AIDS RESEARCH AND HUMAN RETROVIRUSES
Volume 16, Number 18, 2000, pp. 2059–2064
Mary Ann Liebert, Inc.

Pole 2

(13)

Sequence Note

Identification of a Complex *env* Subtype E HIV Type 1 Virus from the Democratic Republic of Congo, Recombinant with A, G, H, J, K, and Unknown Subtypes

NICOLE VIDAL,[1] CLAIRE MULANGA-KABEYA,[1] NZILA NZILAMBI,[2] ERIC DELAPORTE,[1,3] and MARTINE PEETERS[1]

ABSTRACT

Up to now, all known *env* subtype E viruses (CRF01-AE) have had the same mosaic structure with subtype A, and no other *env* subtype E HIV-1 viruses with non-A subtypes in their genomes have been described. In this report we describe the full-length genome sequence of an *env* subtype E isolate with a recombinant genome different from the prototype CRF01-AE strains. The 97CD-KTB49 strain, obtained from a tuberculosis patient in Kinshasa, has a complex mosaic genome involving subtypes A, E, G, H, J, K, and several unknown fragments. The U sequences formed well-separated clusters together with previously described unknown fragments from CRF04-cpx (subtype I), and from Z321, the oldest intersubtype recombinant isolated in 1976 in the Democratic Republic of Congo. The complex recombinant virus from our study is not an isolated strain; partial sequencing of a second strain, 97CD-KFE45, confirmed the breakpoints observed in the 97CD-KTB49 strain in the regions sequenced. The complexity of these recombinant strains suggests a longstanding presence of subtype E in Central Africa.

PHYLOGENETIC ANALYSES of numerous strains of HIV-1, isolated from diverse geographic origins, have revealed three distinct clades of viruses, which have been termed groups M, N, and O. The vast majority of strains found worldwide, and that are responsible for the pandemic, belong to just one of these lineages, group M (for Main). Within group M, subtypes and circulating recombinant forms (CRFs) have been proposed. To be considered a subtype, isolates should resemble each other, and no other existing subtype, across the entire genome, and CRFs should have a similar mosaic genome with the same intersubtype breakpoints. In this light, there are nine subtypes of HIV-1 group M: A, B, C, D, F, G, H, J, and K. All known representatives of what was initially described as subtype E appear in fact to be recombinants of subtypes A and E, and are now designated CRF01-AE. CRF04-cpx viruses correspond to the previously described *env* subtype I viruses, which are complex recombinant viruses involving at least four subtypes.[1]

Subtype E was first designated on the basis of the distinct phylogenetic position of these viruses in *env* trees, but full-length genome analysis revealed that the majority of the genome was subtype A, and the only nonsubtype A sequences are found within (most of) the *env* gene; parts of *vif, vpr,* and *nef*; and the long terminal repeat (LTR).[2,3] However, no virus has yet been described that represents a full-length nonrecombinant, subtype E sequence, perhaps simply because that lineage has become extinct. CRF01-AE viruses are responsible for the AIDS epidemic in Southeast Asia, where it is the predominant circulating strain.[4] CRF01-AE viruses have also been described in Central Africa, mainly in the Central African Republic, where it represented initially more than 15% of the circulating strains,

[1]Laboratoire Retrovirus, IRD, 34032 Montpellier Cedex 1, France.
[2]Projet SIDA, Kinshasa, Democratic Republic of Congo.
[3]CHU, Gui de Chauliac, Montpellier, France.

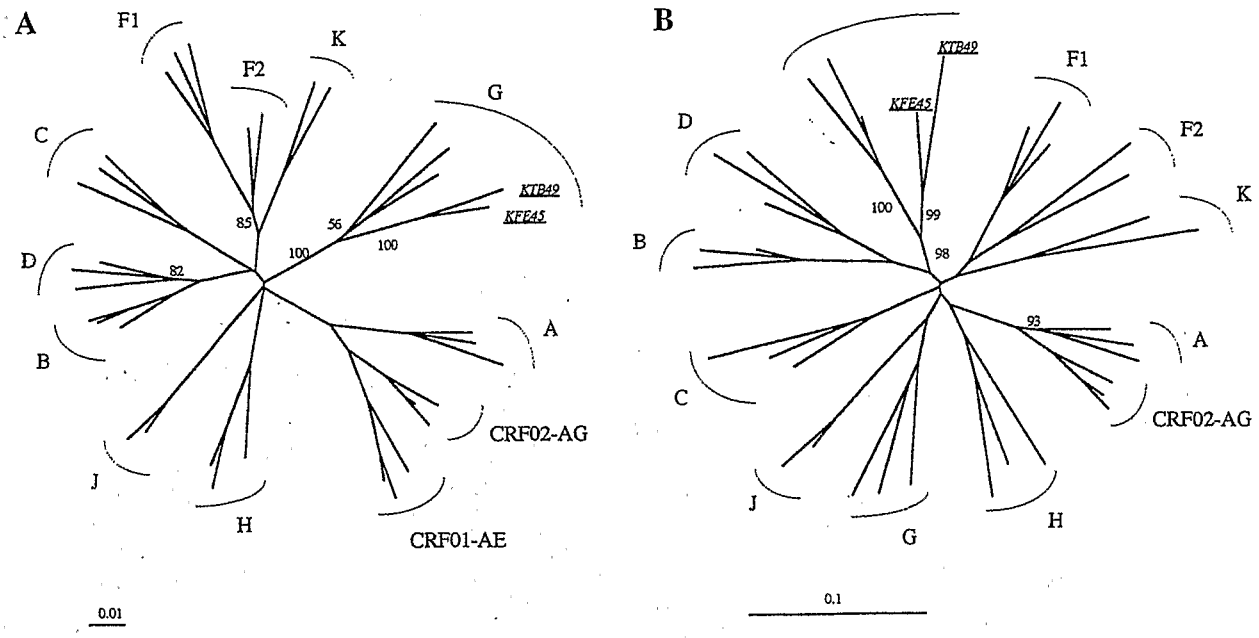


FIG.1. Phylogenetic trees based on 693 and 513 unambiguously aligned nucleotides from the p24 *gag* (**A**) and V3–V5 *env* (**B**) regions. The phylogenetic relationships of the newly derived sequences were estimated from sequence comparisons with previously reported representatives for group M nucleotide acid sequences and were aligned by CLUSTAL W with minor manual adjustments, bearing in mind the protein sequences. Sites where there was a gap in any of the sequences, as well as areas of uncertain alignment, were excluded from all sequence comparisons. Phylogenetic trees were constructed by the neighbor-joining method and the reliability of the branching patterns was assessed by the bootstrap approach implemented by CLUSTAL W.[8] The reference sequences were as follows: A_KE.Q2317, A_UG.92UG037, A_SE.SOSE7253, CRF01_AE_TH.93TH253, CRF01_AE_TH.CM240, CRF01_AE_CF.90CR402, CRF02_AG_DJ.DJ263, CRF02_AG_DJ.DJ264, CRF02_AG_NG.IBNG, B_FR.HXB2R, B_US.JRFL, B_US.WEAU160, C_ET.ETH2220, C_IN.21068, C_BW.96BW0502, D_ZR.84ZR085, D_ZR.NDK, D_UG.94UG114, F1_BR.93BR020, F1_FI.FIN9363, F1_BE.VI850, F2_95CM.MP255, F2_95CM.MP257, G_BE.DRCBL, G_SE.SE6165, G_FI.HH8793, H_BE.VI991, H_BE.VI997, H_CF.90CF056, J_SE.SE92809, J_SE.SE91733, K_96CM.MP535, and K_97CD.EQTB11. All bootstrap values at each of the branches defining a subtype were above 85%.

and sporadic cases have been described in other Central African countries.[5,6]

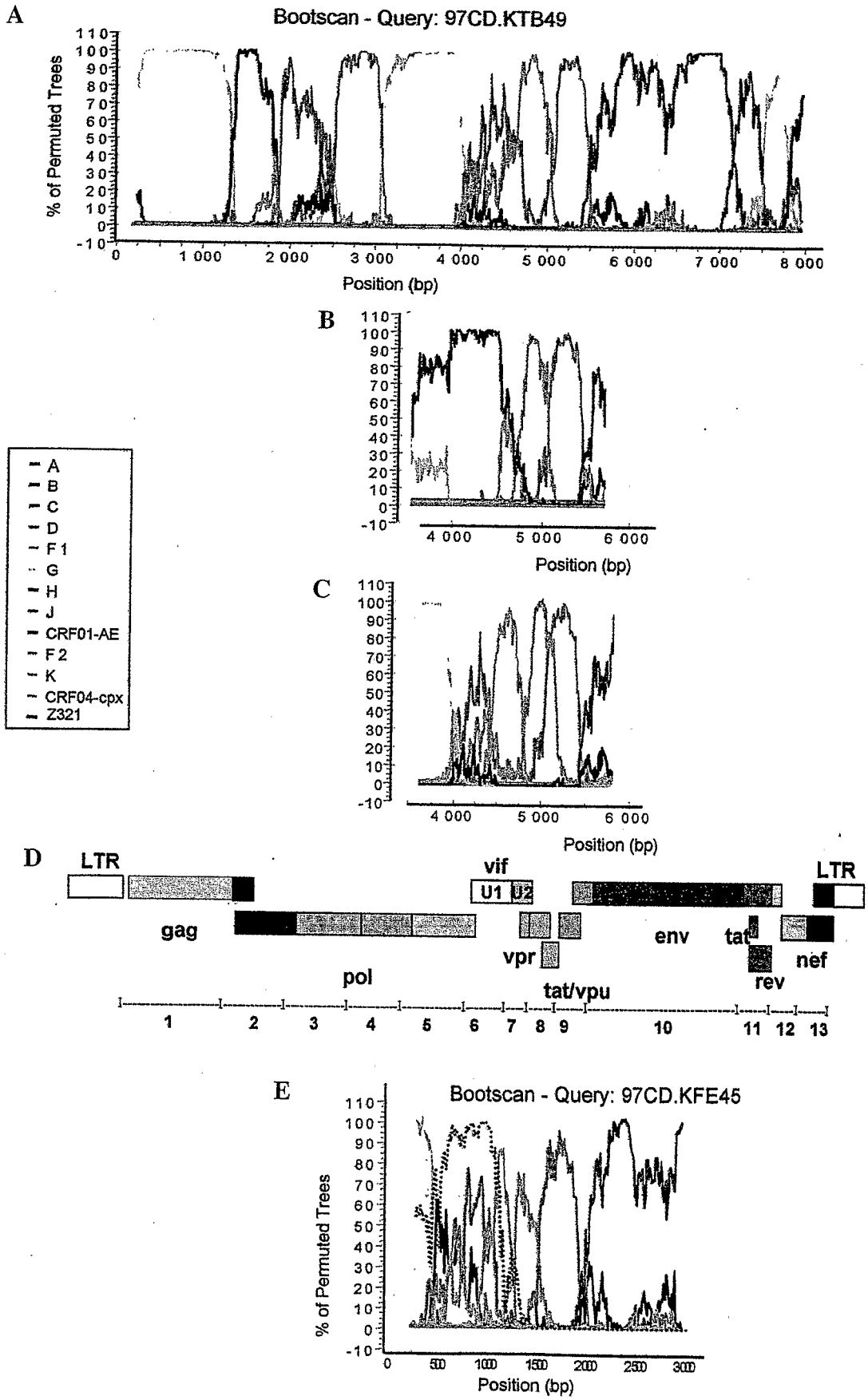
In this report we describe the full-length genome sequence of an *env* subtype E HIV-1 isolate with a recombinant genome different from the prototype CRF01-AE strains.

In 1997 we performed a study identifying the HIV seroprevalence in various population groups and geographic regions from the Democratic Republic of Congo (DRC, formerly Zaire).[7] A 10-ml whole blood sample was collected in an EDTA tube and peripheral mononuclear blood cells (PBMCs) were separated from plasma by ficoll gradient centrifugation. Plasma and cell pellets were stored at −20°C. For the HIV-1-positive samples the genetic subtype was identified in the V3—V5 region of the envelope and in the p24 region from Gag by direct sequencing followed by phylogenetic tree analysis as previously described. This preliminary genetic characterization classified the two samples as divergent subtype E in this region of the envelope and as subtype G in p24 (Fig. 1).

The 97CD-KTB49 sample was obtained from an HIV-positive tuberculosis patient living in Kinshasa, the capital city, and the 97CD-KFE45 strain was from an asymptomatic pregnant women attending an antenatal clinic in Kinshasa.

To elucidate the genomic structure of these viruses, the near full-length sequence was derived directly from PBMCs for the 97CD-KTB49 virus. Provirus was amplified by a series of overlapping nested PCRs amplifying fragments between 2000 and 5000 bp. We used the Boehringer Mannheim (Indianapolis, IN) Long Template Expand DNA polymerase, according to the instructions of the manufacturer. The overlapping PCR fragments were directly sequenced. Sequencing was performed by using cycle sequencing and Big Dye terminator technology on an automated DNA sequencer (ABI 373A stretch; Applied Biosys-

FIG. 2. Bootstrap plot (**A**) showing the complex mosaic structure of the 97CD-KTB49 strain (8936 bp). Windows of 500 bp were analyzed with a 10-bp increment, 100 replicates, and $T/t = 2.0$. The multiple genome alignment used to calculate the consensus references sequences (50% threshold) included the same references sequences as used for the phylogenetic tree analysis of the partial *gag* and *env* sequences in Fig. 1. (**B**) Bootstrap plot of the accessory genes region, using the same references as in (**A**), but with adding Z321 and CRF04-cpx. (**C**) Bootstrap plot of the accessory genes region, using the same references as in (**A**), with omission of Z321. (**D**) Schematic representation of the overall mosaic structure of the 97CD.KTB49 strain. (**E**) Bootstrap plot of the 3500-nucleotide region of the 97CD-KFE45 strain as compared with the mosaic structure of 97CD-KTB49, with a superimposed plot for the Z321 strain.


A
Bootscan - Query: 97CD.KTB49
% of Permuted Trees
Position (bp)
B
Position (bp)
C
Position (bp)
A
B
C
D
F1
G
H
J
CRF01-AE
F2
K
CRF04-cpx
Z321
D
LTR
vif
U1 U2
LTR
gag
env
tat
nef
vpr
rev
pol
tat/vpu
1 2 3 4 5 6 7 8 9 10 11 12 13
E
Bootscan - Query: 97CD.KFE45
% of Permuted Trees
Position (bp)

tems, Foster City, CA). The sequenced fragments were reanalyzed and assembled into contiguous sequences by using the Seqed program (Applied Biosystems).

Examination of potential coding regions revealed the expected reading frames for *gag, pol, vif, vpr, tat, rev, vpu, env,* and *nef.* LTR extremities were not sequenced. None of the genes contained in-frame stop codons, major deletions, insertions, or rearrangements.

The new sequence was aligned against near full-length reference sequences representing the nonrecombinant HIV-1 subtypes and the CRF01-AE strains, using CLUSTAL W[8] with minor manual adjustments, bearing in mind the protein sequences. Regions that could not be aligned unambiguously, because of sequence variability, were omitted from the analysis. Phylogenetic trees based on unambiguously aligned nucleotides were generated by the neighbor-joining approach, and reliability of the branching order was determined by CLUSTAL W, with 1000 bootstraps for each phylogeny. To analyze the recombinant structure of this new virus, several additional analyses were performed. Similarity plots, using SimPlot 2.5 software,[9] determined the percent similarity between selected pairs of sequences by moving a window of 500 bp along the genome alignment in 10-bp increments. The similarity values for each pairwaise comparison were plotted at the midpoint of the 500-bp segment. SimPlot 2.5 software was also used to calculate bootstrap plots. For the bootstrap plots, the SimPlot software performed bootscanning on neighbor-joining trees by using SEQBOOT, DNADIST (with the Kimura two-parameter method and a transition-to-transversion ratio of 2.0), NEIGHBOR, and CONSENSUS from the PHYLIP package for a 500-bp window, moving along the alignment in increments of 10 bp (this approach is similar to the bootscanning method described by Salminen *et al.*[10]) We evaluated 100 replicates for each phylogeny. The bootstrap values for the studied sequences were plotted at the midpoint of each window. The regions that did not cluster with any of the known subtypes were submitted to BLAST analysis (BLASTN 2.0.6 on line; http://hiv-web.lanl.gov/BASIC-BLAST), to see whether they are closely related with previously described unknown fragments in other HIV-1 strains.

The complexity of the new 97CD-KTB49 strain is readily apparent from the similarity plots (data not shown) and the bootscan analyses. Figure 2A represents the bootscan analysis of the new strain against references representing all the nonrecombinant and CRF01-AE HIV-1 viruses. Multiple subtypes are involved in the genomic structure, particularly subtypes G, H, J, K, A, and E. The *gag* gene is almost entirely subtype G; only the 3′ end is subtype H. The structure of the *pol* gene is complex, with alternation of subtypes H, J, and G. The protease gene is subtype H. On the bootstrap and similarity plots, the 5′ end of the reverse transcriptase (RT) gene seemed difficult to classify, but phylogenetic tree analysis, as well as the BLAST analysis of this fragment, classified it as subtype J (Fig. 3, tree 3). The integrase gene belonged to subtype G. The accessory gene region was more complex. The *vif* gene did not cluster with any of the known subtypes; neither did the 5′ end of the *vpr* gene. The 3′ end of *vpr* and the *tat* gene clustered with subtype K, and the *vpu* gene clustered with subtype J.

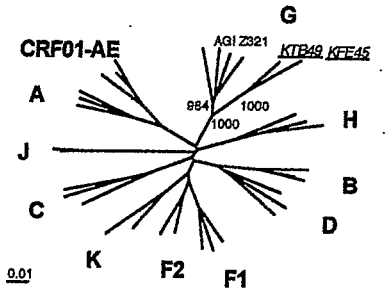
The envelope gene clusters almost entirely with the subtype E fragment from the CRF01-AE prototype strains, except for the 5′ end of gp41, which clearly clusters with subtype A, in contrast to the prototype CRF01-AE, which forms a separate cluster classified as U in the same region under new nomenclature. Overall, the *env* subtype E region was similar between 97CD.KTB49 and the prototype CRF01-AE strains. The *nef* gene is subtype G at the 5′ end and subtype H at the 3′ end. The subtype identifications of the various genomic regions were all confirmed by phylogenetic tree analysis of the corresponding fragments, as shown in Fig. 3.

To better analyze the *vif–vpr* region, for which we could not clearly identify the subtype, we performed a BLAST search. The best match (94% identity) was found with the complex recombinant strain Z321, isolated in 1976 from the Democratic Republic of Congo.[11] It is important to note that the Z321 strain has a recombinant structure in this region,[12,13] with the 3′ end corresponding to the unknown fragment described in CRF04-cpx, and the 5′ end consisting of another unknown fragment. Figure 2B represents a bootscan analysis of the accessory gene region using the same references as in Fig. 2A plus the Z321 strain and CRF04-cpx. From Fig. 2B it is clear that the initially unknown region in 97CD.KTB49 is closely related to Z321, except at a small region where the bootstrap values are lower and the discrimination between Z321 and CRF04-cpx cannot be made, probably because of the close relationship between these two viruses in this region. A bootscan analysis (Fig. 2C) with the same reference strains as in Fig. 2B but with omission of the Z321 strains, confirms that this small fragment corresponds to the unknown fragment common between CRF04-cpx and Z321. The additional bootscan analysis from Fig. 2B and C confirms a similar recombinant structure between Z321 and 97CD-KTB49 in the *vif–vpr* region. Phylogenetic tree analyses confirmed also that the *vif–vpr* fragments consist of two different unknown fragments (Fig. 3, trees 6 and 7). Figure 2D shows the overall mosaic structure of the new 97CD.KTB49 strain.

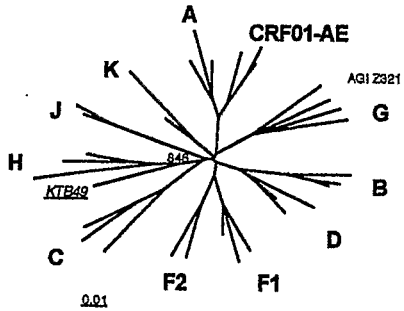
The complex A/E/G/H/J/K/? from our study is not an isolated strain; partial sequencing of the second strain, 97CD-KFE45, confirmed the breakpoints observed in the 97CD-KTB49 strain in the regions sequenced. For the 97CD-KFE45 strain, the sequenced regions were as follows: 1453 bp sequenced in *gag*, 676 bp in the middle of *pol*, 3500 bp starting from the 3′ end of *pol* to the middle of gp120, and 792 bp covering the 3′ end of gp41 to the middle of the *nef* gene. Figure 2E shows the bootscan analysis for the 97CD-KFE45 fragment of 3500 bp in comparison with the overall genomic structure of the 97CD-KTB49 strain. The mosaic structure involved the same subtypes with the same breakpoints as for 97CD.KTB49.

FIG. 3. Phylogenetic trees, confirming the different subtype designations in the 97CD-KTB49 genome; when possible, the second strain, 97CD.KFE45, was included in the analysis. Bootstrap values above 800 at each of the internal branches defining a subtype are not shown, except for the cluster including KTB49 and/or KFE45. The same reference strains as described for Fig. 1 were used, with the difference that CRF02-AG was omitted and Z321 was included and in tree 7, CRF04_CPX was added in the analysis to illustrate the mosaic structure in this particular region of the genome.

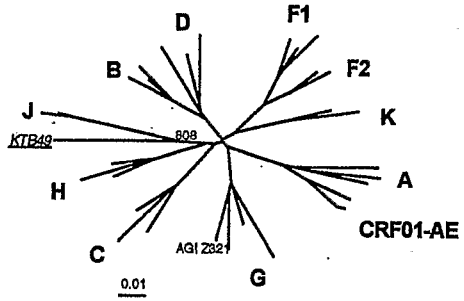
1: 1250 bp



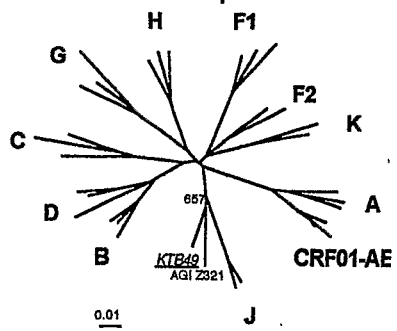
2: 626 bp



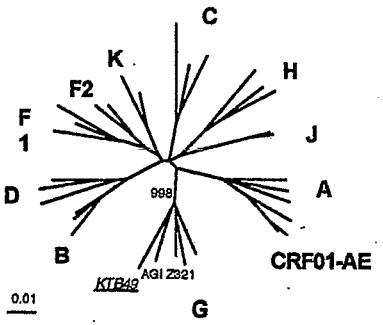
3: 626 bp



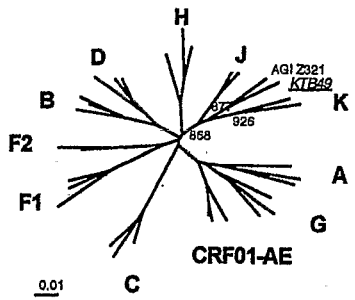
4: 621 bp



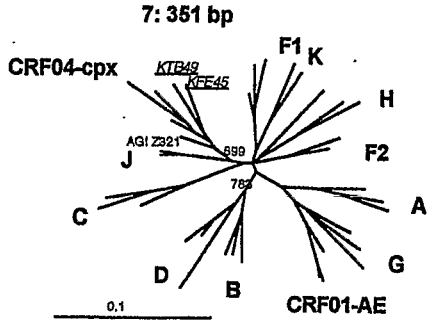
5: 881 bp



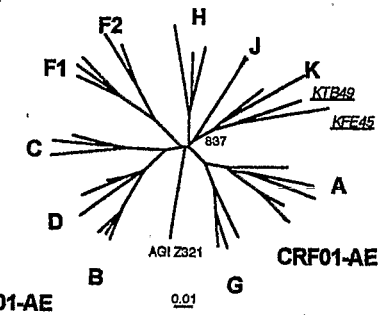
6: 451 bp



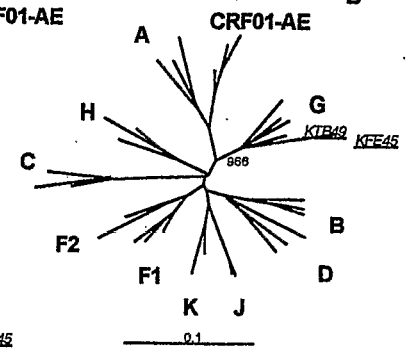
7: 351 bp



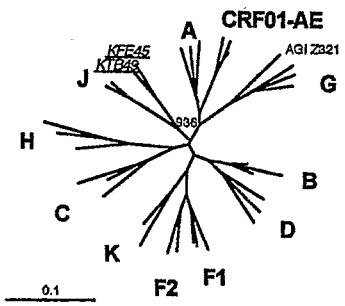
8: 301 bp



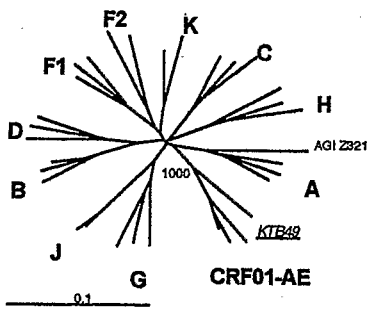
12: 334 bp



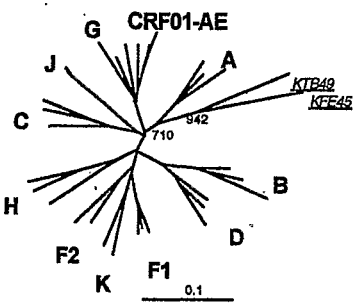
9: 401 bp



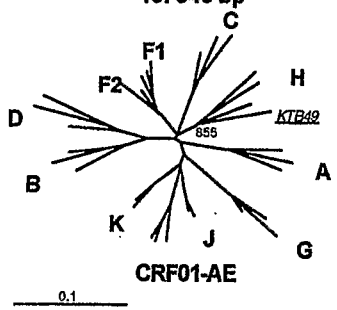
10: 1651 bp



11:323 bp



13: 349 bp



The *vif–vpr* gene has the same structure, with unknown fragments from the CRF04-cpx and Z321 strains. The *gag* fragment was also subtype G, the small *pol* fragment was subtype J, and the partial gp41–*nef* fragment was also A/G recombinant.

Previous reports on genetic subtypes of strains from the DRC suggested a high level of genetic diversity, with all the known subtypes cocirculating and several HIV subtypes having been associated with the DRC through patients living in Europe.[14–16] Mokili and co-workers also reported a high intrasubtype variability.[16] In addition, some of the first African HIV-1 isolates to be characterized, MAL and Z321 (obtained from a stored plasma sample obtained in 1976 in a rural area in the north of the DRC), have been identified as complex recombinants.[12,14,16] The presence of recombinant viruses, early in the AIDS epidemic, confirms that HIV had already been present for a while in this region of Africa.[17]

Overall, the high number of HIV-1 subtypes cocirculating, the high intrasubtype diversity, as well as the complexity of these recombinant strains, are all in agreement with an old and mature epidemic in Central Africa, more particularly in the DRC. The genetic diversity observed in the DRC represents a real challenge for future vaccine development, as well as for efficiency of antiretroviral treatment and diagnostic tests.

ACKNOWLEDGMENTS

This work was cosponsored by grants from the Agence Nationale de Recherches sur le SIDA (ANRS, Projet Sidak) and IRD (Institut de Recherche pour le Dévelopement).

SEQUENCE DATA

The new sequences have been deposited in the GenBank Data Library under the following accession numbers: for 97CD.KFE45: AJ404321 (*gag*), AJ404322 (partial *pol*), AJ404323 (accessory genes to partial *env*), AJ404324 (partial *env* gene to partial *nef* gene). For 97CD.KTB49: AJ404325, complete genome (8936 nucleotides).

REFERENCES

1. Robertson D, Anderson J, Bradac J, *et al.*: HIV-1 nomenclature proposal: A reference guide to HIV-1 classification. In: *Human Retroviruses and AIDS 1999: A Compilation and Analysis of Nucleic Acid and Amino Acid Sequences*. Los Alamos National Laboratory, Los Alamos, New Mexico, 1999.
2. Gao F, Robertson D, Morrison S, *et al.*: The heterosexual human immunodeficiency virus type 1 epidemic in Thailand is caused by an intersubtype (A/E) recombinant of African origin. J Virol 1996;70:7013–7029.
3. Carr J, Salminen M, Koch C, *et al.*: Full-length sequence and mosaic structure of a human immunodeficiency virus type 1 isolate from Thailand. J Virol 1996;70:5935–5943.
4. European Commission and the Joint United Nations Programme on HIV/AIDS: HIV-1 subtypes: Implications for epidemiology, pathogenicity, vaccines and diagnostics. AIDS 1997;11:UNAIDS17–UNAIDS36. [Workshop Report]
5. Muller-Trutwin M, Chaix M, Letourneur F, *et al.*: Increase of HIV-1 subtype A in Central African Republic. J Acquir Immune Defic Syndr 1999;21:164–171.
6. Nkengasong J, Janssens W, Heyndrickx L, *et al.*: Genotypic subtypes of HIV-1 in Cameroon. AIDS 1994;8:1405–1412.
7. Mulanga-Kabeya C, Nzilambi N, Edidi B, *et al.*: Evidence of stable HIV prevalences in selected populations in the Democratic Republic of the Congo. AIDS 1998;12:905–910.
8. Thompson J, Higgins D, and Gibson T: CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 1994;22:4673–4680.
9. Ray, SC: *Simplot for Windows* (version 2.5) Distributed by author (http://www.med.jhu.edu/deptmed/sray/download/).
10. Salminen M, Carr J, Burke D, and McCutchan F: Identification of breakpoints in intergenotypic recombinants of HIV type 1 by bootscanning. AIDS Res Hum Retroviruses 1995;11:1423–1425.
11. Srinivasan A, York D, Butler D, *et al.*: Molecular characterization of HIV-1 isolated from a serum collected in 1976: Nucleotide sequence comparison to recent isolates and generation of hybrid HIV. AIDS Res Hum Retroviruses 1989;5:121–129.
12. Robertson DL, Gao F, Hahn BH, and Sharp PM: Intersubtype recombinant HIV-1 sequences. In: *Human Retroviruses and AIDS 1997: A Compilation and Analysis of Nucleic Acid and Amino Acid Sequences*. Los Alamos National Laboratory, Los Alamos, New Mexico, 1997.
13. Choi D, Dube S, Spicer T, Slade H, Jensen F, and Poiesz B: HIV type 1 isolate Z321, the strain used to make a therapeutic HIV type 1 immunogen, is intersubtype recombinant. AIDS Res Hum Retroviruses 1997;13:357–361.
14. Alizon M, Wain-Hobson S, Montagnier L, and Sonigo P: Genetic variability of the AIDS virus: Nucleotide sequence analysis of two isolates from African patients. Cell 1986;46:63–74.
15. Laukkanen T, Albert J, Liitsola K, *et al.*: Virtually full-length sequences of HIV type 1 subtype J reference strains. AIDS Res Hum Retroviruses 1999;15:293–297.
16. Mokili J, Wade C, Burns S, *et al.*: Genetic heterogeneity of HIV type 1 subtypes in Kimpese, rural Democratic Republic of Congo. AIDS Res Hum Retroviruses 1999;15:655–664.
17. Zhu T, Korber B, Nahmias A, Hooper E, Sharp P, and Ho D: An African HIV-1 sequence from 1959 and implications for the origin of the epidemic. Nature 1998;391:594–597.

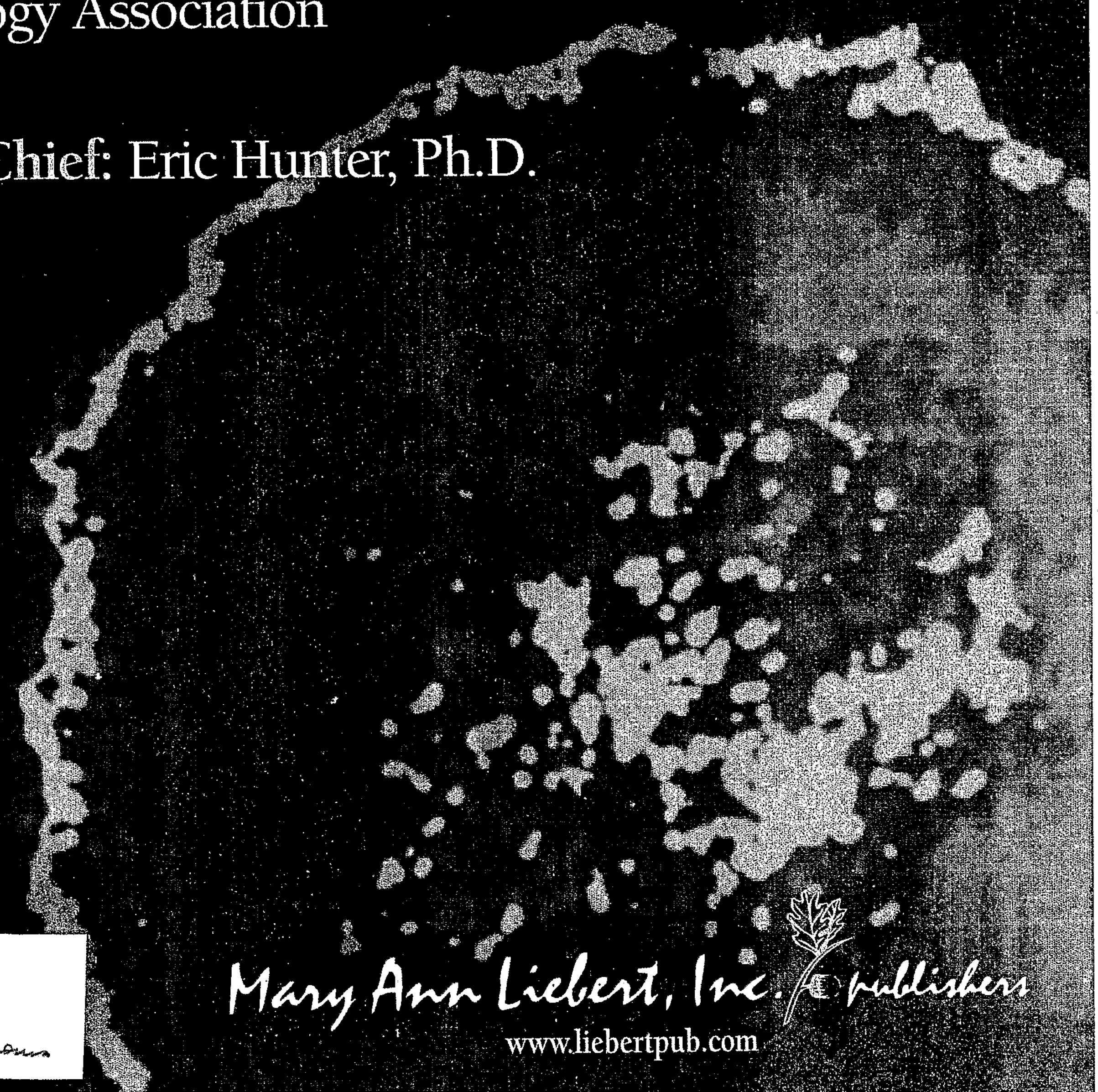
Address reprint requests to:
Martine Peeters
Laboratoire Retrovirus
IRD, BP 5045, 34032
Montpellier Cedex 1, France

E-mail: martine.peeters@mpl.ird.fr

AIDS Research and Human Retroviruses

The Official Journal of the International Retrovirology Association

Editor-in-Chief: Eric Hunter, Ph.D.



Mary Ann Liebert, Inc. publishers
www.liebertpub.com