JAIDS Journal of Acquired Immune Deficiency Syndromes
29:522–530 © 2002 Lippincott Williams & Wilkins, Inc., Philadelphia

CRF06-cpx: A New Circulating Recombinant Form of HIV-1 in West Africa Involving Subtypes A, G, K, and J

*Céline Montavon, †Coumba Toure-Kane, ‡§John N. Nkengasong, *Laurence Vergne, ‖Kurt Hertogs, †Souleymane Mboup, *¶Eric Delaporte, and *Martine Peeters

**Laboratoire Retrovirus, IRD, Montpellier, France; †Hôpital Le Dantec, Dakar, Senegal; ‡Division of HIV/AIDS Prevention, National Center for STD, HIV, and TB Prevention, Centers for Disease Control and Prevention, Atlanta, Georgia, U.S.A.; §Projet RETRO-CI, Abidjan, Ivory Coast; ‖Virco Group of Companies, Mechelen, Belgium; and ¶CHU Gui de Chaulhiac, Montpellier, France*

Summary: Phylogenetic analysis of numerous strains of HIV-1 isolated from diverse geographic origins has revealed three distinct groups of HIV-1: groups M, N, and O. Within group M, subtypes, sub-subtypes and circulating recombinant forms (CRFs) exist. Recently, two near-full-length genomes of similar complex mosaic viruses containing fragments of subtypes A, G, I, and J were described in patients from Burkina Faso (BFP-90) and Mali (95ML-84). Here, we report on the characterization of two additional full-length genome sequences with similar mosaic structure in epidemiologically unlinked individuals from Senegal (97SE-1078) and Mali (95ML-127). Phylogenetic and recombinant analysis confirmed that the previously described strains, BFP-90 and 95ML-84, were indeed a new CRF of HIV-1, which we can now designate as CRF06-cpx. This new CRF fits the complex (cpx) designation, because four different subtypes (A, G, K, and J) were involved in the mosaic genome structure. The fragment in the *pol* gene, which was initially characterized as unknown in the BFP-90 strain and subsequently as subtype I in the 95ML-84 strain, is now, with the recent description of the new K subtype, clearly identified as subtype K. CRF06-cpx circulates in Senegal, Mali, Burkina Faso, Ivory Coast, and Nigeria, although the exact prevalence remains to be determined. Importantly, this new variant has also been documented on other continents (Europe [France] and Australia), showing that these viruses are spreading not only locally but globally.

Key Words: HIV—Subtypes—Recombination—CRF—Africa.

Phylogenetic analysis of numerous strains of HIV-1 isolated from diverse geographic origins has revealed three distinct groups of viruses: groups M, N, and O (1). The majority of strains found worldwide and responsible for the pandemic belong to group M ("main") (2) (HIV Sequence Database Los Alamos National Laboratory; available at http://hiv-web.lanl.gov/). Within group M, subtypes, sub-subtypes, and circulating recombinant forms (CRFs) exist (3). To be considered a subtype, isolates should resemble each other, and no other existing subtype across the entire genome or CRF should have a similar mosaic genome with the same intersubtype breakpoints. Thus, nine subtypes of HIV-1 group M exist (A–D, F–H, J, and K), and several CRFs have been recognized and play a potentially important role in the global epidemic (3). All known representatives of what was initially described as subtype E seem to be recombinants of subtypes A and E; they are now designated as CRF01_AE (4,5) and are responsible for the AIDS epidemic in Southeast Asia. Full-length sequencing of an initially divergent *env* subtype A strain, HIV-1 IbNg,

Address correspondence and reprint requests to Martine Peeters, UR36, Laboratoire Retrovirus, IRD, 911 Avenue Agropolis, B.P. 5045, 34032 Montpellier, Cedex 1, France; e-mail: martine.peeters@mpl.ird.fr
Manuscript received July 11, 2001; accepted December 18, 2001.

revealed that this virus has a complex mosaic genome of alternating subtype A and G sequences (6,7). Because a number of similar viruses have been reported from countries in West Africa and East Africa, this subtype is now designated as CRF02_AG. Recently, it has been shown that these CRF02_AG viruses are predominant in West Africa (8). An epidemic among intravenous drug users in Kaliningrad, Russia involves viruses that are mosaics of subtypes A and B and thus are termed CRF03_AB (9–11). Isolate 94CY032 from Cyprus was designated as the prototype of subtype I based on gp120 sequences (12). Full genome sequencing revealed this virus to be a complex mosaic with multiple breakpoints between regions of several distinct subtypes, including A, G, and I (13), and two similar viruses have been described in epidemiologically unlinked individuals from Greece (14). As is the case with subtype E, no full-length nonrecombinant subtype I virus has yet been identified. These viruses are now designated as CRF04-cpx. Recent reanalysis with previously unavailable complete genome sequences has revealed that these viruses are in fact mosaics with regions associated with subtypes A, G, H, and K and unclassified regions (3). Therefore, subtype I has been removed from the genetic classification system of HIV strains, and the subtype I fragment has been relabeled as unclassified. CRF05-DF has been documented recently in strains originating from the Democratic Republic of Congo (former Zaire) (15).

Because full-length genome sequencing has become technically easier and with the development of a variety of tools to detect the existence of mosaic (i.e., recombinant) genomes, the criteria to propose a novel subtype, sub-subtype, or CRF have been revised. Three near-full-length genomic sequences are preferred; alternatively, two complete genomes in conjunction with partial sequences, including breakpoints, are needed. These strains must be identified in at least 3 individuals without any direct epidemiologic link (3).

Recently, two near-full-length genomes of similar complex mosaic viruses containing fragments of subtypes A, G, I, and J were described in patients from Burkina Faso (BFP-90) and Mali (95ML-84) (16,17). As is the case with the definition of a novel subtype, a third full-length genome sequence or partial sequences confirming the breakpoints from a third epidemiologically unlinked individual are required to designate this as a new CRF (3). In this study, we describe full genome sequences from two additional virus strains with a similar complex mosaic genome obtained from epidemiologically unlinked individuals (1 from Senegal and 1 from Mali), and we have reanalyzed the fragment initially described as subtype I in the previously published sequences. In addition, on the basis of *pol* sequences that cover one of the multiple breakpoints, we have documented the presence of these viruses in the Ivory Coast.

METHODS

Virus Isolates

Full-length genome sequences were obtained for two HIV-1 viruses (95ML-127 and 97SE-1078), starting with DNA extracted from primary uncultured peripheral blood mononuclear lymphocytes. The isolates had been previously classified as subtype A in *gag* (p24) and subtype G in *env* (V3–V5); in addition, the accessory gene region for 95ML-127 was classified as subtype J. Isolate 95ML-127 was obtained from an asymptomatic female sex worker in Bamako, Mali (18). Isolate 97SE-1078 was obtained from a Senegalese woman with clinical symptoms of AIDS, who attended one of the major hospitals in Dakar, Senegal in 1996 and died of AIDS in 1997.

Protease and reverse transcriptase (RT) genes were sequenced starting with plasma samples for nine additional samples. The samples originated from Senegal (n = 3) (99SN-127HPD, 98SN-64HALD, and 98SN-77HALD) and Ivory Coast (n = 4) (99IC-AB0302, 99IC-AB1307, 99IC-AB1545, and 99IC-AB2478), and two were from Burkinabes attending a hospital in France (99FR-MP1426 and 00FR-MP1471).

Complete Genome Sequences

DNA was extracted from primary peripheral blood mononuclear lymphocytes with the QiaAmp DNA Blood Kit (Qiagen, Courtaboeuf, France). For each strain, three overlapping nested polymerase chain reactions (PCRs) were done to obtain the sequence of the entire genome. For the 97SE-1078 strain, a fragment that included the accessory genes, the entire envelope, and the *nef* gene was amplified with h-pol 4235 (CCC TAC AAT CCC CAA AGT CAA GG) (19) and LsiG1 (LTR) (TCA AGG CAA GCT TTA TTG AGG CTT AAG CAG) as outer primers and Vif1 (GGG TTT ATT ACA GGG ACA GCA GAG) (20) and outer 3 (Nef) (AGC ATC TGA GGG TTA GCC ACT) (21) as inner primers. The 3′ end covering half of the *pol* gene and the accessory genes was amplified with polM2 (GAT TTG TAT GTA GGA TCT GA) and Vpu1 (GGT TGG GGT CTG TGG GTA CAC AGG) (20) as outer primers and polM3 (TAT GTA GAT GGG GAG CTA ATA G) and Mvpu3 (TAC TAT RGT CCA CAC AAC TAT) (20) as inner primers. Unintegrated circular DNA was targeted to amplify the rest of the *pol* gene, *gag* gene, and LTR with the following primers: inner 5 (end of gp160) (CTG CAG GAC AGA TAG GGT TAT AGA A) (21) and hpol-4538 (TAC TGC CCC TTC ACC TTT CCA) (19) as outer primers and Mrev (*nef*) (TAA AAG AAA AGG GGG GAC TGG AAG GGC TA) and hpol-4481 (GCT GTC CCT GTA ATA AAC CCG) (19) as inner primers.

The same fragments were amplified for the 95ML-127 strain except for the envelope and *nef* gene, which were amplified separately using EnvA (GGC TTA GGC ATC TCC TAT GGC AGG AAG AA) and EnvN (CTG CCA TTC AGG GAA GTA GCC TTG TGT) (22) and 41-1 (GGG TTC TTG GGA GCA GCA GGA AGC ACT ATG GGC G) (23) and LsiG1 as inner primers on the fragment covering the accessory genes to the *nef* gene.

The Boehringer Expand Long Template DNA polymerase (Roche Diagnostic, Meylan, France) was used according to the instructions of the manufacturer. For the first rounds of amplification, PCR conditions

were as follows: denaturation at 92°C for 3 minutes followed by 34 cycles at 92°C for 10 seconds, 50°C for 30 seconds, and 72°C for 5 minutes and a final extension of 7 minutes at 72°C. The second rounds of amplification were performed under the same cycling conditions except for the polymerization step, where the time was reduced from 5 to 3 minutes.

The amplified products were purified using a QiaQuik Gel Extraction Kit (Qiagen). Nucleotide sequences were obtained by direct sequencing of the amplified DNA using the inner primers of each PCR and several primers encompassing the entire fragments.

Only the gp160 fragment of the 95ML-127 strain was cloned in pGEM plasmid with the pGEM-T Easy Vector System II (Promega, Lyon, France) before sequencing. Cycle sequencing was done using Big Dye (Applied Biosystem, Courtaboeuf, France) terminator chemistry according to the instructions of the manufacturer. Electrophoresis and data collection were performed on an ABI 373A stretch automatic sequencer with XL upgrade (Applied Biosystem).

The sequenced fragments were assembled in contiguous sequences, and a consensus of the strands was formed using the Seqed program (Applied Biosystems).

pol Sequences

Viral RNA was isolated from plasma with the QIAmp Viral RNA Kit (Qiagen) and was transcribed to cDNA as previously described for samples from Senegal and France (24). Briefly, RNA was retrotranscribed to cDNA using Expand RT (Roche Diagnostic) with primer IN3 (TCT ATB CCA TCT AAA AAT AGT ACT TTC CTG ATT CC) according to the instructions of the manufacturer. This cDNA was then amplified with a seminested PCR using outer primers G25REV (GCA AGA GTT TTG GCT GAA GCA ATG AG) and IN3 and inner primers AV150 (GTG GAA AGG AAG GAC ACC AAA TGA AAG) and IN3, with the Expand Long Template PCR DNA polymerase (Roche Diagnostic). We used buffer 3 and temperature conditions of 94°C for 3 minutes, 94°C for 20 seconds, 50°C for 30 seconds, and 72°C for 3 minutes repeated for 34 additional cycles starting from step 2, with a final extension of 5 minutes. After purification, the amplified products were directly sequenced as described previously.

pol sequences of the strains from Ivory Coast were performed in collaboration with the Virco laboratories (Virco NV, Mechelen, Belgium; Virco Ltd, Cambridge, U.K.). Briefly, the viral RNA was reverse transcribed, and a 2.2-kb fragment encoding the protease and RT regions was then amplified by nested PCR using PCR primers and conditions as described by Hertogs et al. (25). This genetic material was subsequently used in sequencing experiments. The PCR products obtained from patient plasma samples were sequenced by dideoxynucleoside-based sequence analysis. Samples were sequenced using the Big Dye terminator kit and resolved on an ABI 377 DNA sequencer (PE Biosystem), as previously described (26).

Phylogenetic Tree Analysis

Phylogenetic relations of the newly derived viruses were estimated from sequence comparisons with previously reported representatives of group M. Nucleotide sequences for the near-full-length genome and the partial *pol* sequences were aligned using CLUSTAL_W (27) with minor manual adjustments, bearing in mind the protein sequences. Sites with gaps in any of the sequences as well as areas of uncertain alignment were excluded from all sequence comparisons. Phylogenetic trees of the entire genome and the *pol* gene were constructed by means of the neighbor joining method, and reliability of the branching orders was implemented by CLUSTAL_W using the bootstrap approach. TreeView (Roderic Page, Division of Environmental and Evolutionary Biology, University of Glasgow, U.K.) was used to draw trees for illustrations. Genetic distances were calculated using the Kimura two-parameter method, with a transition weight of 0.5 (28).

The new sequences were aligned and compared with the following reference sequences: subtype A (A-U455, A-UG037, and A-Q23), subtype B (B-RF, B-OYI, B-JRFL, and B-HXB2), subtype C (C-ETH2220 and C-92BR025), subtype D (D-84ZR085, D-94UG114, D-NDK, and D-ELI), sub-subtype F1 (F-93BR20, F-BE.VI850, F-FI.FIN9363, and F1-96FR.MP411), sub-subtype F2 (F2-95CM.MP255 and F2-CM.MP257), subtype G (G-SE6165, G-HH8793, G-92NG083, and G-92NG003), subtype H (H-90CR056, H-BE.VI991, and H-BE.VI997), subtype J (J-SE.SE92809 and J-SE.SE91733), and subtype K (K-97ZR.EQTB11 and K-96CM.MP535).

Analysis for Intersubtype Mosaicism

To analyze the recombinant structure of the new viruses, several additional analyses were performed. Using Simplot 2.5 software (written by Stuart Ray and distributed by the author at: http://www.med.jhu.edu/deptmed/sray/), similarity plots were developed to determine the percentage of similarity between selected pairs of sequences by moving a 500–base pair (bp) window along the genome alignment at 50-bp increments. Similarity values for each pair-wise comparison were plotted at the midpoint of the 500-bp segment.

The Simplot 2.5 software was also used to calculate bootscan plots. For the bootscan plots, the SimPlot software performed bootscanning on parsimony trees using SEQBOOT, DNAPARS, and CONSENSUS from the Phylip package (available at: http://evolution.genetics.washington.edu/phylip.html) for a 500-bp window moving along the alignment at increments of 50 bp [this approach is similar to the bootscanning method described by Salminen et al. (29)]. We evaluated 100 replicates for each phylogeny. The bootstrap values for the studied sequences were plotted at the midpoint of each window.

In the similarity and bootscan plots, the new sequences were compared with consensus sequences (50% threshold) representing the nonrecombinant subtypes from the same alignment as the phylogenetic tree, which was ungapped.

RESULTS

Analysis of the Full-Length Sequences

To verify whether the *gag* subtype A and *env* subtype G HIV-1 viruses that cluster with the prototype BFP-90 and 95ML-84 strains in these genomic regions have a similar complex mosaic genome, we sequenced the full-length genomes for two such strains. One virus was obtained from a symptomatic patient in Dakar, Senegal (97SE-1078), and a second one was obtained from an asymptomatic female sex worker in Bamako, Mali (95ML-127).

The two HIV-1 genomes were sequenced entirely, including the LTR extremities, 9808 bp for 97SE-1078 and 9719 bp for 95ML-127. All reading frames were open

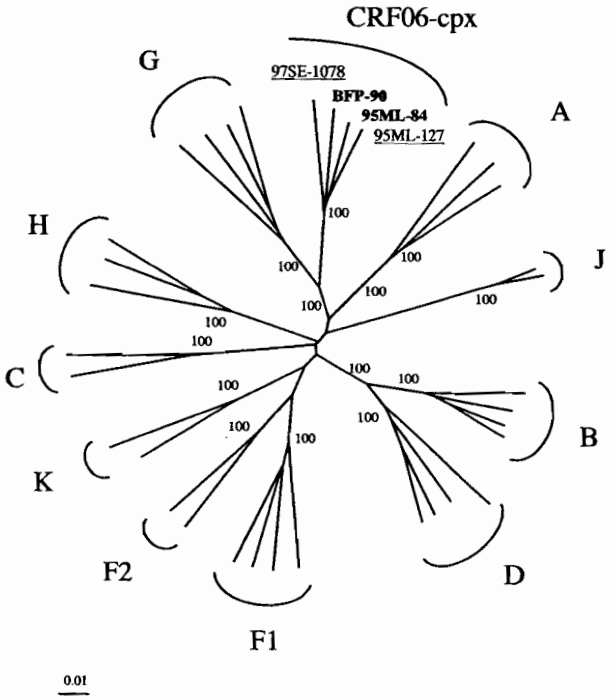


FIG. 1. Phylogenetic tree derived from a near-full-length genome alignment (7944 base pairs). The analysis and reference strains used are described in Methods section.

and of complete length. None of the genomes had major deletions or rearrangements. We also completed the near-full-length sequence of 95ML-84 by additional sequencing of the LTR extremities. In the LTR, the duplication after the TCF1-α site described in BFP-90 (16) was also present in 97SE-1078 and was increased by 9 bp; it was absent in 95ML-84 and 95ML-127. All previously described sites in LTR, TCF1-α, two NF-κB and three SP1, the TATA box, and the poly-A signal were present in the newly described isolates. The transactivation responsive region (TAR) element was conserved among the four viruses.

To determine the phylogenetic relations of the newly characterized viruses, we constructed evolutionary trees from the near-full-length genomes with reference strains from the database (see Methods section). The phylogenetic tree showed that 95ML-127 and 97SE-1078 formed a separate cluster with the BFP-90 and 95ML-84 prototype strains supported by high bootstrap values (100%) (Fig. 1).

Recombinant Analysis

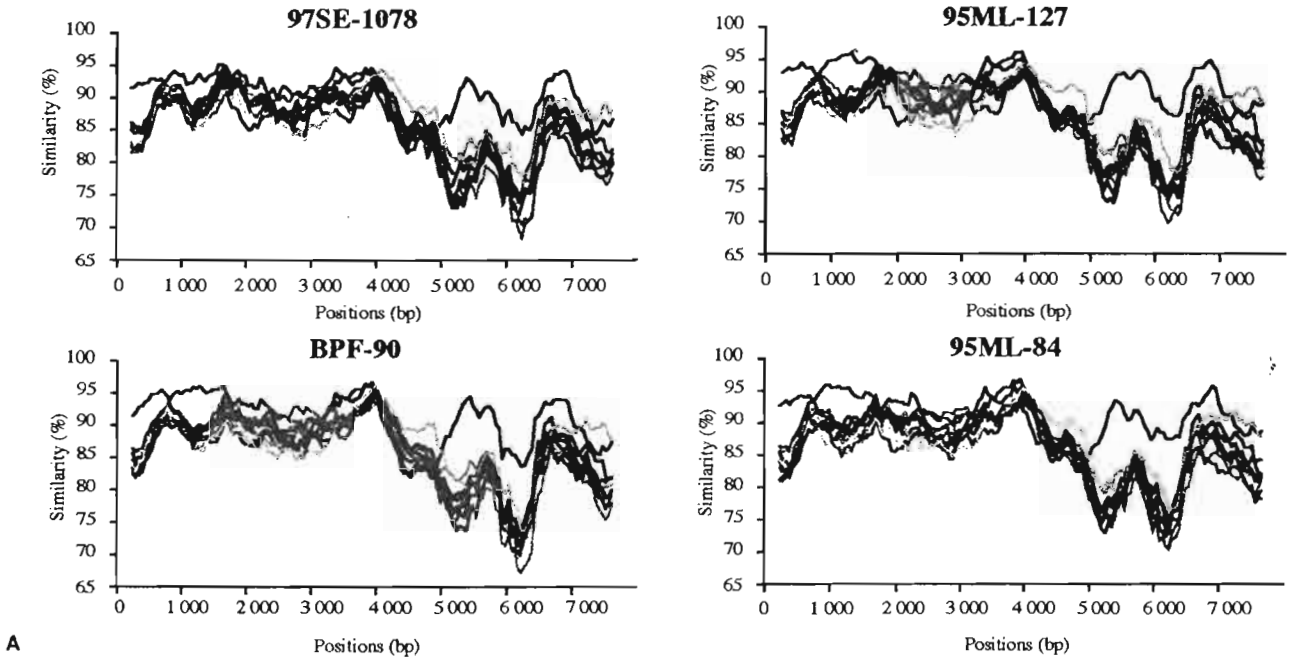
To determine whether 95ML-127 and 97SE-1078 have the same intersubtype recombinant structure as BFP-90 and 95ML-84, similarity plots were developed and bootscan analysis was done. Genomic alignment for the similarity plots was the same as that used for the phylogenetic tree construction. When the new sequences were plotted against references of all the nonrecombinant subtypes, we observed similar disproportionate levels of sequence identity with subtypes A, G, K, and J along their genome for the two sequences, which is consistent with the pattern observed when the BFP-90 and 95ML-84 prototype viruses were plotted against the same reference strains (Fig. 2A). The fragment previously described as subtype I in *pol* was now identified as subtype K. This result is consistent with the new evaluation of subtype I (3).

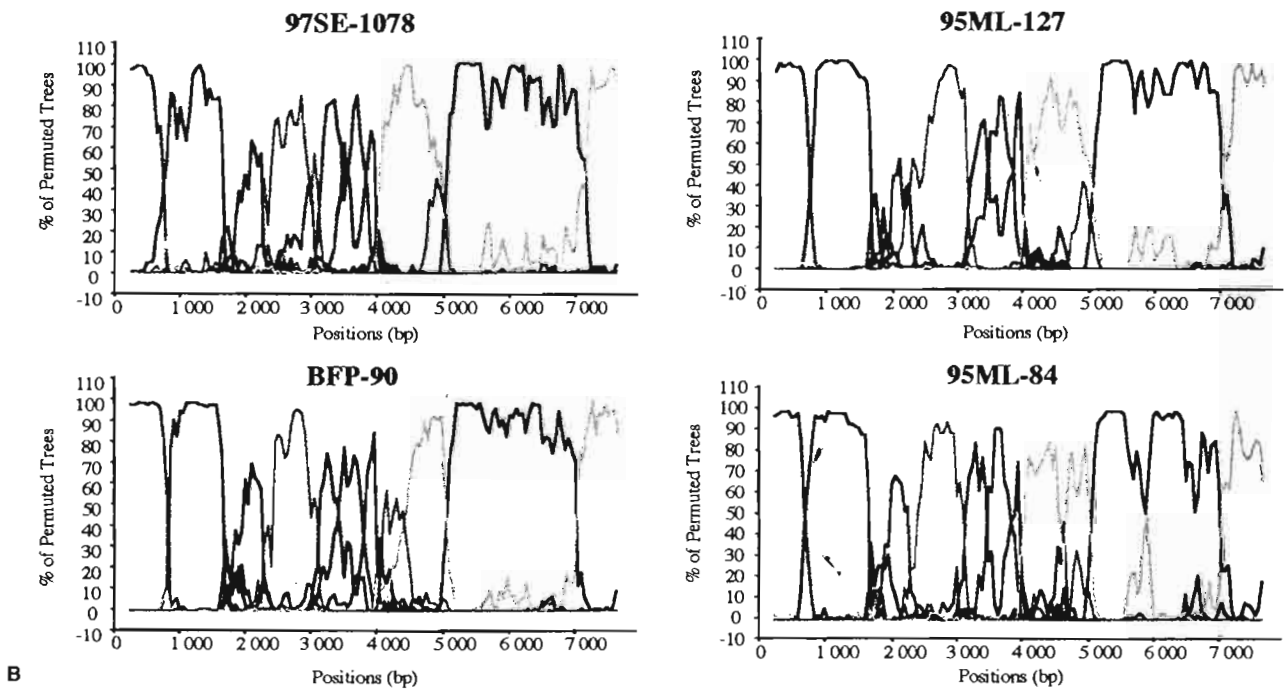
We systematically evaluated bootstrap values supporting independent monophyletic phylogeny in different parts of the genome using the same multiple genome alignment generated for the similarity plots and the phylogenetic tree of Figures 1 and 2A The results of the bootscan analysis are shown in Figure 2B. The complexity of these new strains is readily apparent, with alternating segments clustering with subtypes A, G, K, and J similar to the complexity observed for the BFP-90 and 95ML-84 prototype strains. The bootscan results are in agreement with similarity plots.

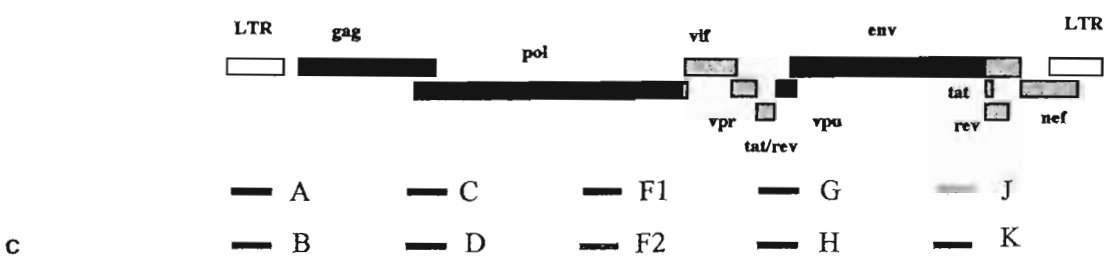
The locations of the breakpoints in the sequences were indicated on the similarity and bootscan plots, and the sequences were then broken down to different independent fragments corresponding to the inferred recombined regions and analyzed phylogenetically. This analysis confirmed the breakpoints and subtype designations, including the fragment in *pol* corresponding to the 5′ RT region for which the bootscan values were less significant (data not shown).

The overall genomic structure of this group of four viruses is shown in Figure 2C. The *gag* gene is subtype A over the segment coding p17 and part of p24, with the remaining portion being subtype G. The *pol* gene is subtype G in protease and the 5′ end of the RT region, whereas the 3′ end of the RT gene is classified as subtype K, and the integrase region is subtype G in the 5′ end and subtype A in the 3′ end. In the accessory gene region, the *vif*, *vpr*, and *tat* genes are subtype J. The *vpu* gene and almost the entire envelope gene are subtype G. The 3′ end of the envelope starting at the second exons of *tat* and *rev* as well as the entire *nef* gene are again subtype J.

In addition to the initial report describing the BFP-90 virus as an A/G/J/? recombinant and the 95ML-84 strain as an A/G/I/J recombinant, our data show that the region previously identified as unclassified or subtype I in *pol* corresponds to subtype K.


97SE-1078
95ML-127
BPF-90
95ML-84
Similarity (%)
Positions (bp)
A


97SE-1078
95ML-127
BFP-90
95ML-84
% of Permuted Trees
Positions (bp)
B


LTR
gag
pol
vif
env
LTR
tat
vpr
vpu
rev
nef
tat/rev
A
B
C
D
F1
F2
G
H
J
K
C

Identification of Related HIV-1 Strains in Other West African Countries

Phylogenetic tree analysis from smaller fragments dispersed over the genome showed that the 95ML-84, BFP-90, 97SE-1078, and 95ML-127 strains consistently form a separate and well-supported cluster. Genetic characterization of a *pol* fragment covering the protease and RT genes in HIV-1 samples from Senegal (99SN-127HPD, 98SN-64HALD, and 98SN-77HALD) and Ivory Coast (99IC-AB0302, 99IC-AB1307, 99IC-AB1545, and 99IC-AB2478) and from patients from Burkina Faso attending a hospital in France (99FR-MP1426 and 00FR-MP1471) showed that these strains formed a separate cluster, with 95ML-84, BFP-90, 97SE-1078, and 95ML-127 supported by high bootstrap values (97%) in the region studied (Fig. 3A).

As shown in Figure 2C, this region of the genome is recombinant with a breakpoint between G and K in the middle of the RT gene. Bootscanning was done on these nine *pol* sequences with sliding windows of 400 bp moving at 50-bp steps against nonrecombinant consensus reference sequences representing subtypes G, K, and B. This more detailed analysis showed similar breakpoints between the K and G sequences for eight of the nine strains (see Fig. 3B). Similar results were obtained when all the nonrecombinant subtypes were present in the analysis (data not shown). For sequence 99IC-AB1307, the breakpoint was at the same position, but the subtype designation of the 3′ end of the RT gene was less evident. When the BFP-90 strain was included in the bootscan analysis, the fragment was closest to this sequence over the entire fragment (data not shown), and phylogenetic tree analysis confirmed that the fragment clustered with subtype K.

We can extrapolate with a relatively high probability that these nine strains have the same mosaic A/G/K/J genome. Moreover, they were all subtype A in the p24 region of the *gag* gene and subtype G in the V3–V5 region of the *env* gene similar to the prototype strains BFP-90 and 95ML-84 (data not shown) and formed a separate and well-defined cluster within subtypes A and G.

DISCUSSION

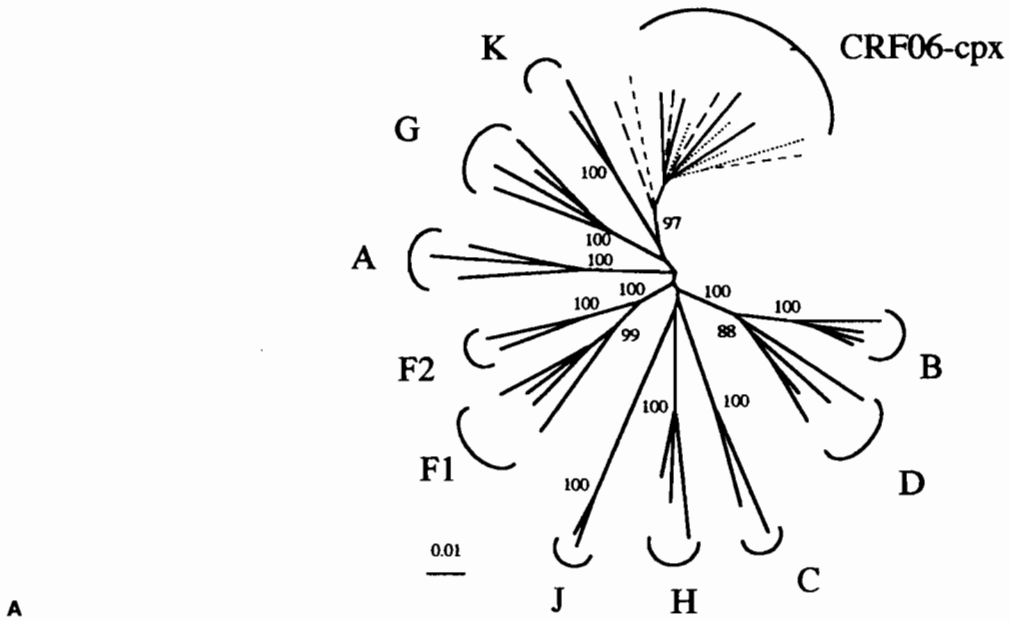
On the basis of the newly proposed nomenclature, the formal requirement for assigning a new CRF is the existence of at least three epidemiologically independent complete genome sequences that share the same recombinant structure and form a monophyletic cluster in all regions of the genome or two full-length genome sequences plus some smaller fragments that cluster with the full-length genome sequences and share identical breakpoints. The two additional full-length genome sequences described in this study confirm that the previously described strains, BFP-90 and 95ML-84, are indeed a new CRF of HIV-1, which we can now designate as CRF06-cpx, because four different subtypes were involved in the mosaic genome structure. This new CRF is composed of successive fragments of subtypes A, G, K, and J. The fragment in the *pol* gene, which was initially characterized as unknown in the BFP-90 strain and subsequently as subtype I in the 95ML-84 strain, is now clearly identified as subtype K. Subtype K sequences were unavailable at the time of the initial description of BFP-90, the prototype strain of this CRF.

Among the six known CRFs, subtype A is absent only in CRF05-FD. In addition, subtype G is present in three CRFs: CRF02-AG, CRF04-cpx, and CRF06-cpx.

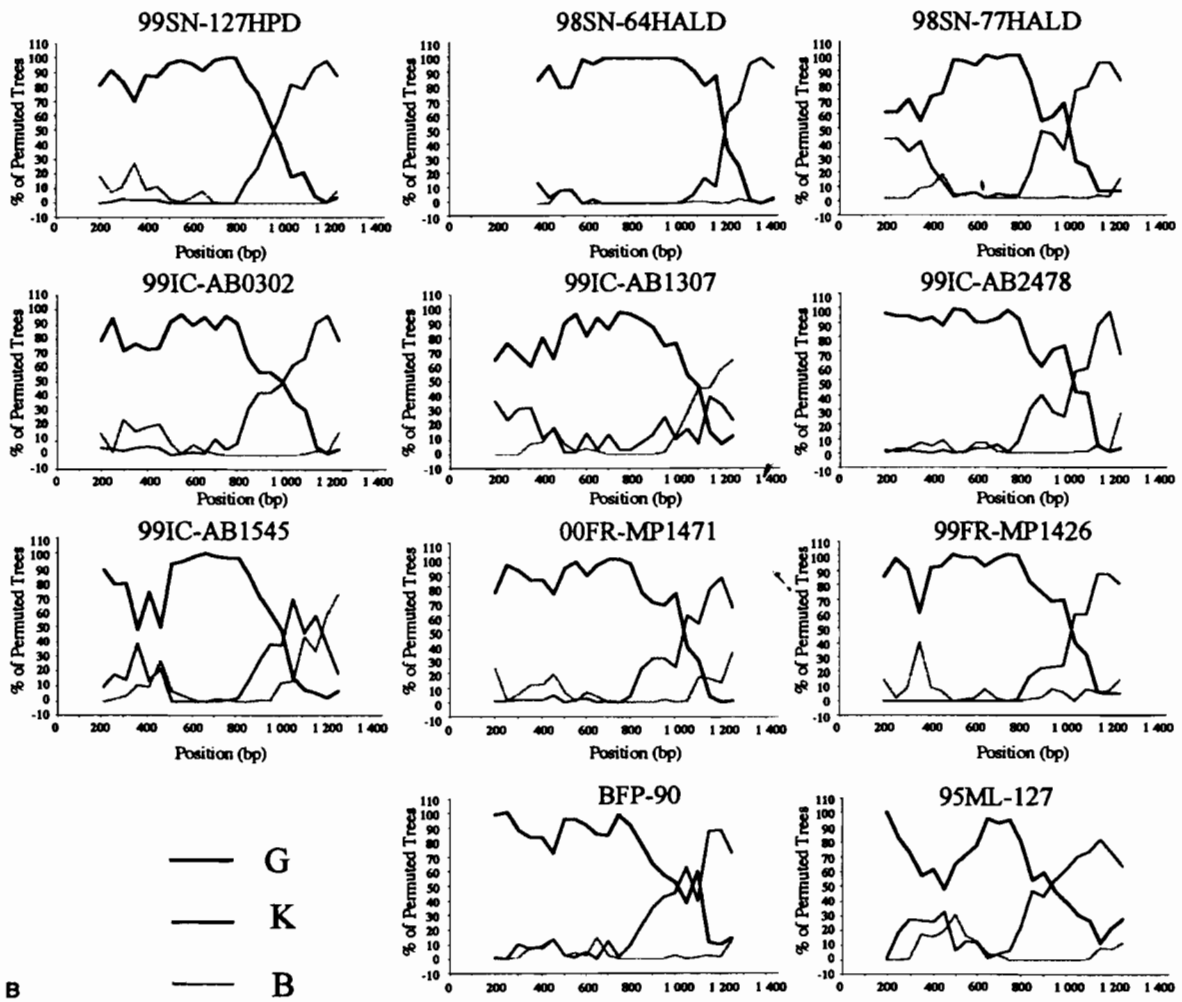
The presence of subtypes A and G in many CRFs in Africa is likely related to the high prevalence of subtype A (approximately 80%) in West Africa and Central Africa (approximately 50%). Subtype G is the second prevalent subtype documented in these regions of Africa, with a prevalence ranging from 10% to more than 30% (2,30).

In this and our previous study, we also report evidence for the presence of CRF06-cpx in several West African countries (17). CRF06-cpx circulates in Senegal, Mali, Burkina Faso, Ivory Coast, and Nigeria. The exact preva-

FIG. 2. Recombinant analysis of 95ML-127 and 97SE-1078. **(A)** Similarity plots comparing 95ML-127, 97SE-1078, BFP-90, and 95ML-84 with the consensus sequences of each nonrecombinant group M subtype. Similarity plots using Simplot 2.5 software determined the percentage of similarity between selected pairs of sequences by moving a 500-base pair (bp) window along the genome alignment at 50-bp increments. The similarity values for each pair-wise comparison were plotted at the midpoint of the 500-bp segment. Similar disproportionate levels of sequence identity with subtypes A (red), G (blue), K (orange), and J (yellow) were observed along their genome consistent with the pattern of BFP-90 and 95ML-84. **(B)** Bootscan analysis was done to evaluate systematically the bootstrap values supporting independent monophyletic phylogeny in different parts of the genome. For the bootscan plots, Simplot 2.5 software performed bootscanning on parsimony trees by using SEQBOOT, DNAPARS, and CONSENSUS from the Phylip package for a 500-bp window moving along the alignment at 50-bp increments. The bootstrap values for the studied sequences were plotted at the midpoint of each window. Successive A, G, K, and J fragments were observed. The pattern was the same as for BFP-90 and 95ML-84. **(C)** Schematic representation of the mosaic structure of the CRF06-cpx genome. The bootstrap values supporting the subtype assignments in the subregion trees were all greater than 80%.


K
CRF06-cpx
G
A
F2
F1
B
D
C
J
H
0.01
100
97
99
88

A


99SN-127HPD
98SN-64HALD
98SN-77HALD
99IC-AB0302
99IC-AB1307
99IC-AB2478
99IC-AB1545
00FR-MP1471
99FR-MP1426
BFP-90
95ML-127
% of Permuted Trees
Position (bp)
G
K
B

B

FIG. 3. Phylogenetic and recombinant analysis of nine partial *pol* sequences. **(A)** Phylogenetic tree from partial *pol* sequences (1452 base pairs [bp]). Our nine strains—99SN-127HPD, 98SN-64HALD, and 98SN-77HALD from Senegal (large gray dashes); 99IC-AB0302, 99IC-AB1307, 99IC-AB1545, and 99IC-AB2478 from Ivory Coast (small clear gray dashes); and 99FR-MP1426 and 00FR-MP1471 from patients infected in Burkina Faso who attended a hospital in France (gray dots)—form a cluster around BFP-90, 95ML-84, 95ML-127, and 97SE-1078 (gray lines) with a high bootstrap value (97%). **(B)** Bootscan analysis of the nine *pol* sequences with sliding windows of 400 bp moving at 50-bp steps compared with nonrecombinant consensus reference sequences representing subtypes G, K, and B. Similar results were obtained when all the nonrecombinant subtypes were included in the analysis.

lence of this virus remains to be determined, however. Importantly, this new variant was also introduced on other continents. Two cases were documented in patients from Burkina Faso attending a hospital in France, and the prototype strain BFP-90 was isolated in Australia from a patient who was also from Burkina Faso. This indicates that these viruses are spreading not only locally but globally.

In any geographic region, the proportion of recombinant viruses depends on a number of factors, including the prevalence rates of different subtypes, the probability that certain population groups acquire multiple infections and transmit their viruses further, and the fitness of any mosaic viruses generated. The frequency of recombinant viruses is almost certain to increase, however. Once it has occurred, recombination cannot be undone; thus, the frequency of "pure" subtype strains is likely to decrease. It is increasingly evident that the distribution of HIV-1 genetic subtypes is a dynamic and unpredictable process. The geographic distribution of subtypes is evolving, and intermixing of HIV-1 variants is inevitable. Recombinant viruses are already contributing substantially to the global pandemic, and the frequency and complexity of recombinant viruses are likely to increase as the various subtypes spread to all continents and viruses that are already mosaic are involved in further recombination events.

It is critical to determine whether there are biological differences among the various subtypes and CRFs, and it is important to track the molecular epidemiology of these viruses. More studies are needed to understand the role and implications of recombinant viruses in the global HIV evolution. It is important to study in detail the impact of viral recombination on viral properties, because recombination may introduce genetic and biological consequences that are far greater than those resulting from the steady accumulation of single mutations. It remains to be determined what impact the diversity caused by the formation of CRFs may have on the development of a broadly efficacious AIDS vaccine.

Acknowledgments: This work was sponsored by grants from the Agence Nationale de Recherches sur le SIDA (ANRS, Projet Sidak). C.M. is the recipient of a doctoral fellowship from ANRS.

SEQUENCES DATA

The new sequences have been deposited in the European Molecular Biology Laboratory (EMBL) Data Library under the following accession numbers:

Complete genome:
95ML-127: AJ288982
97SE-1078: AJ288981
Partial *pol* sequences:
98SN-64HALD: AJ286983
98SN-77HALD: AJ293337
99SN-127HPD: AJ293336
99FR-MP1426: AJ287033
00FR-MP1471: AJ293338
99IC-AB0302: AJ419631
99IC-AB1307: AJ419633
99IC-AB1545: AJ419634
99IC-AB2478: AJ419632

REFERENCES

1. Simon F, Mauclere P, Roques P, et al. Identification of a new human immunodeficiency virus type 1 distinct from group M and group O. *Nat Med* 1998;4:1032–7.
2. Peeters M, Sharp PM. Genetic diversity of HIV-1: the moving target. AIDS 2000, a year in review. *AIDS* 2000;14(suppl 3): S129–40.
3. Robertson DL, Anderson JP, Bradac JA, et al. HIV-1 nomenclature proposal: a reference guide to HIV-1 classification. In Kuiken CL, Foley B, Hahn B, et al., eds. *Human retroviruses and AIDS 1999: a compilation and analysis of nucleic acid and amino acid sequences*. Los Alamos, CA: 1999:492–505.
4. Carr JK, Salminen MO, Koch C, et al. Full-length sequence and mosaic structure of a human immunodeficiency virus type 1 isolate from Thailand. *J Virol* 1996;70:5935–43.
5. Gao F, Robertson DL, Morrison SG, 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* 1996b;70:7013–29.
6. Howard TM, Rasheed S. Genomic structure and nucleotide sequence analysis of a new HIV type 1 subtype A strain from Nigeria. *AIDS Res Hum Retroviruses* 1996;12:1413–25.
7. Carr JK, Salminen MO, Albert J, et al. Full genome sequences of human immunodeficiency virus type 1 subtypes G and A/G intersubtype recombinants. *Virology* 1998;247:22–31.
8. Montavon C, Toure-Kane C, Liegeois F, et al. Most *env* and *gag* subtype A HIV-1 viruses circulating in West and West Central Africa are similar to the prototype AG recombinant virus IBNG. *J Acquir Immune Defic Syndr* 2000;23:363–74.
9. Liitsola K, Tashkinova I, Laukkanen T, et al. HIV-1 genetic sub-

type A/B recombinant strain causing an explosive epidemic in injecting drug users in Kaliningrad. *AIDS* 1998;12:1907–19.
10. Bobkov A, Kazennova E, Selimova L, et al. A sudden epidemic of HIV type 1 among injecting drug users in the former Soviet Union: identification of subtype A, subtype B, and novel gagA/envB recombinants. *AIDS Res Hum Retroviruses* 1998;14:669–76.
11. Lukashov VV, Huismans R, Rakhmanova AG, et al. Circulation of subtype A and gagA/envB recombinant HIV type 1 strains among injecting drug users in St. Petersburg, Russia, correlates with geographical origin of infections. *AIDS Res Hum Retroviruses* 1999; 15:1577–83.
12. Kostrikis LG, Bagdades E, Cao Y, et al. Genetic analysis of human immunodeficiency virus type 1 strains from patients in Cyprus: identification of a new subtype designated subtype I. *J Virol* 1995; 69:122–30.
13. Gao F, Robertson DL, Carruthers CD, et al. 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* 1998;72:10234–41.
14. Nasioulas G, Paraskevis D, Magiorkinis E, et al. Molecular analysis of the full-length genome of HIV type 1 subtype I: evidence of A/G/I recombination. *AIDS Res Hum Retroviruses* 1999;15: 745–58.
15. Laukkanen T, Carr JK, Janssens W, et al. Virtually full-length subtype F and F/D recombinant HIV-1 from Africa and South America. *Virology* 2000;269:95–104.
16. Oelrichs RB, Workman C, Laukkanen T, et al. A novel subtype A/G/J recombinant full-length HIV type 1 genome from Burkina Faso. *AIDS Res Hum Retroviruses* 1998;14:1495–500.
17. Montavon C, Bibollet-Ruche F, Robertson D, et al. The identification of a complex A/G/I/J recombinant HIV type 1 virus in various West African countries. *AIDS Res Hum Retroviruses* 1999; 15:1707–12.
18. Peeters M, Koumare B, Mulanga C, et al. Genetic subtypes of HIV type 1 and HIV type 2 strains in commercial sex workers from Bamako, Mali. *AIDS Res Hum Retroviruses* 1998;14:51–8.
19. Fransen K, Zhong P, De Beenhouwer H, et al. Design and evaluation of new, highly sensitive and specific primers for polymerase chain reaction detection of HIV-1 infected primary lymphocytes [erratum]. *Mol Cell Probes* 1995;9:373.
20. Bibollet-Ruche F, Loussert-Ajaka I, Simon F, et al. Genetic characterization of accessory genes from human immunodeficiency virus type 1 group O strains. *AIDS Res Hum Retroviruses* 1998a;14:951–61.
21. Artenstein AW, Hegerich PA, Beyrer C, et al. Sequences and phylogenetic analysis of the *nef* gene from Thai subjects harboring subtype E HIV-1. *AIDS Res Hum Retroviruses* 1996;12:557–60.
22. Gao F, Morrison SG, Robertson DL, et al. Molecular cloning and analysis of functional envelope genes from human immunodeficiency virus type 1 sequence subtypes A through G. The WHO and NIAID Networks for HIV Isolation and Characterization. *J Virol* 1996a;70:1651–67.
23. Bibollet-Ruche F, Peeters M, Mboup S, et al. Molecular characterization of the envelope transmembrane glycoprotein of 13 new human immunodeficiency virus type 1 group O strains from six different African countries. *AIDS Res Hum Retroviruses* 1998b;14:1281–5.
24. Vergne L, Peeters M, Mpoudi-Ngole E, et al. Genetic diversity of protease and reverse transcriptase sequences of non-B HIV-1 strains: evidence for many minor drug resistant mutations in treatment-naïve individuals. *J Clin Microbiol* 2000;38:3919–25.
25. Hertogs K, de Bethune MP, Miller V, et al. A rapid method for simultaneous detection of phenotypic resistance to inhibitors of protease and reverse transcriptase in recombinant human immunodeficiency virus type 1 isolates from patients treated with antiretroviral drugs. *Antimicrob Agents Chemother* 1998;42:269–76.
26. Larder BA, Kohli A, Kellam P, et al. Quantitative detection of HIV-1 drug resistance mutations by automated DNA sequencing. *Nature* 1993;365:671–3.
27. Thompson JD, Higgins DG, Gibson TJ. 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;22:4673–80.
28. Kimura M. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J Mol Evol* 1980;16:111–20.
29. Salminen MO, Carr JK, Burke DS, et al. Identification of breakpoints in intergenotypic recombinants of HIV type 1 by bootscanning. *AIDS Res Hum Retroviruses* 1995;11:423–5.
30. Janssens W, Buve A, Nkengasong JN. The puzzle of HIV-1 subtypes in Africa [editorial]. *AIDS* 1997;11:705–12.