1581–1585.
18. **Hampl, H., D. Sawitzky, M. Stöffler-Meilicke, A. Groh, M. Schmitt, J. Eberle, and L. Gürtler.** 1995. First case of HIV-1 subtype 0 infection in Germany. Infection **23:**369–370.
19. **Hu, W.-S., and H. M. Temin.** 1990. Retroviral recombination and reverse transcription. Science **250:**1227–1233.
20. **Janssens, W., L. Heyndrickx, K. Fransen, M. Temmerman, A. Leonaers, T. Ivens, J. Motte, P. Piot, and G. van der Groen.** 1994. Genetic variability of HIV type 1 in Kenya. AIDS Res. Hum. Retroviruses **10:**1577–1579.
21. **Katz, R. A., and A. M. Skalka.** 1990. Generation of diversity in retroviruses. Annu. Rev. Genet. **24:**409–445.
22. **Kellam, P., and B. A. Larder.** 1995. Retroviral recombination can lead to linkage of reverse transcriptase mutations that confer increased zidovudine resistance. J. Virol. **69:**669–674.
23. **Kostrikis, L. G., E. Bagdades, Y. Z. Cao, L. Q. Zhang, D. Dimitriou, and D. D. Ho.** 1995. Genetic analysis of human immunodeficiency virus type 1 strains from patients in Cyprus: identification of a new subtype designated subtype I. J. Virol. **69:**6122–6130.
24. **Leitner, T., D. Escanilla, S. Marquina, J. Wahlberg, C. Brostrom, H. B. Hansson, M. Uhlen, and J. Albert.** 1995. Biological and molecular characterization of subtype D, G, and A/D recombinant HIV-1 transmissions in Sweden. Virology **209:**136–146.
25. **Li, H., L. Schmidt, M. H. Wei, T. Hustad, M. I. Lerman, B. Zbar, and K. Tory.** 1993. Three tetranucleotide polymorphisms for loci: D3S1352; D3S1358; D3S1359. Hum. Mol. Genet. **2:**1327.
26. **Liitsola, K., M. Salminen, T. Laukkanen, and P. Leinikki.** Distribution of

genetic subtypes of HIV-1 in Finland. AIDS Res. Hum. Retroviruses, in press.
27. **Louwagie, J., W. Janssens, J. Mascola, L. Heyndrickx, P. Hegerich, G. van der Groen, F. E. McCutchan, and D. S. Burke.** 1995. Genetic diversity of the envelope glycoprotein from human immunodeficiency virus type 1 isolates of African origin. J. Virol. **69:**263–271.
28. **Louwagie, J., F. E. McCutchan, M. Peeters, T. P. Brennan, E. Sanders-Buell, G. A. Eddy, G. van der Groen, K. Fransen, G.-M. Gershy-Damet, R. Deleys, and D. S. Burke.** 1993. Phylogenetic analysis of *gag* genes from 70 international HIV-1 isolates provides evidence for multiple genotypes. AIDS **7:**769–780.
29. **Lukashov, V. V., M. T. E. Cornelissen, J. Goudsmit, M. N. Papuashvilli, P. G. Rytik, R. M. Khaitov, E. V. Karamov, and F. de Wolf.** 1995. Simultaneous introduction of distinct HIV-1 subtypes into different risk groups in Russia, Byelorussia and Lithuania. AIDS **9:**435–439.
30. **Mansky, L. M., and H. M. Temin.** 1995. Lower in vivo mutation rate of human immunodeficiency virus type 1 than that predicted from the fidelity of purified reverse transcriptase. J. Virol. **69:**5087–5094.
31. **Maynard Smith, J.** 1992. Analyzing the mosaic structure of genes. J. Mol. Evol. **34:**126–129.
32. **Moller, A., E. Meyer, and B. Brinkmann.** 1994. Different types of structural variation in STRs: HumFES/FPS, HumVWA and HumD21S11. Int. J. Legal Med. **106:**319–323.
33. **Myers, G., et al.** 1995. Human retroviruses and AIDS. Theoretical Biology and Biophysics, Las Alamos National Laboratory, Los Alamos, N.Mex.
34. **Olsen, G. J., H. Matsuda, R. Hagstrom, and R. Overbeek.** 1994. fastDNAML: a tool for construction of phylogenetic trees of DNA sequences using maximum likelihood. Comput. Appl. Biosci. **10:**41–48.
35. **Preston, B. D., B. J. Poiesz, and L. A. Loeb.** 1988. Fidelity of HIV-1 reverse transcriptase. Science **242:**1168–1171.
36. **Puers, C., H. A. Hammond, L. Jin, C. T. Caskey, and J. W. Schumm.** 1993. Identification of repeat sequence heterogeneity at the polymorphic short tandem repeat locus HUMTH01[AATG]n and reassignment of alleles in population analysis by using a locus-specific allelic ladder. Am. J. Hum. Genet. **53:**953–958.
37. **Robertson, D. L., B. H. Hahn, and P. M. Sharp.** 1995. Recombination in AIDS viruses. J. Mol. Evol. **40:**249–259.
38. **Robertson, D. L., P. M. Sharp, F. E. McCutchan, and B. H. Hahn.** 1995. Recombination in HIV-1. Nature **374:**124–126.
39. **Sabino, E. C., E. G. Shpaer, M. G. Morgado, B. T. M. Korber, R. S. Diaz, V. Bongertz, S. Cavalcante, B. Galvao-Castro, J. I. Mullins, and A. Mayer.** 1994. Identification of human immunodeficiency virus type 1 envelope genes recombinant between subtypes B and F in two epidemiologically linked individuals from Brazil. J. Virol. **68:**6340–6346.
40. **Salminen, M., A. Nykanen, H. Brummer-Korvenkontio, M. L. Kantanen, K. Liitsola, and P. Leinikki.** 1993. Molecular epidemiology of HIV-1 based on phylogenetic analysis of *in vivo gag* p7/p9 direct sequences. Virology **195:** 185–194.
41. **Salminen, M. O., J. K. Carr, D. S. Burke, and F. E. McCutchan.** 1995. Identification of breakpoints in intergenotypic recombinants of HIV-1 by bootscanning. AIDS Res. Hum. Retroviruses **11:**1423–1425.
42. **Salminen, M. O., B. Johansson, A. Sonnerborg, S. Ayehunie, D. Gotte, P. O. Leinikki, D. S. Burke, and F. E. McCutchan.** 1996. Full-length sequence of an Ethiopian human immunodeficiency virus type-1 (HIV-1) isolate of genetic subtype C. AIDS Res. Hum. Retroviruses **12:**1329–1339.
43. **Salminen, M. O., C. Koch, E. Sanders-Buell, P. K. Ehrenberg, N. L. Michael, J. K. Carr, D. S. Burke, and F. E. McCutchan.** 1995. Recovery of virtually full-length HIV-1 provirus of diverse subtypes from primary virus cultures using the polymerase chain reaction. Virology **213:**80–86.
44. **Stuhlmann, H., and P. Berg.** 1992. Homologous recombination of copackaged retrovirus RNAs during reverse transcription. J. Virol. **66:**2378–2388.
45. **Thompson, J. D., D. G. Higgins, and T. J. Gibson.** 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. **22:**4673–4680.
46. **Tumas, K. M., J. M. Poszgay, N. Avidan, S. J. Ksiazek, B. Overmoyer, K. J. Blank, and M. B. Prystowsky.** 1993. Loss of antigenic epitopes as the result of env gene recombination in retrovirus-induced leukemia in immunocompetent mice. Virology **192:**587–595.
47. **Vanden Haesevelde, M., J. L. Decourt, R. J. De Leys, B. Vanderborght, G. van der Groen, H. van Heuversvin, and E. Saman.** 1994. Genomic cloning and complete sequence analysis of a highly divergent African human immunodeficiency virus isolate. J. Virol. **68:**1586–1596.
48. **Voevodin, A., K. A. Crandall, P. Seth, and S. Al Mufti.** 1996. HIV type 1 subtypes B and C from new regions of India and Indian and Ethiopian expatriates in Kuwait. AIDS Res. Hum. Retroviruses **12:**641–643.
49. **Yang, Y. L., G. Q. Wang, K. Dorman, and A. H. Kaplan.** 1996. Long polymerase chain reaction amplification of heterogeneous HIV type 1 templates produces recombination at a relatively high frequency. AIDS Res. Hum. Retroviruses **12:**303–306.
50. **Zhu, T., N. Wang, A. Carr, S. Wolinsky, and D. D. Ho.** 1995. Evidence for coinfection by multiple strains of human immunodeficiency virus type 1 subtype B in an acute seroconvertor. J. Virol. **69:**1324–1327.