Virology **269,** 95–104 (2000)
doi:10.1006/viro.2000.0214, available online at http://www.idealibrary.com on IDEAL®

Virtually Full-Length Subtype F and F/D Recombinant HIV-1 from Africa and South America

Tiina Laukkanen,* Jean K. Carr,† Wouter Janssens,‡ Kirsi Liitsola,* Deanna Gotte,† Francine E. McCutchan,† Eline Op de Coul,§ Marion Cornelissen,¶ Leo Heyndrickx,‡ Guido van der Groen,‡ and Mika O. Salminen*,[1]

*HIV Laboratory, Department of Infectious Disease Epidemiology, National Public Health Institute, Mannerheimintie 166, FIN-00300 Helsinki, Finland; †Henry M. Jackson Foundation, Rockville, Maryland 20850; ‡Department of Microbiology, Institute of Tropical Medicine, Antwerp, Belgium; §Division of Public Health and Environment, Municipal Health Service, Nieuwe Achtergracht 100, 1018 WT Amsterdam, The Netherlands; and ¶Department of Human Retrovirology, Academic Medical Centre, Amsterdam, The Netherlands

Received November 8, 1999; accepted January 21, 2000

For reliable classification of HIV-1 strains appropriate reference sequences are needed. The HIV-1 genetic subtype F has a wide geographic spread, causing significant epidemics in South America, Africa, and some regions of Europe. Previously only two full-length sequences of each of the HIV-1 subtype F subclusters F1 and F2 have been described. To extend the knowledge of subtype F variation on a complete genome level, three new virtually full-length F1 sequences were cloned and sequenced, two from Africa and one from South America. Comparison of the new and previously described sequences showed that monophyletic clustering of the subcluster F1 of subtype F is consistent and highly supported in all genome regions. Two additional full-length strains were shown to be mosaics of subtypes F and D. These epidemiologically unrelated F/D sequences showed similar chimeric structure, suggesting that they may represent a previously undescribed circulating recombinant form (CRF). This was supported by partial sequences from three additional unlinked F/D recombinants. Genetic distances in the phylogenetic trees suggest that the recombination event leading to the putative CRF occurred relatively long ago, close to the divergence of the F1 and F2 subclusters. Furthermore, all five F/D recombinants are linked to the Democratic Republic of Congo, suggesting that the original recombination event took place in central Africa.

INTRODUCTION

A high degree of variation among HIV-1 strains has resulted in the establishment of multiple parallel evolutionary lineages. Strains of HIV-1 are phylogenetically divided into three groups, M (major), O (outlier), and N (non-M-non-O), of which the M group viruses are most widespread and form the majority of the HIV/AIDS pandemic. The M group can be further divided into several different subtypes (A–K) (Korber *et al.,* 1998; Triques *et al.,* 2000). Subtype F is furthermore subdivided into two subclusters, or subsubtypes, designated F1 and F2. For most subtypes, there are now a sufficient number of complete genome sequences available to demonstrate that the phylogenetic subtype classification is consistent in all parts of the genome.

Subtype F strains have been found in significant numbers in South American countries such as Brazil and Argentina (Potts *et al.,* 1993; Morgado *et al.,* 1994; Sabino *et al.,* 1996, Louwagie *et al.,* 1994; Gao *et al.,* 1996b; Lukashov *et al.,* 1996; Marquina *et al.,* 1996; Campodonico *et al.,* 1996), in Africa in the Democratic Republic of Congo (DRC) and Cameroon (Nkengasong *et al.,* 1994; Lukashov *et al.,* 1996), and in Europe in Romania, Russia, Cyprus, and France (Dumitrescu *et al.,* 1994; Liitsola *et al.,* 1996, Kostrikis *et al.,* 1995; Simon *et al.,* 1996). See also the HIV Sequence Database (http://hiv-web.lanl.gov; Korber *et al.,* 1998). The Cameroonian isolates and one strain from the DRC form the F2 subcluster within the F cluster, whereas the other strains fall into the F1 subcluster (Nkengasong *et al.,* 1994; Lukashov *et al.,* 1996; Triques *et al.,* 1999, 2000; Peeters *et al.,* 1998).

However, as with most HIV-1 sequences (independent of the subtype) available in GenBank, the majority of the available subtype F sequences are short sequence fragments (usually *env* or *gag*). So far only two full-length strains of subtype F1 and two of subtype F2 have been cloned and sequenced (Gao *et al.,* 1998b; Triques *et al.,* 2000). Often HIV-1 strain subtypes are determined by sequencing only one or a few short segments of the genome. Such a typing strategy overlooks the fact that in addition to evolving by accumulation of point mutations, several studies conducted in recent years have shown that recombination between subtypes may be quite common (Robertson *et al.,* 1995a).

Piecemeal sequencing may not accurately describe the genetic makeup of a virus, since the recombination events might have occurred elsewhere in the genome than in the sequenced parts. This has frequently resulted in a strain that was originally classified as belonging to a single subtype later proving to be a mosaic of two or

[1] To whom correspondence and reprint requests should be addressed. Fax: +358-9-4744 8461. E-mail: Mika.Salminen@ktl.fi.

0042-6822/00 $35.00

more subtypes after a full-genome sequence was produced (Robertson *et al.,* 1995b; Carr *et al.,* 1996; Gao *et al.,* 1996a,b, 1998a).

In addition to reports of independent unique chimeric strains, some epidemics caused by recombinant HIV-1 strains have been encountered (Carr *et al.,* 1996; 1998; Gao *et al.,* 1996a; Liitsola *et al.,* 1998). In these cases the recombinants behave like the genetic subtypes, forming monophyletic clusters in phylogenetic analyses. Such epidemic recombinants form a significant contribution to the molecular epidemiology of HIV and have recently been given the collective designation of a circulating recombinant form (CRF). Currently three such CRFs have been described: the AE_{CM240} CRF prevalent in Southeast Asia, the AG_{IbNG} CRF prevalent in western and central Africa, and the AB_{KAL153} CRF, which is found in several Eastern European countries (Liitsola *et al.,* 1998, submitted for publication, a). The common factor in all CRFs described to date is that one of the parental viruses belongs to subtype A. No CRFs that would not include subtype A-derived regions have yet been reported.

To allow a better definition of the F subtype and recombinants involving subtype F we have cloned and sequenced three virtually complete subtype F viruses and two F/D recombinants and compared them to existing sequences. The two F/D recombinants were also compared to epidemiologically unlinked F/D recombinant strains from The Netherlands and Belgium (Op de Coul *et al.,* 1999).

RESULTS

Full-length clones

The lengths of the cloned sequences were as follows: F9363, 8925 bp; BZ126, 8962 bp; VI850, 8945 bp; VI1310, 9083 bp; and VI961, 9005 bp. The open reading frames for the genes were intact in all sequences, except in VI961, in which *pol* was prematurely truncated due to a frameshift mutation at position 3696, and in F9363, due to a point mutation resulting in a premature stop codon at 1473 in *pol.* All strains had two NFκB sites, three SP-1 sites, a normal TATA box, and a typical 3-nt bulge in the TAR stem region. However, the secondary structure prediction for the TAR top loop for one of the isolates (VI850) indicated a 4-nt top loop instead of the normal 6-nt loop. Whether this has any consequences for the biological properties of the virus is uncertain. There was some minor variation in the length of *gag* products in some of the strains compared to what is usually seen. In BZ126 and VI850 the Gag protein was six amino acids longer than it usually is in subtype F strains.

Phylogenetic analyses

The phylogenetic tree based on full-length sequences shows the clustering of the subtype F strains together with the F1 reference strain BR020 as well as the independent clustering of the F/D mosaic strains (Fig. 1). A recently published F1 sequence (Triques *et al.,* 2000) was not included in our analyses, but preliminary analysis indicates that it is a member of the F1 subtype. The bootstrap values supporting subtype and CRF clades and intraclade subclusters (AG_{IbNG} and AE_{CM240} CRFs, the F1 and F2 subclusters within the F clade) were significant.

Detailed analysis of the newly derived complete genome sequences

Similarity plotting and bootscanning analysis of the newly derived BZ126, F9363, and VI850 complete genome sequences showed them to be most closely related to the BR020 subtype F reference sequence in all genome regions and therefore verified their classification as new reference strains for subtype F subcluster F1 (Fig. 2). A similar analysis of the virtually full-length sequences of isolates VI1310 and VI961 revealed them to be F/D mosaics and to possess multiple recombination sites throughout the genome. Figure 3 compares the similarity and bootscan plots. The locations of the breakpoints in the two sequences are shown. A short gap region in the beginning of the sequence in the similarity plots represents a region for which no genetic data were available (due to the PCR strategy). The sequence of the 5′ LTR U3 was not present in the cloned provirus but was reconstructed based on the 3′ LTR sequence for the purpose of clarity. The breakpoints in VI1310 and VI961 are mostly similar but some slight discrepancies can be observed.

The sequences were next broken into 12 independent fragments corresponding to the inferred recombined regions and separately analyzed phylogenetically (Fig. 4). This analysis confirmed the inferred breakpoints for most regions. However, the fragment corresponding roughly to the *pol* p10 protease coding region—the third fragment in Fig. 3—could not be verified to belong to subtype F by this analysis (tree not shown), although the bootscan seems to indicate such a result. In the similarity plot the region was unclassified, and it is possible that it is too short to be analyzed reliably. In some of the trees the bootstrap values are probably affected by placing of the other sequences. For example, strain UG114 seems to cause a lowering of the value for the D cluster in Fig. 4C (the value is 71% without UG114; not shown). This might reflect the close relationship between the B and the D subtypes. Also, in some other fragments the short length as well as the inconsistent clustering of strains representing different clades leads to low bootstrap values. An example of this is the *vif–vpr* region, which is problematic in subtypes A and G and the AE_{CM240} CRF that tend to form a supercluster in this region (Carr *et al.,*

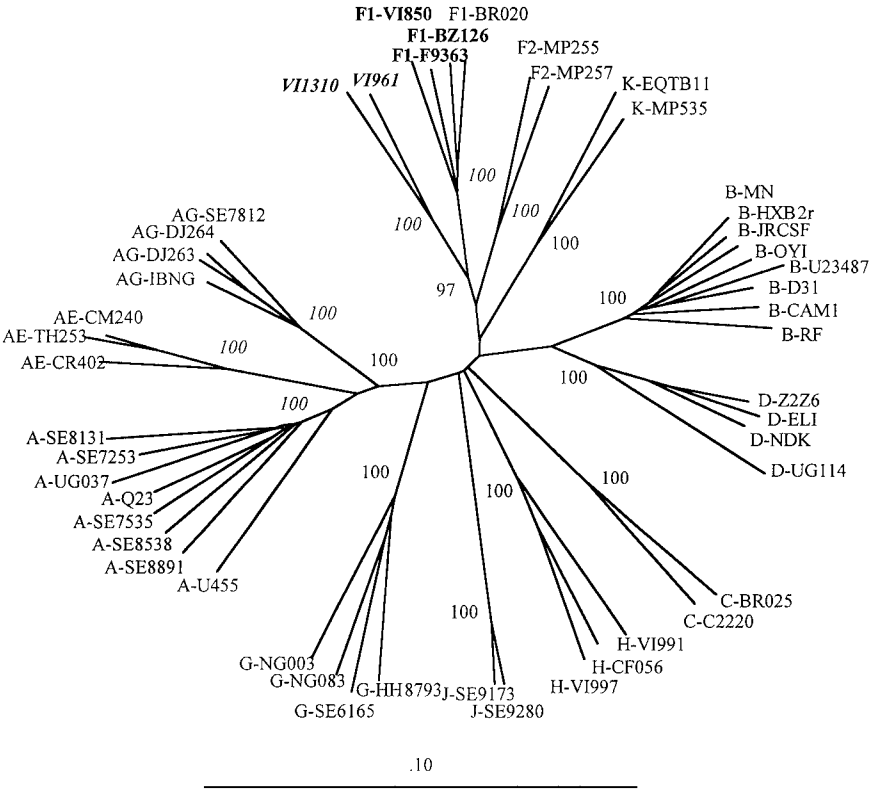


FIG. 1. Phylogenetic analysis based on full-length sequences. The letter specifying the clade is shown in front of the isolate name. The new full-length F subtype sequences are shown in boldface type. The F/D mosaic sequences are shown in boldface italics. Bootstrap values for each clade and intraclade subclusters (italics) are shown next to the nodes. Method used: DNADIST Kimura two-parameter, 100 replicates.

1998). This explains the cross-clustering of subtypes A and G in Figs. 4G and 4H.

The seventh fragment (the end of *pol* and the beginning of *vif*) was supported by a low bootstrap value (65%) in the phylogenetic analysis (Fig. 4F). The bootscan picture suggests that the region might have a short fragment of subtype D in the middle (particularly in the isolate VI1310, Fig. 3B), although this could not be confirmed by individual fragment phylogenetic analysis. Using the split decomposition analysis as implemented in Splitstree supported the classification of this region as recombinant (not shown). However, due to the low bootstrap value and uncertain similarity plot the whole fragment is shown as "unclassified" in the genome picture in the middle (striped). The eighth and ninth fragments were analyzed phylogenetically separately because in VI1310 they are different from each other: the eighth fragment is classified as subtype D, whereas the ninth remains unclassified (shown as striped; this isolate was excluded from Fig. 4H). In Figs. 3C and 3D the corresponding fragments (fragments 8 and 9) of the sequence VI961 are shown as a contiguous region belonging to subtype D.

In summary, the LTR R and U3 region of isolates VI1310 and VI961 consists of subtype F and *gag* of subtype D. The *pol* region consists of intervening subtype D and F fragments with some unclassified regions. *Vif* is unclassified and D in both strains, but although *vpr* and the first exon of *tat* are also D in the strain VI961, they remain unclassified in VI1310. *Vpu* and *env* are F in both strains except for a small fragment of D in the gp41 region (*tat* and *rev* second exons and the region coding for the transmembrane part of gp41). *Nef* is derived from subtype F. In short, these F/D mosaics seem to consist of close to equal proportions of subtypes F and D.

Comparative analysis of five distinct F/D mosaics in *gag–pol* and *env* regions

The isolates BEVI1206 and BEVI1267 were analyzed earlier by *env* sequencing and were classified as F (Heyndrickx *et al.,* 1998). BEVI1206 was obtained from a Belgian woman who probably was infected in the DRC and BEVI1267 was obtained from a Belgian woman who acquired the virus from a seropositive partner from the DRC. Later, these strains were analyzed in the *gag–pol* region along with another strain, R890820, and all were found to be structurally similar F/D mosaics (Op de Coul *et al.,* 1999). R890820 was isolated from a Dutch man who had a female partner from the DRC. No epidemiological link was evident between the three cases based on available background information.

Sequences from these individuals were compared to the newly obtained full-length F/D sequences, VI1310

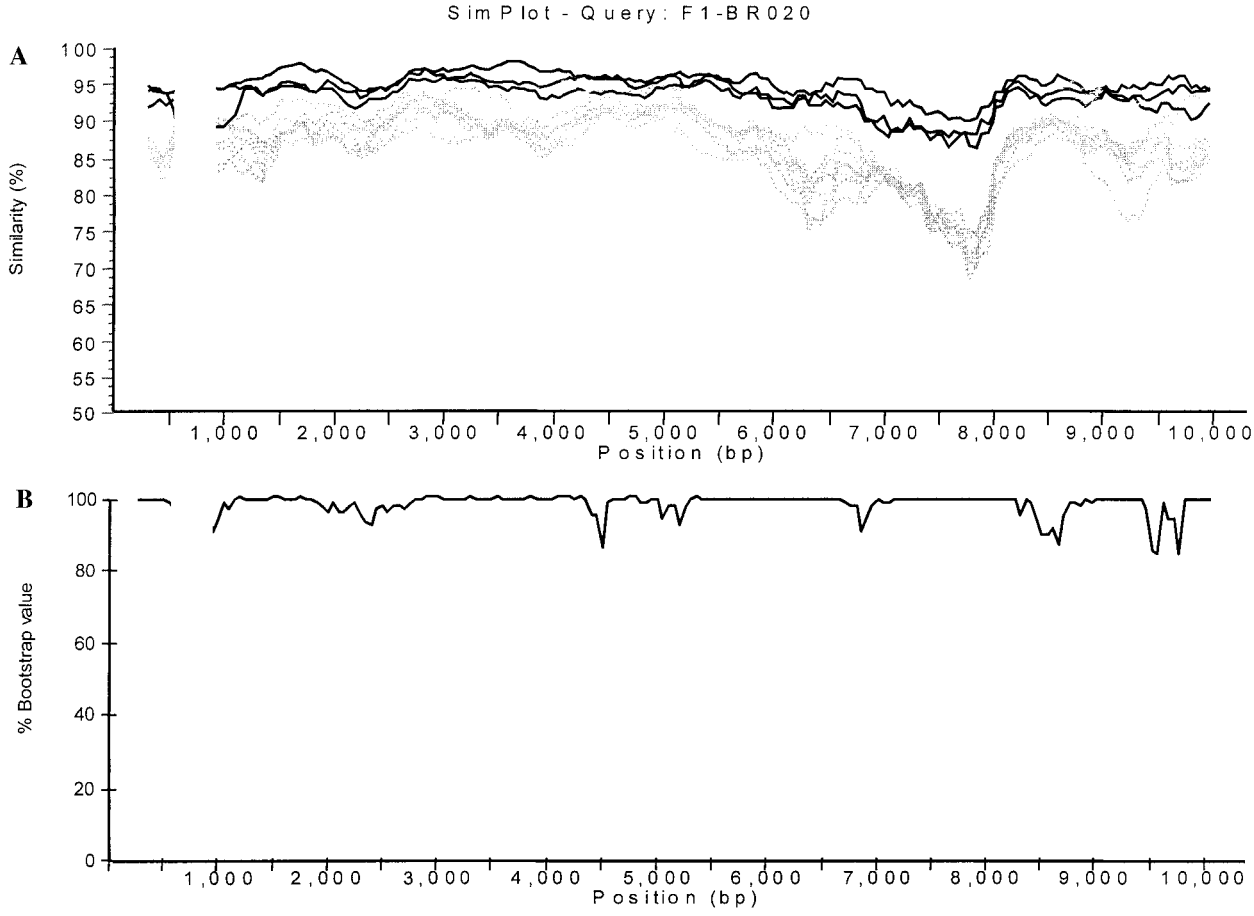


FIG. 2. Similarity plot (A) and bootscan (B) of the three new subtype F subgroup F1 complete genomes compared to the consensus reference sequences. Similarity plots and bootscanning were performed as described under Materials and Methods, except that a window of 600 bp was employed for reasons of clarity. Positions indicate alignment coordinates. (A) The BR020 reference was compared to the newly derived F1 sequences (solid black lines) and to consensus reference sequences of subtypes A–K (gray dotted lines). (B) The bootstrap value of the F1 cluster of four sequences was analyzed along an alignment of all reference sequences excluding F2 using neighbour-joining analysis. This was performed using the bootscanning package for Linux.

and VI961, in the corresponding genomic regions (*env* V3 and *gag–pol*). The clustering of all five sequences is similar in all regions analyzed (Figs. 5A–5C). The breakpoint in the *gag–pol* region where the sequence changes from D to F was found to be located at the same site in all five sequences (Fig. 5D; position 1980–2722 in HXB2r). The fragment may contain two other possible recombination sites, first at position 1340, which corresponds to 2148 in HXB2r (from D to F), and again at position 1420, which corresponds 2192 in HXB2r (from F to D), but there is some minor variation in the exact locations (see Fig. 5D). However, this region is too short to allow for reliable interpretation. The results indicate that all five strains are likely to represent members of the same recombinant HIV-1 evolutionary lineage.

DISCUSSION

This article describes the first full-length F/D intersubtype recombinant isolates and extends the known variation of the F subtype subcluster F1 by providing three new virtually full-length reference sequences originating in Kenya, Brazil, and the DRC. The full-length sequences of all these strains belong to the F1 subcluster in all genomic regions and consistently cluster with the previously described full-length subcluster F1 reference strain BR020. The phylogenetic structure of the F clade, and the F1 subcluster in particular, is now on a more solid basis than before. Recent work by Triques *et al.* (2000), which demonstrated that subtype F does not form a homogenous clade, but is divided in two distinct subclusters, is further supported by the analysis presented here.

In addition, two full-length sequences with an intersubtype mosaic pattern consisting of intervening fragments of subtypes F and D were described and their pattern of recombination and relation to other previously described recombinant viruses of subtypes F and D was analyzed. In all the analyses, the virtually complete genome sequences of the two F/D recombinant strains

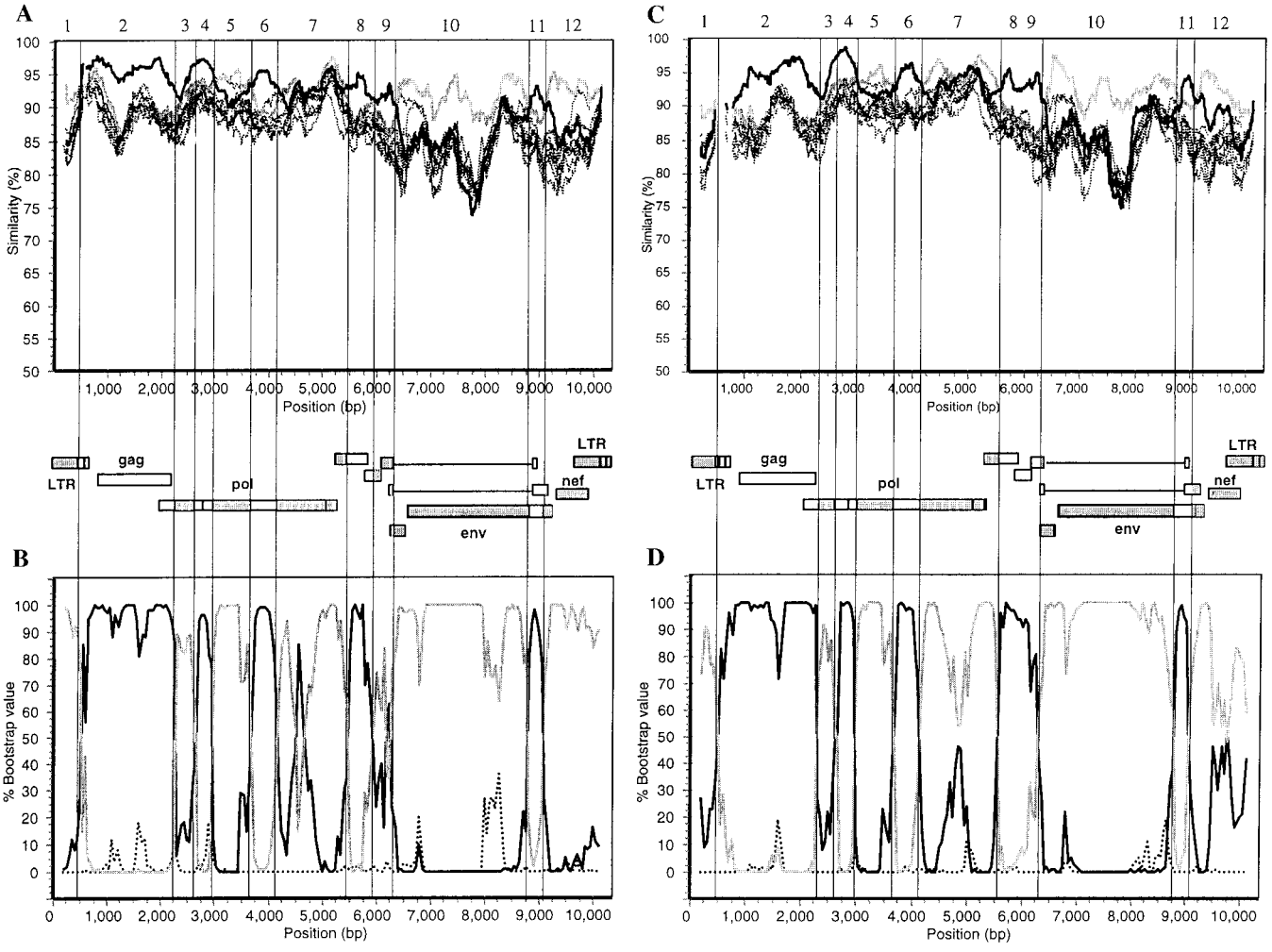
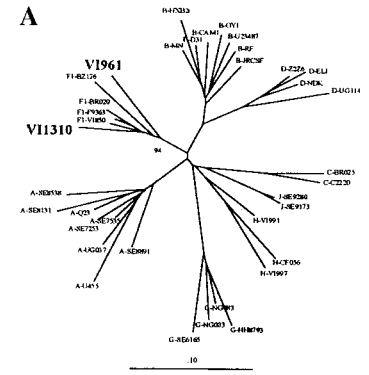


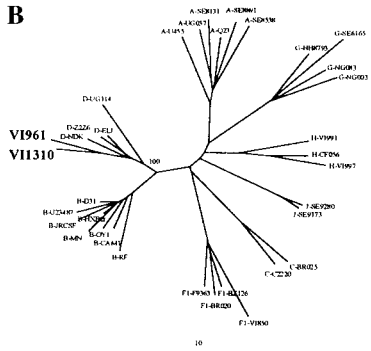
FIG. 3. Similarity plot and bootscan analyses of the new full-length F/D mosaics. Black solid line depicts subtype D, gray line depicts subtype F. In similarity plots the subtypes are shown as dotted lines, in bootscan pictures the outgroup (subtype C) is shown as a dotted line. The locations of the breakpoints in the genome are shown in the middle as in HXBr2. Gray regions depict subtype F, white regions depict subtype D, striped regions represent unclassified regions or regions whose classification was uncertain. (A and B) Analyses of the strain VI1310. (C and D) Analyses of the strain VI961.

resemble one another to a great extent. The genome segment patterns and breakpoint locations are very similar (Fig. 3). A common source of infection was first considered based on the similar genetic structure of the isolates, but there was no evidence for an epidemiological link between the individuals. On the contrary, the background information suggested that the infections were independent. However, the original source of the virus could be traced to the DRC in all cases. This brings up the possibility that the viruses represent an evolutionary lineage that has been established from a point source since the original recombination event occurred. The likely location of this event would be Central Africa and possibly even the DRC, where multiple subtypes (including F and D) are known to circulate (Peeters *et al.,* 1998). The DRC has been reported to have a relatively high prevalence (5.3%) of subtype D compared to many other African countries. The prevalence for subtype F is also high (8.9%) in the DRC as well as in neighboring countries such as Cameroon (14.4%) and Gabon (3.2%). It can be predicted that in a country with several genetic subtypes of HIV-1 circulating, the chance of a dual infection with two different subtypes is high and thus the likelihood for generating intersubtype recombinant strains is also high.

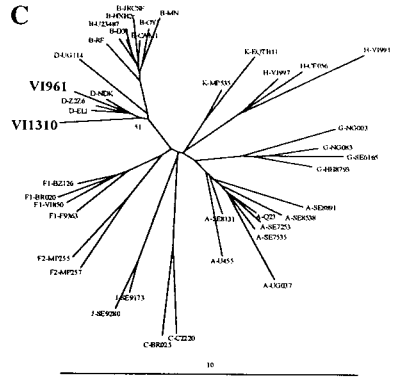
Another case of three independent, yet highly similar F/D mosaic isolates found by others supports the existence of an F/D evolutionary lineage originating in Central Africa (Op de Coul *et al.,* 1999). In this study all of the isolates also could be traced back to the DRC. In comparisons of the *gag–pol* and V3 sequences from these strains and the newly described F/D mosaics all sequences were found to be highly related. In phylogenetic analysis, some variation between the five F/D strains was found, but it is of a magnitude similar to that seen within the other CRF subclusters, in particular the AE_{CM240}



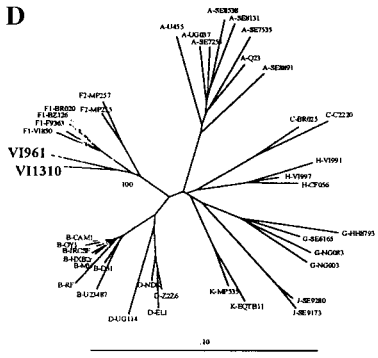
1: 1-357



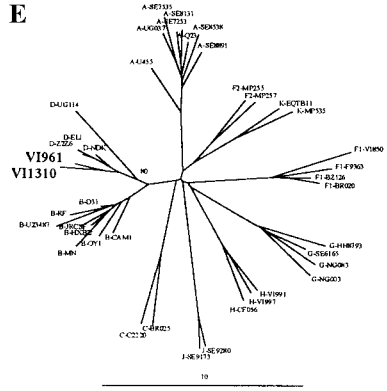
2: 358-2122



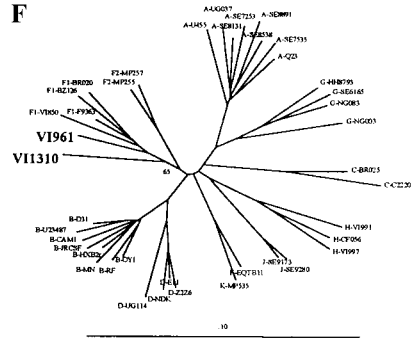
3: 2230-2715



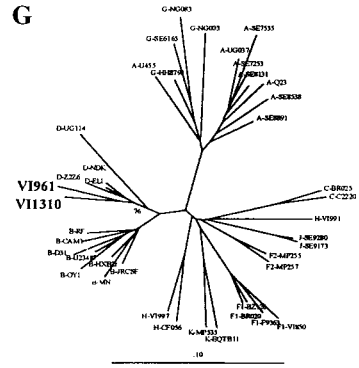
4: 2716-3435



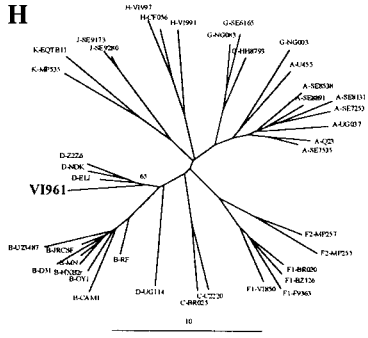
5: 3436-3912



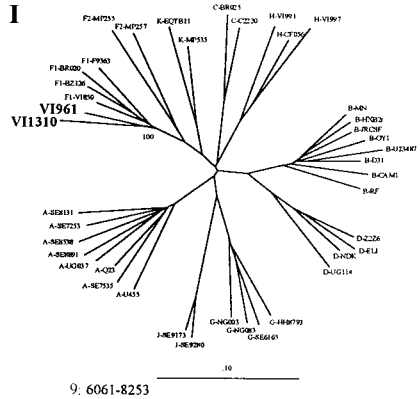
6: 3913-5254



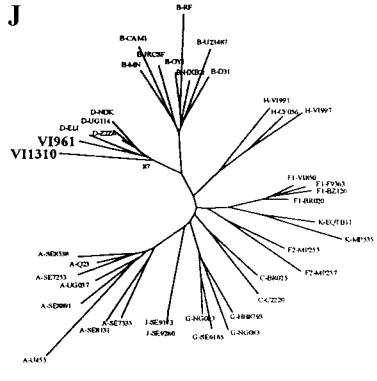
7: 5255-5654



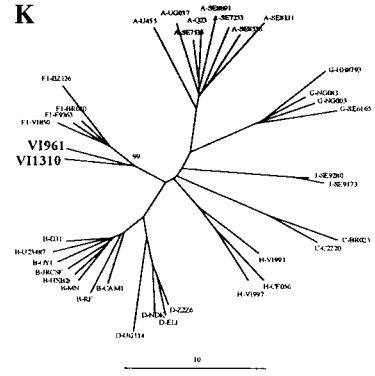
8: 5655-6053



9: 6061-8253



10: 8254-8572



11: 8591-9491

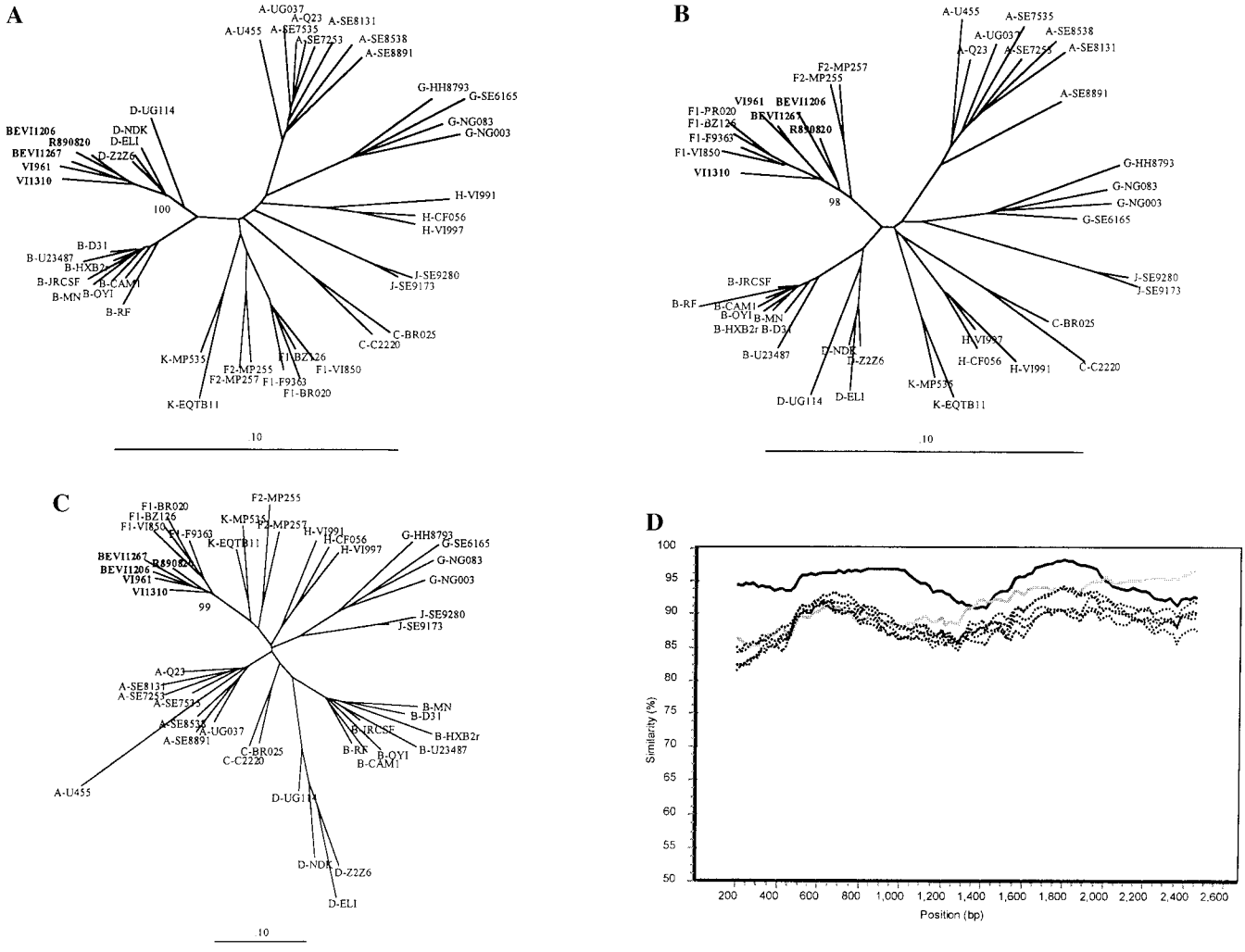


FIG. 5. Phylogenetic analyses comparing the F/D mosaic isolates BEVI1206, BEVI1267, R890820, VI961, and VI1310. (A) An analysis of a fragment of 1980 bp from the *gag–pol* region (location in HXBr2: 859–2838). (B) An analysis of the next 520 bp (location in HXBr2: 2839–3416). (C) An analysis of an *env* fragment of 260 bp (location in HXBr2: 7032–7301). Bootstrap values for the specific clades are shown next to the node. Bootstrap values for the other subtypes are not shown, but were significant. (D) A similarity plot analysis of the *gag–pol* region using the five F/D mosaics as a consensus query. Black line depicts subtype D, gray line depicts subtype F. The breakpoint is located at around base 1980 (location 2838 in HXBr2). Other subtypes are shown by dotted lines.

and AG_{IbNG} lineages (Fig. 1). The independent epidemiological origins of the isolates and the high level of similarity between them coupled with a monophyletic clustering in phylogenetic trees would suggest that this F/D mosaic is a representative of a fairly prevalent evolutionary lineage, which might in the future be classified as a new CRF. However, the formal requirement for assigning a new CRF is the existence of at least three epidemiologically independent complete genome sequences that share their recombinant structure and form a monophyletic cluster in all regions of the genome. Since we have sequenced only two complete genomes, the formal clas-

FIG. 4. Phylogenetic analyses of individual fragments (numbered according to Fig. 3) of the F/D mosaic strains based on recombination points in the genome. Recombinant sequences are shown in large font. The bootstrap value supporting the clustering is indicated next to the node. Some of the reference sequences were excluded from the first two analyses and from the last analysis because no sequence data from those regions were available. (A) 5′ LTR U3. (B) *gag* (p17, p24, p7, p6). The fragment corresponding to *pol* p10 was too short to be analyzed reliably and thus is not represented here. (C) The beginning of *pol* p51. The bootstrap value for the D cluster is 71% without D-UG114 (data not shown). (D) The middle region of *pol* p51. (E) The end of p51 and the first half of the RNaseH region. (F) The end of *pol* and the first half of *vif.* The bootstrap value for the F1 subcluster (excluding F2) is 84% (data not shown). (G) The end of *vif* region and the first half of *vpr.* (H) The end of *vpr* and the first exons of *tat* and *rev.* The value for the D cluster is given without D-UG114. F/D isolate VI1310 was excluded because its subtype cannot be determined in this region. (I) *Vpu,* 4/5 of *env.* (J) Part of *env* gp41. (K) The end of *env* gp41, *nef,* 3′ LTR. Genomic locations represented as in HXB2r. Between some of the analyzed fragments, short gap regions were excluded.

sification of the FD CRF will have to await sequencing of another similar virus isolate.

In summary the findings of this study extend the information available for subtype F viruses, verify their division into two distinct subgroups, F1 and F2, and provide evidence for the possible existence of the first CRF that does not contain subtype A-like sequences. The results increase our understanding of HIV-1 molecular epidemiology and variation and may have implications for the design of future vaccines. If HIV subtypes must be considered in vaccine development, CRFs are in the same position as the subtypes and therefore information on their prevalence is as valuable as that of the subtypes.

MATERIALS AND METHODS

Viral isolates

Full-length clones were produced and completely sequenced from five short-term PBMC coculture isolates (isolated according to UNAIDS Virus Isolation Guidelines) from Belgium (isolates VI1310, VI961, and VI850, all linked to the DRC), Finland (F9363 linked to Kenya), and Brazil (BZ126). The isolates had been previously classified as members of subtype F based on phylogenetic analyses of short fragments or by heteroduplex mobility assay (Delwart *et al.,* 1993; Louwagie *et al.,* 1994; Heyndrickx *et al.,* 1998; Van de Peer *et al.,* 1996; Liitsola *et al.,* submitted for publication, b). The isolate VI1310 was obtained from a Belgian woman, reportedly infected by a man who frequently traveled to the DRC. The individual infected with VI961 was the partner of a seropositive man from the DRC. The isolate VI850 is from a Belgian man infected by his wife, who acquired the infection in the DRC. The isolate F9363 is from a Finnish man, who most likely was infected by a Kenyan woman. Strain BZ126 is from a Brazilian individual infected in Brazil. Total cellular DNA was isolated from the cells of the PBMC HIV coculture at the peak of HIV antigen production using a commercially available kit (Qiagen Blood and Cell Culture Kit, Qiagen GmbH, Hilden, Germany).

PCR and cloning of the amplified fragments

PCR amplification and cloning were performed essentially as previously described (Salminen *et al.,* 1995a). Briefly, virtually full-length genomes were amplified using the Expand Long Template PCR Kit (Boehringer Mannheim, Mannheim, Germany) using buffer 2 since it turned out to give optimal amplification and cloning results. Primers MSF12 5′-AAATCTCTAGCAGTGGCGCCCGAACAG-3′ (location in HXB2r, 623–649, includes a unique *Bbe*I site) and MSR5 5′-GCATGCGCCCTCAAGGCAAGCTTTATTGAGGCT-3′ (location in HXB2r, 9606–9638, includes a unique *Bgl*I site) were used. These primers amplify a region of 9003 bp in HXB2r. In some cases the first-round PCR was not successful, so an additional seminested PCR was carried out with primers MSF14, 5′-TCTCTCGACGCAGGACTCGGCTTG-3′ (location in HXB2r, 682–705), and MSR5, which amplify a fragment of 8945 bp. The first-round primers are complementary to the tRNA primer binding site, located at the 5′ end of the genome just before the *gag* leader region, and to the polyadenylation site at the 3′ end, which is located between the R and the U5 regions of the 3′ LTR.

The thermocycle program for the first-round PCR was 94°C for 2 min, 94°C for 10 s, 60°C for 30 s, 68°C for 8 min, repeated for 9 additional cycles starting from step 2; 94°C for 10 s, 55°C for 30 s, 68°C for 8 min, repeated 19 times, 72°C for 30 min (end segment). Total cellular DNA (50–250 ng) from HIV-infected cells in a volume of 1 μl TE buffer (10 mM Tris, 1 mM EDTA, pH 8.0) was used as template in a PCR of 50 μl (or in the case of the seminested PCR, 1 μl of the first-round product), after which the products of 10 successful repeat reactions were pooled and the amplified DNA was gel-purified using the QIAquick Gel Extraction Kit (Qiagen GmbH). All other reaction components were assembled according to the recommendations of the manufacturer. Purified DNA was TA-cloned using the pCR 2.1 vector (Invitrogen, San Diego, CA) following the manufacturer's instructions, but it was incubated at 15°C overnight in the ligation reaction. Plasmids were heat-shock transformed into STBL-2 Max Efficiency cells (Gibco BRL Life Technologies, Inchinnan, Scotland). Clones were chosen for further studies based on positive hybridization with radioactively labeled probes and later by digestion with restriction enzymes. Plasmid DNA was purified and prepared with the Qiagen Plasmid Maxi Kit (Qiagen GmbH).

Full-length clones were sequenced completely by primer walking using cycle-sequencing with dye-deoxy terminators and an ABI 377 automatic sequencer (Perkin–Elmer Applied Biosystems Division, Foster City, CA).

Phylogenetic analyses

New full-length sequences were aligned with and compared to existing full-length reference sequences. These included the following sequences: subtype A, (UG037, U455, SE7253, SE7535, Q23, SE8131, SE8891, SE8538), CRF AE_{CM240} (TH253, CM240, CR402), CRF AG_{IbNG} (IbNG, DJ263, DJ264, SE7812); subtype B (CAM1, RF, OYI, D31, U23487, HXB2r, MN, JRCSF); subtype C (BR025, C2220); subtype D (ELI, NDK, Z2Z6, UG114); subtype F, subcluster F1 (BR020), subcluster F2 (MP255, MP257); subtype G (HH8793, G6165, NG083, NG003); subtype H (CF056, VI991, VI997); subtype J; and subtype K (EQTB11, MP535).

Phylogenetic analyses were performed and bootstrap values (100 replicates) calculated for each subtype branch using ClustalW applying the Kimura two-parameter model of evolution, or the SEQBOOT, DNADIST (with the F84 maximum-likelihood model), and NEIGHBOR or

FITCH programs of the Phylip software package (version 3.572c, Felsenstein, 1981, 1991). In order to determine the approximate breakpoints in mosaic sequences, bootscanning and similarity plot analyses were carried out with SimPlot (version 2.5) in a sliding window of 400 (or 600) bases (Salminen *et al.,* 1995b; Ray, 1998/1999). Based on the inferred breakpoint locations, the sequence alignment was broken into corresponding fragments that were phylogenetically analyzed individually. Some regions for which the above methods did not unambiguously provide classification were further analyzed using the program Splitstree (version 2.1.1.), which in contrast to most phylogenetic software is able to show alternative tree topologies simultaneously (Huson, 1998).

Sequence accession numbers

The full-length sequences of VI1310, VI961, VI850, F9363, and BZ126 have been deposited with GenBank under accession numbers AF193253, AF076998, AF077336, AF075703, and L22083, respectively.

ACKNOWLEDGMENTS

The work presented here was supported partly by the Henry M. Jackson Foundation for the Advancement of Military Medicine and the UNAIDS Virus Characterization Network, as well as the European Union EASP-Concerted Action Network.

REFERENCES

Bianchard, A., Ferris, S., Chamaret, S., Guétard, D., and Montagnier, L. (1998). Molecular evidence for nosocomial transmission of human immunodeficiency virus from a surgeon to one of his patients. *J. Virol.* **72,** 4537–4540.

Campodonico, M., Janssens, W., Heyndrickx, L., Fransen, K., Leonaers, A., Fay, F. F., Taborda, M., van der Groen, G., and Fay, O. H. (1996). HIV type 1 subtypes in Argentina and genetic heterogeneity of the V3 region. *AIDS Res. Hum. Retroviruses* **12,** 79–81.

Carr, J. K., Salminen, M. O., Koch, C., Gotte, D., Artenstein, A. W., Hegerich, P. A., St. Louis, D., Burke, D. S., and McCutchan, F. E. (1996). Full-length sequence and mosaic structure of a human immunodeficiency virus type 1 isolate from Thailand. *J. Virol.* **70,** 5935–5943.

Carr, J. K., Salminen, M. O., Albert, J., Sanders-Buell, E., Gotte, D., Birx, D. L., and McCutchan, F. E. (1998). Full genome sequences of human immunodeficiency virus type 1 subtypes G and A/G intersubtype recombinants. *Virology* **247,** 22–31.

Delwart, E. L., Shpaer, E. G., Louwagie, J., McCutchan, F. E., Grez, M., Rübsamen-Waigmann, H., and Mullins, J. I. (1993). Genetic relationships determined by a DNA heteroduplex mobility assay: Analysis of HIV-1 *env* genes. *Science* **262,** 1257–1261.

Dumitrescu, O., Kalish, M. L., Kliks, S. C., Bandea, C. I., and Levy, J. A. (1994). Characterization of human immunodeficiency virus type 1 isolates from children in Romania: Identification of a new envelope subtype. *J. Infect. Dis.* **169,** 281–288.

Felsenstein, J. (1981). Evolutionary tree from DNA sequences: A maximum likelihood approach. *J. Mol. Evol.* **17,** 368–376.

Felsenstein, J. (1991). "PHYLIP Manual," Version 3.4, University Herbarium, University of California, Berkeley, CA.

Gao, F., Robertson, D. L., Morrison, S. G., Hui, H., Craig, S., Decker, J., Fultz, P. N., Girard, M., Shaw, G. M., Hahn, B. H., and Sharp, P. M. (1996a). 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.

Gao, F., Morrison, S. G., Robertson, D. L., Thornton, C. L., Craig, S., Karlsson, G., Sodroski, J., Morgado, M., Galvao-Castro, B., von Briesen, H., Beddows, S., Weber, J., Sharp, P. M., Shaw, G. M., Hahn, B. H., and The WHO and NIAID Networks for HIV Isolation and Characterization (1996b). Molecular cloning and analysis of functional envelope genes from human immunodeficiency virus type 1 sequence subtypes A through G. *J. Virol.* **70,** 1651–1667.

Gao, F., Robertson, D. L., Carruthers, C. D., Li, Y., Bailes, E., Kostrikis, L. G., Salminen, M. O., Bibollet-Ruche, F., Peeters, M., Ho, D. D., Shaw, G. M., Sharp, P. M., and Hahn, B. H. (1998a). An isolate of human immunodeficiency virus type 1 originally classified as subtype I represents a complex mosaic comprising three different group M subtypes (A, G, and I). *J. Virol.* **72,** 10234–10241.

Gao, F., Robertson, D. L., Carruthers, C. D., Morrison, S. G., Jian, B., Chen, Y., Barre-Sinoussi, F., Girard, M., Srinivasan, A., Abimiku, A. G., Shaw, G. M., Sharp, P. M., and Hahn, B. H. (1998b). A comprehensive panel of near-full-length clones and reference sequences for non-subtype B isolates of human immunodeficiency virus type 1. *J. Virol.* **72,** 5680–5698.

Heyndrickx, L., Janssens, W., Coppens, S., Vereecken, K., Willems, B., Fransen, K., Colebunders, R., Vandenbruaene, M., and van der Groen, G. (1998). HIV type 1 C2V3 *env* diversity among Belgian individuals. *AIDS Res. Hum. Retroviruses* **14,** 1291–1296.

Huson, D. H. (1998). SplitsTree: Analyzing and visualizing evolutionary data. *Bioinformatics* **14,** 68–73.

Korber, B., Hahn, B., Foley, B., Mellors, J. W., Leitner, T., Myers, G., McCutchan, F., and Kuiken, C. L. (1998). "Human Retroviruses and AIDS 1998: A Compilation and Analysis of Nucleic Acid and Amino Acid Sequences," http://hiv-web.lanl.gov. Theoretical Biology and Biophysics Group, Los Alamos National Laboratory, Los Alamos, NM.

Kostrikis, L. G., Bagdades, E., Cao, Y., Zhang, L., Dimitriou, D., and Ho, D. D. (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.

Liitsola, K., Laukkanen, T., Denisova, A., Grishkevichius, A., Smolskaja, T., Ustina, V., Vlasov, N., Leinikki, P., and Salminen, M. O. (1996). Genetic characterization of HIV-1 strains in the Baltic countries and Russia. *Scand. J. Infect. Dis.* **28,** 537–541.

Liitsola, K., Tashkinova, I., Laukkanen, T., Korovina, G., Smolskaja, T., Momot, O., Mashkilleyson, N., Chaplinskas, S., Brummer-Korvenkontio, H., Vanhatalo, J., Leinikki, P., and Salminen, M. O. (1998). HIV-1 genetic subtype A/B recombinant strain causing an explosive epidemic in injecting drug users in Kaliningrad. *AIDS* **12,** 1907–1919.

Liitsola, K., Holm, K., Bobkov, A., Pokrovsky, V., Smolskaya, T., Leinikki, P., Osmanov, S., Salminen, M., and the UNAIDS Virus Isolation Network An AB recombinant and its parental HIV-1 strains in the area of the former Soviet Union: Low requirements for sequence identity in recombination. Submitted for publication, a.

Liitsola, K., Laukkanen, T., Leinikki, P., and Salminen, M. Genetic variability of HIV-1 in Finland. Submitted for publication, b.

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

Louwagie, J., Delwart, E. L., Mullins, J. I., McCutchan, F. E., Eddy, G., and Burke, D. S. (1994). Genetic analysis of HIV-1 isolates from Brazil reveals presence of two distinct genetic subtypes. *AIDS Res. Hum. Retroviruses* **10,** 561–567.

Lukashov, V. V., Kuiken, C. L., Boer, K., and Goudsmit, J. (1996). HIV type 1 subtypes in the Netherlands circulating among women originating from AIDS-endemic regions. *AIDS Res. Hum. Retroviruses* **12,** 951–953.

Marquina, S., Leitner, T., Rabinovich, R. D., Benetucci, J., Libonatti, O.,

and Albert, J. (1996). Coexistence of subtypes B, F, and B/F *env* recombinant of HIV type 1 in Buenos Aires, Argentina. *AIDS Res. Hum. Retroviruses* **12,** 1651–1654.

Morgado, M. G., Sabino, E. C., Shpaer, E. G., Bongertz, V., Brigido, L., Guimaraes, M. D. C., Castilho, E. A., Galvao-Castro, B., Mullins, J. I., Hendry, R. M., and Mayer, A. (1994). V3 region polymorphisms in HIV-1 from Brazil: Prevalence of subtype B strains divergent from North American/European prototype and detection of subtype F. *AIDS Res. Hum. Retroviruses* **10,** 569–576.

Nkengasong, J. N., Janssens, W., Heyndrickx, L., Fransen, K., Ndumbe, P. M., Motte, J., Leonaers, A., Ngolle, M., Ayuk, J., Piot, P., and van der Groen, G. (1994). Genotypic subtypes of HIV-1 in Cameroon. *AIDS* **8,** 1405–1412.

Op de Coul, E., van der Schoot, A., Goudsmit, J., van den Burg, R., Janssens, W., Heyndrickx, L., van der Groen, G., and Cornelissen, M. (1999). Independent introduction of transmissible F/D recombinant HIV-1 from Africa into Belgium and the Netherlands. Submitted for publication.

Peeters, M., Esu-Williams, E., Nzila, E., Mpoudi, E., Mboup, S., Saragosti, S., and Delaporte, E. (1998). Molecular epidemiology of HIV-1 genetic subtypes in West and Central Africa. Presented at 12th World AIDS Conference, Geneva, Switzerland, 1998. [Presentation 430, Poster 11163]

Potts, K. E., Kalish, M. L., Lott, T., Orloff, G., Luo, C-C., Bernard, M-A., Alves, C. B., Badaro, R., Suleiman, J., Ferreira, O., Schochetman, G., Johnson, W. D., Jr., Ou, C-Y., Ho, J. L., and the Brazilian Collaborative AIDS Research Group (1993). Genetic heterogeneity of the V3 region of the HIV-1 envelope glycoprotein in Brazil. *AIDS* **7,** 1191–1197.

Ray Stuart, C. (1998/1999) "SimPlot," Division of Infectious Diseases, Johns Hopkins University School of Medicine, Baltimore, MD. [software and documentation]

Robertson, D. L., Sharp, P. M., McCutchan, F. E., and Hahn, B. H. (1995a). Recombination in HIV-1. *Nature* **374,** 124–126.

Robertson, D. L., Hahn, B. H., and Sharp, P. M. (1995b). Recombination in AIDS viruses. *J. Mol. Evol.* **40,** 249–259.

Sabino, E. C., Diaz, R. S., Brigido, L. F., Learn, G. H., Mullins, J. I., Reingold, A. L., Duarte, A. J. S., Mayer, A., and Busch, M. P. (1996). Distribution of HIV-1 subtypes seen in an AIDS clinic in Sao Paulo City, Brazil. *AIDS* **10,** 1579–1584.

Sabino, E. C., Shpaer, E. G., Morgado, M. G., Korber, B. T. M., Diaz, R. S., Bongertz, V., Cavalcante, S., Galvao-Castro, B., Mullins, J. I., and Mayer, A. (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.

Salminen, M. O., Koch, C., Sanders-Buell, E., Ehrenberg, P. K., Michael, N. L., Carr, J. K., Burke, D. S., and McCutchan, F. E. (1995a). Recovery of virtually full length HIV-1 provirus of diverse subtypes from primary virus cultures using the polymerase chain reaction. *Virology* **213,** 80–86.

Salminen, M. O., Carr, J. K., Burke, D. S., and McCutchan, F. E. (1995b). Identification of breakpoints in intergenotypic recombinants of HIV type 1 by bootscanning. *AIDS Res. Hum. Retroviruses* **11,** 1423–1425.

Simon, F., Loussert-Ajaka, I., Damond, F., Saragosti, S., Barin, F., and Brun-Vézinet, F. (1996). HIV type 1 diversity in northern Paris, France. *AIDS Res. Hum. Retroviruses* **12,** 1427–1433.

Triques, K., Bourgeois, A., Saragosti, S., Vidal, N., Mpoudi-Ngole, E., Nzilambi, N., Apetrei, C., Ekwalanga, M., Delaporte, E., and Peeters, M. (1999). High diversity of HIV-1 subtype F strains in Central Africa. *Virology* **259,** 99–109.

Triques, K., Bourgeois, A., Vidal, N., Mpoudi-Ngole, E., Mulanga-Kabeya, C., Nailambi, N., Torimiro, N., Saman, E., Delaporte, E., and Peeters, M. (2000). Near-full-length genome sequencing of divergent African HIV-1 subtype F viruses leads to the identification of a new HIV-1 subtype designated K. *AIDS Res. Hum. Retroviruses* **16,** 139–151.

Van de Peer, Y., Janssens, W., Heyndrickx, L., Fransen, K., van der Groen, G., and De Wachter, R. (1996). Phylogenetic analysis of the env gene of HIV-1 isolates taking into account individual nucleotide substitution rates. *AIDS* **10,** 1485–1494.

World Health Organization (1998). WHO/UNAIDS Guidelines for Standard HIV Isolation Procedures, 1998 update.