Virology **286,** 168–181 (2001)
doi:10.1006/viro.2001.0976, available online at http://www.idealibrary.com on IDEAL®

The AG Recombinant IbNG and Novel Strains of Group M HIV-1 Are Common in Cameroon

Jean K. Carr,*[,1] Judith N. Torimiro,† Nathan D. Wolfe,‡ Mpoudi Ngole Eitel,† Bohye Kim,* Eric Sanders-Buell,* Linda L. Jagodzinski,* Deanna Gotte,* Donald S. Burke,‡ Deborah L. Birx,§ and Francine E. McCutchan*

*Henry M. Jackson Foundation, Rockville, Maryland 20850; †Hopital Militaire de Yaounde, Yaounde, Cameroon; ‡Johns Hopkins School of Public Health, Baltimore, Maryland 21205; and §Walter Reed Army Institute of Research, Rockville, Maryland 20850

Received January 16, 2001; returned to author for revision March 27, 2001; accepted April 23, 2001

The genetic diversity of group M HIV-1 is highest in west central Africa. Blood samples from four locations in Cameroon were collected to determine the molecular epidemiology of HIV-1. The C2-V5 region of envelope was sequenced from 39 of the 40 samples collected, and 7 samples were sequenced across the genome. All strains belonged to group M of HIV-1. The circulating recombinant form CRF02_AG (IbNG) was the most common strain (22/39, 56%). Two of these were confirmed by full genome analysis. Four samples (4/39, 10%) clustered with the sub-subtype F2 and one of these was confirmed by full genome sequencing. Recombinant forms, each different but containing subtype A, accounted for the next most common form (7/39, 18%). Among these recombinants, those combining subtypes A and G were the most common (4/7, 57%). Also found were 3 subtype A, 2 subtype G, and 1 subtype B strain. Many recombination break points were shared between IbNG and the other AG recombinants, though none of these other AG recombinants included IbNG as a parent. This suggests that there was an ancestral AG recombinant that gave rise to CRF02_AG (IbNG), the successful circulating recombinant form, and to others that were less successful and are now rare. © 2001 Academic Press

INTRODUCTION

The high genetic diversity of group M HIV-1 is well established and the patterns of diversity are complex. In particular, the distribution of HIV-1 genetic subtypes on the African continent is a puzzle (Janssens *et al.,* 1997). The distribution or composition of subtypes on the western side of Africa is dramatically different from that on the eastern side. On the eastern side of Africa, subtype C is the most common and is associated with a high prevalence of infection in southern Africa and in Ethiopia (Hussein *et al.,* 2000; Karim and Karim, 1999). In Kenya and Uganda the most common subtypes are A and D, with very little subtype C (Kaleebu *et al.,* 2000; Neilson *et al.,* 1999). Only one circulating recombinant form (CRF), CRF10_CD (BFL061), has been identified on the eastern side of Africa. On the western coast of Africa, the most common HIV-1 strain is the circulating AG recombinant CRF02_AG (IbNG) (Andersson *et al.,* 1999; Carr *et al.,* 1999, 1998b; Mboudjeka *et al.,* 1999). The first full-length prototype for this CRF is the isolate "IbNG" which came from Ibadan, Nigeria (Howard and Rasheed, 1996; Robertson *et al.,* 2000).

Characterization of HIV-1 subtypes in African countries is ongoing but incomplete, particularly in west central Africa, where the diversity of subtypes is the highest. There have been several studies of the genetic diversity of HIV-1 in Cameroon in the past decade (Fonjungo *et al.,* 2000; Mboudjeka *et al.,* 1999; Nkengasong *et al.,* 1994; Takehisa *et al.,* 1998). In 1993, a sample was drawn from 17 patients in the capitol, Yaounde, and 61% of them were subtype A in env, consistent with CRF02_AG (Nkengasong *et al.,* 1994). In 1994 and 1995 a sample of 40 patients was drawn from hospitals throughout the country, and 65% were subtype A in env, and 60% also clustered with IbNG in pol (Takehisa *et al.,* 1998). In 1999, another study reported that 64% of the sample was CRF02_AG (Mboudjeka *et al.,* 1999). These studies concur in the view that the most prevalent subtype of HIV-1 in Cameroon is CRF02_AG. Several other studies found viruses that cluster significantly with subtype F but also were clearly distinct from subtype F. This genetic group has been designated sub-subtype "F2" (Triques *et al.,* 2000).

To further describe the genetic diversity in Cameroon, 40 new samples were collected from seropositive volunteers in rural and urban locations: (1) Douala, a large seaport city, (2) Yaounde, the capitol of the country, (3) Bertoua, a town in the southeast, and (4) a small village near Bafoussam in the west. Genetic diversity was high in all locations, both urban and rural. CRF01_AG and other AG recombinants with structures similar to that of CRF02_AG formed the most prevalent strains throughout Cameroon.

The views and opinions expressed herein do not necessarily reflect those of the U.S. Army or the Department of Defense.

[1] To whom correspondence and reprint requests should be addressed at Henry M. Jackson Foundation, 1 Taft Ct., Suite 250, Rockville, MD 20850. E-mail: jcarr@hivresearch.org.

0042-6822/01 $35.00
Copyright © 2001 by Academic Press
All rights of reproduction in any form reserved.

TABLE 1

Study Subjects and Subtype Designation

Viral sample	Location	Age	Sex	Transmission risk[a]	Viral load[b]	Region sequenced	Genotype	Novel[c]
CM52607	Yaounde	37	Female	Het	758,382	C2-V5 of env	G	
CM52608	Yaounde	25	Male	Het + scar	108,000	C2-V5 of env	CRF02.AG	
CM52610	Yaounde	nk	Female	nk	124,450	C2-V5 of env	G	
CM52876	Yaounde	37	Male	Het	183,685	Full-length genome	AF2G recomb	✓
CM52877	Yaounde	22	Male	Het	809,148	C2-V5 of env	CRF02.AG	
CM52878	Yaounde	37	Male	Het	1,841,084	C2-V5 of env	F2	
CM52879	Yaounde	31	Male	Het	126,022	C2-V5 of env	F2	
CM52881	Yaounde	25	Female	Het	127,241	C2-V5 of env	CRF02.AG	
CM52882	Yaounde	31	Female	Het	21,678	Unable to amplify	nd	
CM52883	Yaounde	36	Female	Het	72,692	C2-V5 of env	CRF02.AG	
CM52884	Yaounde	48	Male	Het	52,884	C2-V5 of env	CRF02.AG	
CM52885	Yaounde	43	Female	Het	9,145,867	Full-length genome	CRF02.AG	
CM52886	Yaounde	35	Female	Het	1,212,712	C2-V5 of env	CRF02.AG	
CM53122	Bertoua	46	Male	Transf.	591,004	gag through env	AE recomb	✓
CM53123	Bertoua	45	Male	Het	103,605	C2-V5 of env, protease	A	
CM53124	Bertoua	48	Male	Hom	105,174	C2-V5 of env, protease	CRF02.AG	
CM53125	Bertoua	44	Male	Het	nd	C2-V5 of env	CRF02.AG	
CM53136	Bertoua	nk	Male	Het	277,048	C2-V5-V4 of env	CRF02.AG	
CM53137	Bertoua	32	Male	Het	100,305	C2-V5 of env	CRF02.AG	
CM53138	Bertoua	39	Male	Hom	83,823	C2-V5 of env	CRF02.AG	
CM53375	Koutaba	39	Male	Het	146,309	env	AD recomb[d]	✓
CM53377	Koutaba	40	Male	Het	771,429	C2-V5 of env	CRF02.AG	
CM53379	Koutaba	46	Male	Het	153,488	Full-length genome	AFGHJU recomb	✓
CM53380	Koutaba	53	Male	Het	331,620	C2-V5 of env	CRF02.AG	
CM53385	Koutaba	48	Male	Het	596,278	C2-V5 of env	F2	
CM53386	Koutaba	35	Male	Het	66,409	C2-V5 of env, p24	CRF02.AG	
CM53388	Koutaba	37	Male	Het	16,205	C2-V5 of env	CRF02.AG	
CM53392	Koutaba	44	Male	Het	459,305	Full-length genome	AG recomb	✓
CM53393	Koutaba	30	Male	Het	1,747	C2-V5 of env	A	
CM53396	Koutaba	29	Male	Het	174,605	env	AGJ recomb	✓
CM53657	Douala	40	Male	Het	306,682	Full-length genome	F2	
CM53658	Douala	41	Male	Het	204,276	Full-length genome	CRF02.AG	
CM53659	Douala	43	Male	Het	1,350,000	C2-V5 of env	CRF02.AG	
CM53661	Douala	31	Male	Het	260,730	Protease, C2-V5 of env	AJ recomb[d]	✓
CM53662	Douala	45	Male	Hom	608,654	C2-V5 of env	B	
CM53663	Douala	28	Male	Het	15,104	C2-V5 of env	A	
CM53665	Douala	45	Male	Het	65,087	C2-V5 of env	CRF02.AG	
CM53666	Douala	35	Male	Het	66,367	C2-V5 of env	CRF02.AG	
CM53667	Douala	30	Male	Het	168,179	C2-V5 of env	CRF02.AG	
CM53668	Douala	31	Male	Het	310,340	C2-V5 of env	CRF02.AG	

[a] Het, heterosexual; Hom, homosexual; Transf, blood transfusion; scar, scarification; nk, not known.
[b] RNA copies/ml.
[c] 'Novel', first strain identified with this structure.
[d] Data not presented.

RESULTS

Military personnel in four locations of Cameroon were offered HIV testing and 40 tested seropositive. The seroprevalence rate among the volunteers ranged from 8.5 to 14.2% depending on the location. Because volunteers for testing were self-selected, the seroprevalence rates did not represent true population prevalences. In addition, the geographic location where the sampling was done did not necessarily reflect of the location of risk since the volunteers were transient military personnel, not long-term residents of the area. Most of the seropositive individuals were male, aged 30 to 50, and reported heterosexual risk factors (Table 1). The viral loads as measured by the Roche v1.5 assay, which is of equivalent sensitivity for most, if not all, HIV-1 subtypes (Michael *et al.,* 1999), were at levels normally found in untreated populations. The genotype of each sample is listed in Table 1 and refers only to the region(s) sequenced. Seven of the samples were new recombinant forms, designated "novel" because they have not been docu-

mented before and were each found in only a single individual.

The distribution of the genetic strains of HIV-1 in this population was examined according the demographic information available. No association was seen between genetic subtype and age or sex. In terms of geographic location there were no discernable differences between the sites. Koutaba, the most rural of the sites, was found to have the highest number of novel strains (4/7). Since the sample population consists of military personnel stationed in specific locations rather than an indigenous population, it was not surprising that there was uniformity across sites.

The V2-V5 region of envelope was amplified in 39 of 40 specimens, and these were sequenced. The remaining sample, CM52882, amplified by integrase primers used to distinguish M group from O group HIV-1 (Heyndrickx *et al.,* 1998), typed as group M. A phylogenetic tree with the genetic relationships between the 39 Cameroon partial envelope sequences and reference strains is shown in Fig. 1A. The majority of the sequences fell into subtype A. The C2-V5 region of env in CRF02_AG is subtype A; the 600-bp C2-V5 region is not sufficient to completely distinguish subtype A from CRF02_AG. For this reason, the sequences contained in the overall cluster with subtype A were analyzed individually. The 22 strains with a black box following the name (Fig. 1A) clustered significantly with IbNG (bootstrap value >70%), distinct from subtype A, in this analysis, representing 56% (22/39) of the sample. Additionally, 8 viruses did not cluster significantly with IbNG but were within subtype A; these were provisionally considered nonrecombinant subtype A (open squares, Fig. 1A). Four of the sequences (10%) clustered with subtype F2, a sub-subtype of F that has been found only in Cameroon (Triques *et al.,* 2000). Two viruses were subtype G (5%); the remaining strains clustered with subtypes B, D, and H.

Selected viral specimens were chosen for full genome characterization. First, representatives of CRF02_AG, the most abundant genotype in the sample, were examined (Fig. 1B). Two viruses that were classified as CRF02_AG on partial envelope analysis, CM53658 and CM52885, also clustered significantly with the reference sequences of this CRF, IbNG and DJ263 (bootstrap, 100%) in full genome analysis. Figures 2A and 2B show the distance scans of these viruses with the CRF02_AG prototype, IbNG. Both are closer to IbNG than they are to any other subtype throughout the genome. There are now six full genomes of CRF02_AG, one from Nigeria (Howard and Rasheed, 1996), two from Djibouti (Carr *et al.,* 1998b), one from Ivory Coast (Carr *et al.,* 1999), and two from Cameroon (this report).

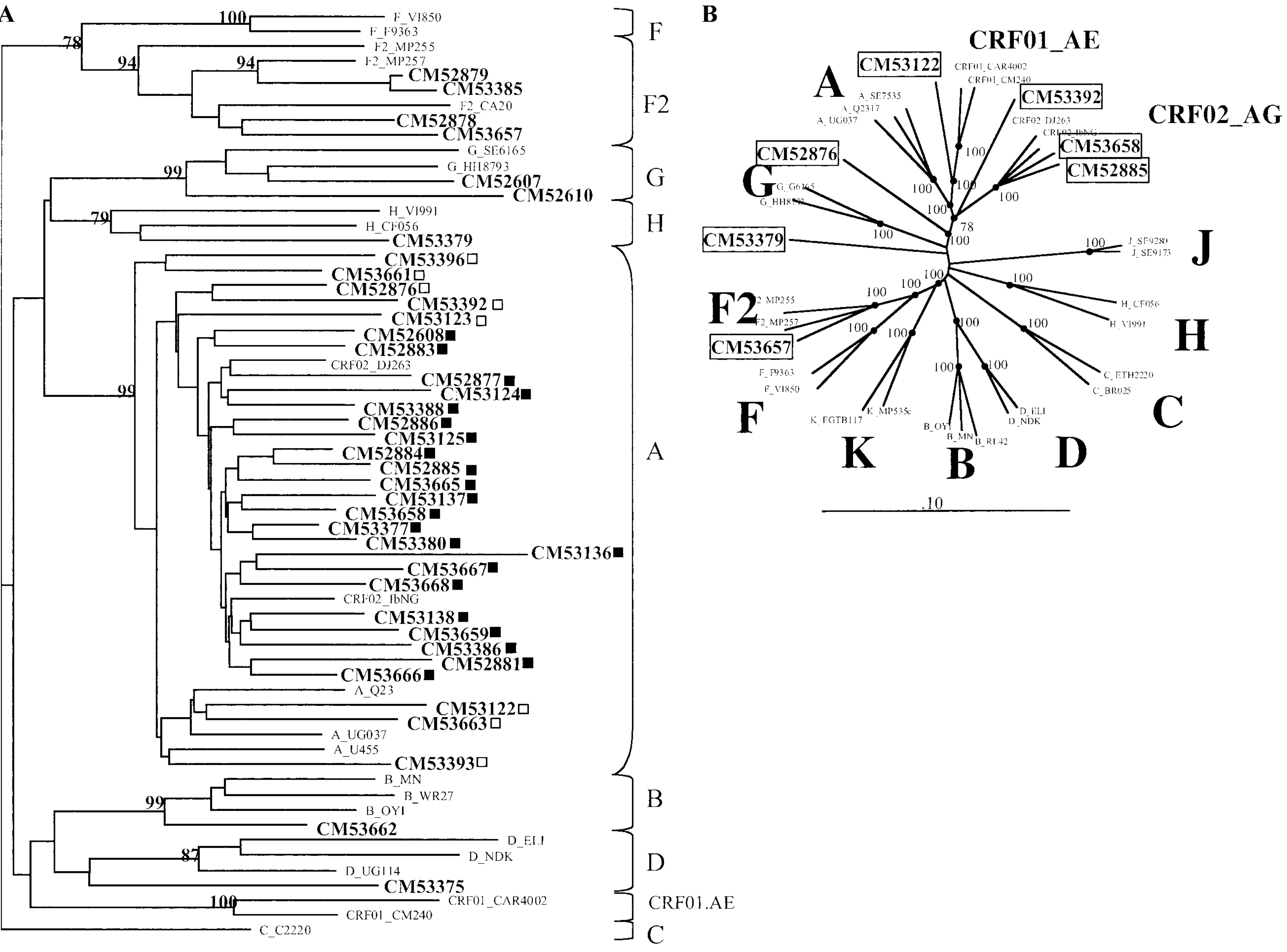
After CRF02_AG, the next most common strain in this sample set was F2, a subcluster of subtype F that has been found in Cameroon (Triques *et al.,* 2000), which formed 10% of the collection. As shown in Fig. 1B, CM53657 was sequenced in full and grouped with the reference F2 viruses, MP255 and MP257 (bootstrap, 100%). The results of the distance scan are shown in Fig. 3C, and it is clear that this sequence is similar to MP255, the F2 prototype, throughout the genome.

Four additional viruses that were within subtype A in C2-V5 but not classified as CRF02_AG (Fig. 1A) were sequenced in full. They were CM53122, CM52876, CM53379, and CM53392. Three were outliers in the full genome tree (Fig. 1B), different from each other and from all other known HIV-1 strains. The fourth strain, CM53122, formed a deep branch on the CRF01_AE lineage.

The recombinant break point analysis of these four strains is shown in Figs. 3 through 5, respectively. CM53122 was found to be a novel recombinant of CRF01_AE, a virus whose principal location in Africa is the Central African Republic. The genome of CM53122 can be divided into three segments (Fig. 3). From gag to mid-env, the virus clusters with CRF01_AE (bootstrap, 100%). This portion of the genome belongs to the CRF01_AE branch, rather than the nonrecombinant branch, of subtype A (Fig. 3A, left). In mid-env, the CM53122 sequence no longer resembles the CRF01_AE sequence. This piece joins nonrecombinant subtype A and extends from mid-gp120 to mid-gp41 (Fig. 3A, middle) (bootstrap, 100%). The 3′ end of the CM53122 sequence is CRF01.AE (Fig. 3A, right). This strain may represent a "backcross" between CRF01.AE and nonrecombinant subtype A.

Full genome analysis of CM53392, shown in Fig. 4, revealed a very interesting virus. Break point analysis of

FIG. 1. (A) Phylogenetic analysis of C2-V5 of env. Thirty-nine C2-V5 envelope sequences (shown in bold) from HIV-1 in Cameroon were phylogenetically analyzed with reference sequences from other subtypes in a neighbor joining tree with parsimony bootstrap (Felsenstein, 1989). A genetic distance of 10% is shown at the scale on the bottom of the figure. The bootstrap value of the node is placed by each significant node. Samples that clustered significantly (bootstrap >70%) with CRF02_AG in individual analyses are marked with a filled square; those which clustered with subtype A but not in the CRF02_AG subcluster are marked with an open square. (B) Phylogenetic analysis of nearly full-length CM53122, CM53392, CM52876, CM53379, CM53657, CM52885, and CM53658 genomes. Six Cameroon virtually full genome sequences and one sequence which was slightly shorter at the 3′ end (CM53122) were phylogenetically analyzed with reference sequences from each subtype and CRF01 and CRF02 in a neighbor joining tree with parsimony bootstrap (Felsenstein, 1989). A genetic distance of 10% is shown at the scale on the bottom of the figure. CM53657 joins with sub-subtype F2 (bootstrap, 100%), CM52885 and CM53658 cluster with CRF02_AG (bootstrap = 100%), and CM53122 clusters with CRF01_AE (see Results for full analysis of the strains).


A
F_VI850
F_F9363
F2_MP255
F2_MP257
CM52879
CM53385
F2_CA20
CM52878
CM53657
G_SE6165
G_HH8793
CM52607
CM52610
H_VI991
H_CF056
CM53379
CM53396□
CM53661□
CM52876□
CM53392 □
CM53123 □
CM52608■
CM52883■
CRF02_DJ263
CM52877■
CM53124■
CM53388■
CM52886■
CM53125■
CM52884■
CM52885 ■
CM53665 ■
CM53137■
CM53658 ■
CM53377■
CM53380 ■
CM53136■
CM53667■
CM53668■
CRF02_IbNG
CM53138 ■
CM53659■
CM53386 ■
CM52881■
CM53666■
A_Q23
CM53122□
CM53663□
A_UG037
A_U455
CM53393□
B_MN
B_WR27
B_OYI
CM53662
D_ELI
D_NDK
D_UG114
CM53375
CRF01_CAR4002
CRF01_CM240
C_C2220
0.10
F
F2
G
H
A
B
D
CRF01.AE
C
B
CRF01_AE
CRF02_AG
CM53122
CM53392
CM53658
CM52885
CM52876
CM53379
CM53657
A
G
J
F2
H
F
K
B
D
C
.10

the virus (not shown) suggested that it was an AG recombinant containing fragments of subtype A and subtype G as shown. Each fragment was confirmed by a significant bootstrap value (bootstrap >70%) in a phylogenetic analysis of the fragment with reference strains from all subtypes (Fig. 4). At the bottom of the figure the break point structure of CRF02_AG as represented by the prototypic IbNG is shown for comparison. The break point structure of CM53392 is almost identical to that of IbNG, with the exception of one fragment near the 3′ end of pol, which is subtype A in CM53392 and subtype G in IbNG (Fig. 4, second tree from the left, second row). Even though CM53392 is an AG recombinant with a nearly identical structure to IbNG, the full genome phylogenetic analysis shows that CM53392 is distinct from CRF02_AG (Fig. 1B). Careful examination of the subregion phylogenetic trees in Fig. 5 shows that while CM53392 clusters significantly in each segment with subtype A or subtype G, respectively, it is not actually contained in the subcluster with IbNG and the other CRF02 viruses (Fig. 4, all trees). The existence of this virus would suggest that other viruses with nearly the same recombinant break point pattern as CRF02 may have predated the epidemic spread of CRF02_AG in west and west central Africa.

The analysis of recombinant CM52876 is shown in Fig. 5A. The virus is a mosaic of subtypes A, G, and F2. There are three segments of the genome that cluster with subtype F2. In each case, the viral sequence clusters with the F2 subcluster of subtype F. Three segments cluster with subtype A but do not join the IbNG cluster and three other segments are from subtype G. This virus is a recombinant between subtypes A, F2, and G.

Recombinant analysis of CM53379 is shown in Fig. 5B. CM53379 is the most complex full genome analyzed in this study. It has identifiable segments derived from subtypes A, F, G, H, and J. A segment encompassing most of pol, vif, and vpr could not be classified despite detailed analysis with all known subtypes and CRF. The close interconnections of subtypes A and G can be seen in the last tree of the virus, where there is one large cluster containing subtype A, subtype G, CRF01, CRF02, and CM53379. Of note, the full envelope gene of CM53379, which contains segments of subtypes G, H, and A, clustered closely (data not shown) with a previously partially sequenced virus from Cameroon, CA13 (Nkengasong *et al.,* 1994). CM53379 could represent a CRF in Cameroon but at least two additional full genomes would be needed to establish this. In summary, the seven full genomes sequenced represented CRF02_AG (2), F2 (1), a novel AE recombinant related to CRF01_AE, and three other novel, complex AG-containing recombinant strains.

Another novel virus, CM53396, was explored by sequencing a 3.3-kb segment containing vpu, env, and part of nef. Phylogenetic analysis revealed it as a complex AGJ recombinant (Fig. 6A). Vpu was derived from subtype J, a segment near the beginning of gp120 derived from subtype G, and the remainder of gp120 was subtype A. The intracellular portion of gp41 was subtype J and the first 300 bp of nef were subtype A. The extracellular portion of gp41 was unclassifiable. Phylogenetic analysis of the gp120 region of CM53396 with available sequences from other countries showed that it clustered significantly (bootstrap, 96%) with three other viruses: CA1 from Cameroon (Nkengasong *et al.,* 1994), CAR4054 from the Central African Republic (Murphy *et al.,* 1993), and NG1923 from Nigeria (McCutchan *et al.,* 1999) (Fig. 8B). Also shown in the phylogenetic tree are the AGJ recombinants recently designated CRF06_AGJ (BFP90). It is clear that CM53396 is not a member of that CRF, but the presence of samples from other west central African countries that cluster with CM53396 in gp120 would suggest that it is also a CRF. Official designation of CRF status must await the full genomes of at least two more representatives of this recombinant, as mentioned above.

Two of the viruses from the collection were subjected to partial sequencing and were also found to be novel (Table 1, data not shown). The full envelope gene of CM53375 was sequenced and consisted of a mosaic between subtypes A and D. Partial envelope sequencing of CM53661 revealed it to be a subtype A sequence. The protease gene for CM53661 was subtype J, however. It is therefore considered a novel AJ recombinant.

CRF02 and the four novel AG-containing recombinants found in Cameroon all have break points at similar locations in the genome (Fig. 7): at the frame-shift location between gag and pol, at the active site of RT, at the end of pol (or beginning of vif), at the beginning and end of vpu, and at the membrane spanning region of gp41. In addition, many of the viruses have the same subtype in particular genome positions. For example, three of the viruses have gag from subtype A and three have a subtype G vpu. This could be attributed to a selective advantage to having certain genes from certain subtypes or to an ancestral relationship between the viruses, among other possibilities.

In summary, the majority of the strains (56%) in this sample set from Cameroon were the circulating recombinant form CRF02_AG, similar to the virus isolated in Nigeria, IbNG. Ten percent of the sample was the sub-subtype F2, 8% were subtype A other than IbNG, 5% were subtype G, and 3% were subtype B. Of note, the only subtype B infection was in one of the three people reporting homosexual risk factors. At least 18% of the viruses were novel recombinants first described here (Table 1, Fig. 7). This percentage is, in fact, a minimum estimate, since not all of the viral sequences were sequenced in full and there are undoubtedly other recombinants among those that are considered at the moment to be nonrecombinant. No recombinants were found between CRF02_AG and any other subtype, though most of the recombinants were, like CRF02_AG, recombinant be-

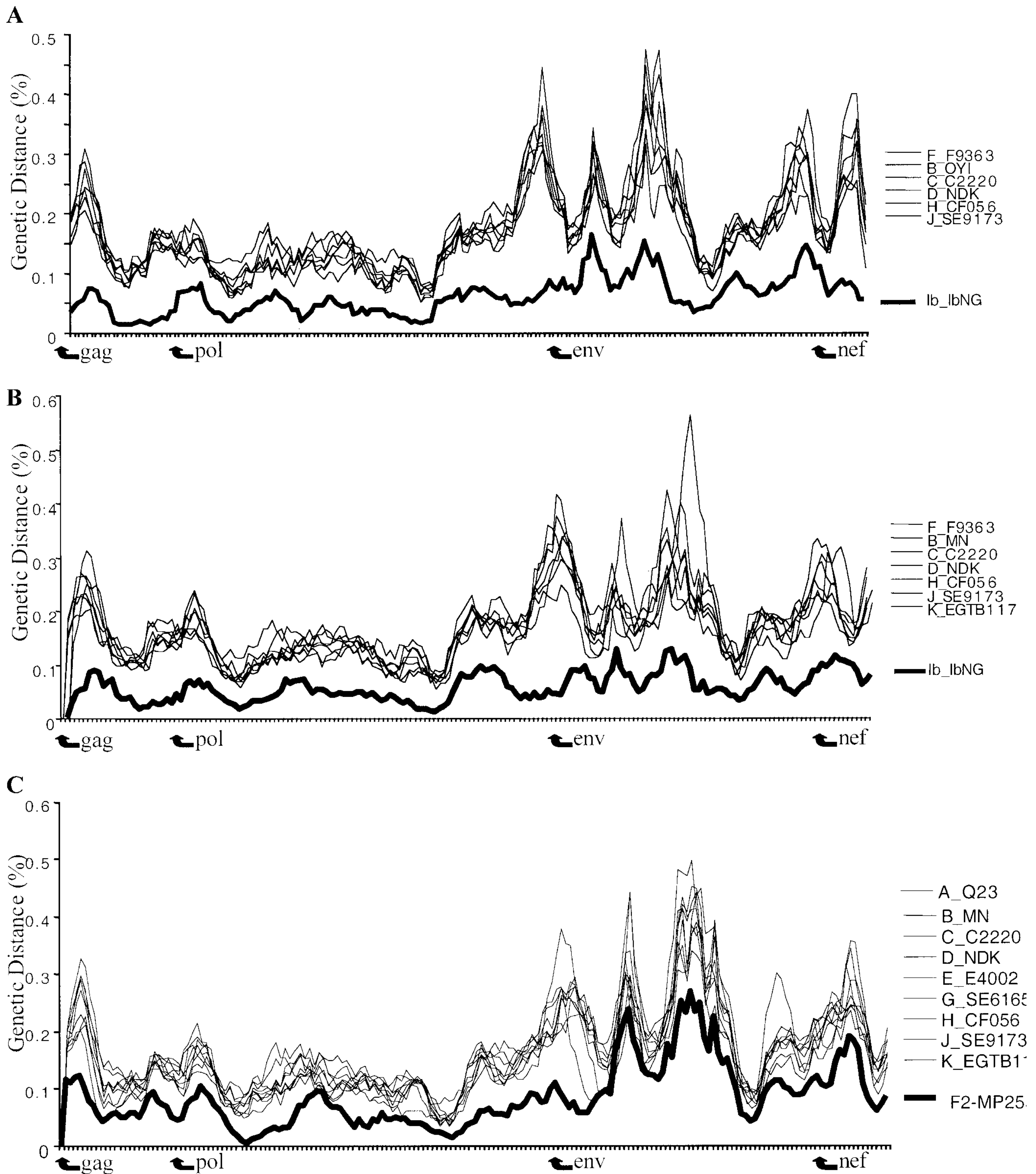


FIG. 2. Distance scan analysis of CM52885, CM53658, and CM53657. (A and B) Distance scans of CM52885 and CM53658, respectively. The pairwise distances between CM52885 (A) or CM53658 (B) and each subtype are plotted across the genome. Distances with CRF02_AG sample IbNG (thick black line) and other subtypes (thin black lines) are shown. (C) Distance scan of CM53657. The pairwise distances between CM53657 and each subtype are plotted across the genome of the virus. Distances with sub-subtype F2 MP255 (thick black line) and other subtypes (thin black lines) are shown.

tween subtypes A and G. The structure of strain CM53392 suggests that the IbNG pattern is widely represented and that CRF02_AG (IbNG) represents the epidemic spread of one of many similar recombinants. There was no evidence for any viral sequences outside of the M group of HIV-1.

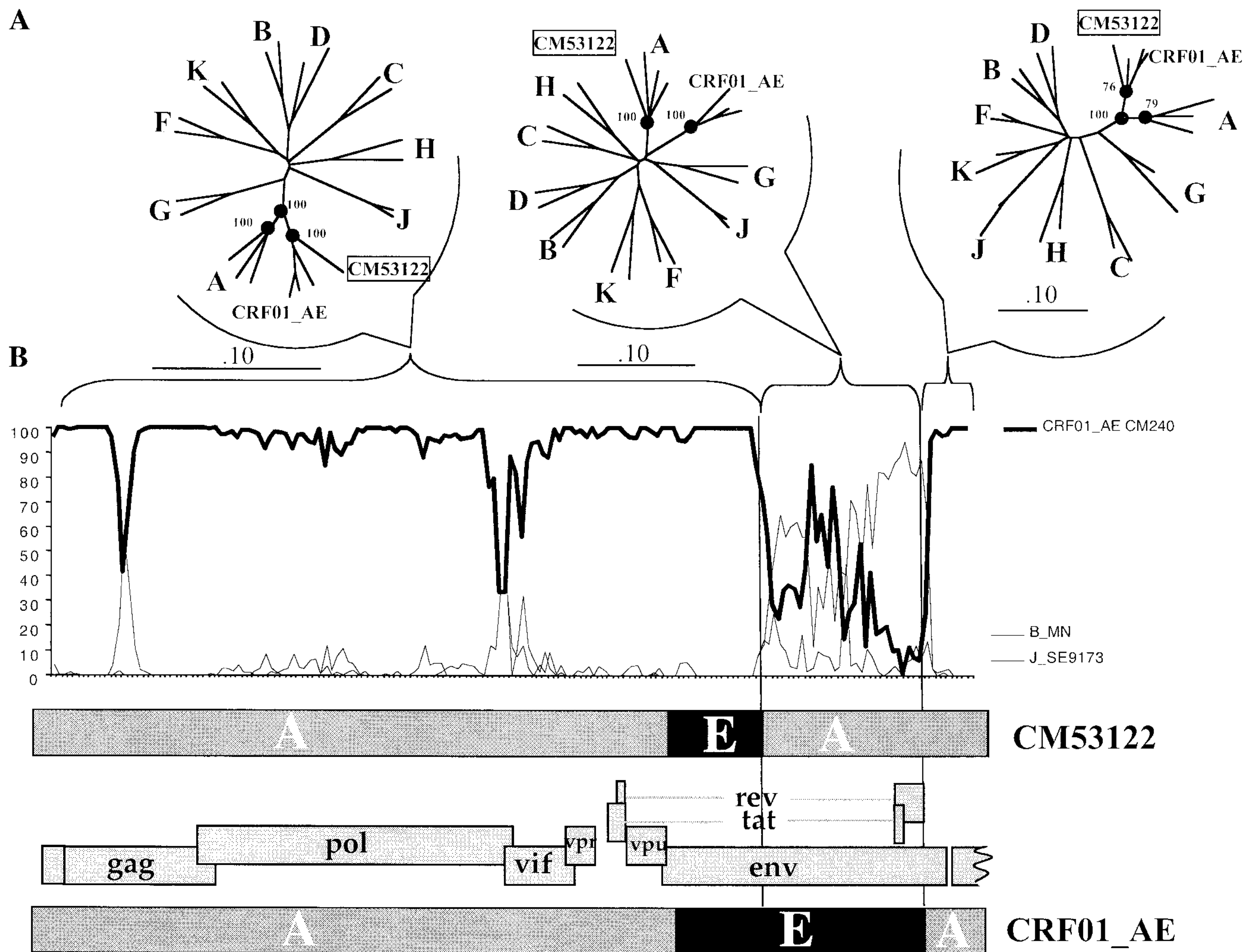


FIG. 3. Subtype analysis of CM53122. (A) Phylogenetic trees for CM53122 in the different regions of the genome. Neighbor joining trees with parsimony bootstrap values are shown for three segments of the genome. CM53122 clusters with CRF01.AE in the first segment (nt 796–7072 of HXB-2), with subtype A in the second segment (nt 7073–8567 of HXB-2), and with CRF01.AE in the third segment (nt 8568–9012 of HXB-2). (B) Bootscan of CM53122 with CRF01.AE, subtype B and subtype J strains. The bootstrap values joining CM53122 with CRF01.AE (CM240) are shown as a thick black line. The bootstrap values joining CM53122 with subtype B and subtype J are thin black lines. Below the panel is a diagram of the structure of the genome in relationship to CRF01_AE.

DISCUSSION

Forty blood samples from four locations in Cameroon were subjected to partial envelope amplification by PCR and 39 (98%) were successfully amplified. The most common genetic form was the AG recombinant, CRF02_AG, which resembles the virus from Ibadan, Nigeria, IbNG (Carr *et al.,* 1998b; Howard and Rasheed, 1996). Twenty-two samples, or 56% (Table 1 and Fig. 1A) were classified as CRF02_AG by partial envelope sequencing and two were confirmed by full genome sequencing (Figs. 1B, 2A, and 2B). Other studies have also found that the most common form of HIV-1 in Cameroon is CRF02_AG (Mboudjeka *et al.,* 1999; Takehisa *et al.,* 1998). In addition to its prominence in Cameroon, this CRF is probably also the most common form of HIV-1 in Guinea-Bissau (Andersson *et al.,* 1999), the Ivory Coast (Ellenberger *et al.,* 1999), and Nigeria (McCutchan *et al.,* 1999). The next most common strain in this sample set is the sub-subtype F2, of which there were 4 (10%). This genetic form of HIV-1 has recently been defined by full genome sequencing of several viruses from Cameroon (Triques *et al.,* 2000), and one of the four viruses in this collection was confirmed as F2 by full genome analysis (Figs. 1B and 3C).

Seven of the samples in this collection (18%) were novel viruses not previously described (Table 1 and Fig. 7). Two of the samples, CM53379 (Fig. 5) and CM53396 (Fig. 6), clustered with other partial genome sequences from central west Africa and may represent newly discovered CRFs. The remaining five were novel recombinants unlike any other sequences currently available.

A relatively comprehensive analysis of 39 HIV-1 sero-

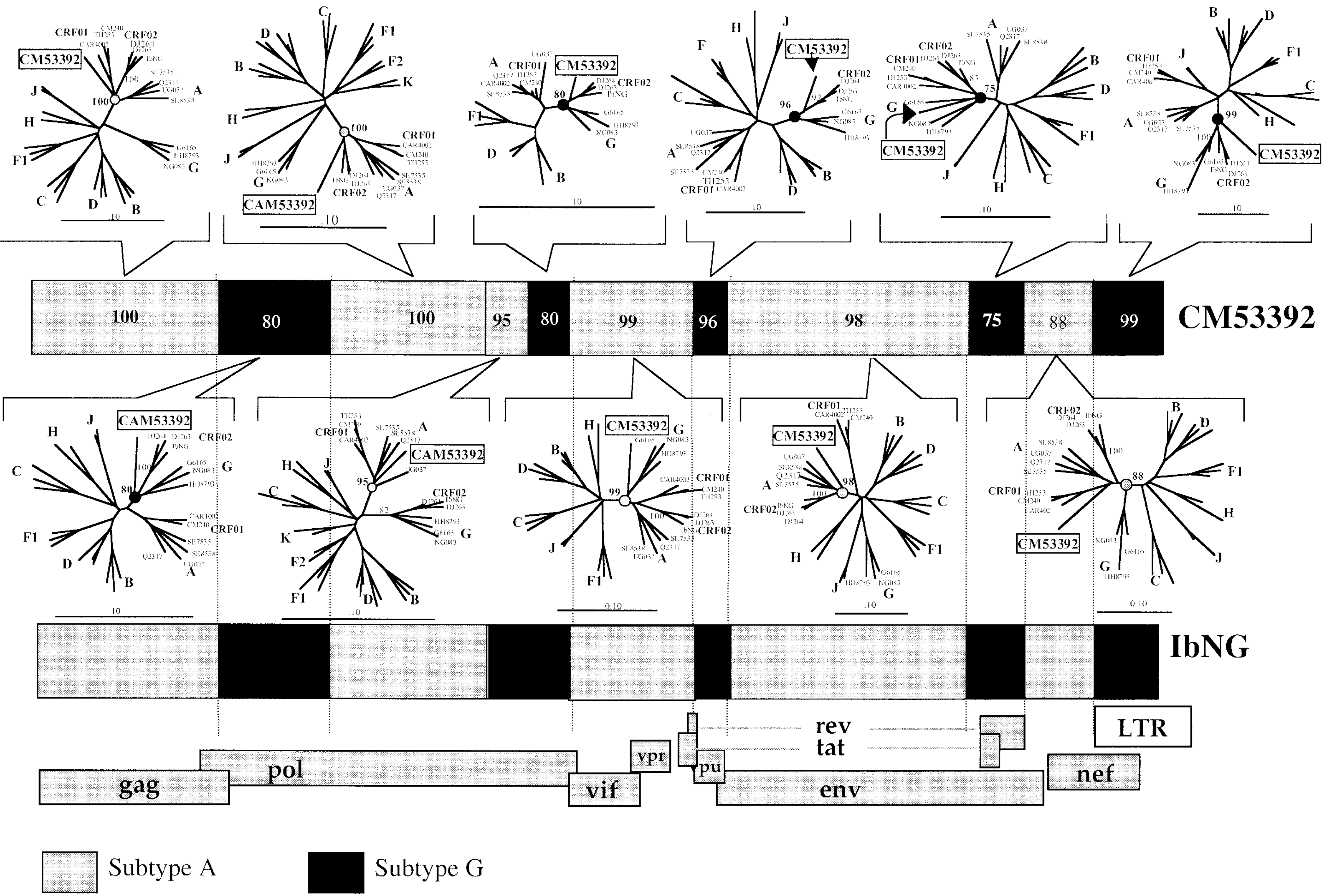
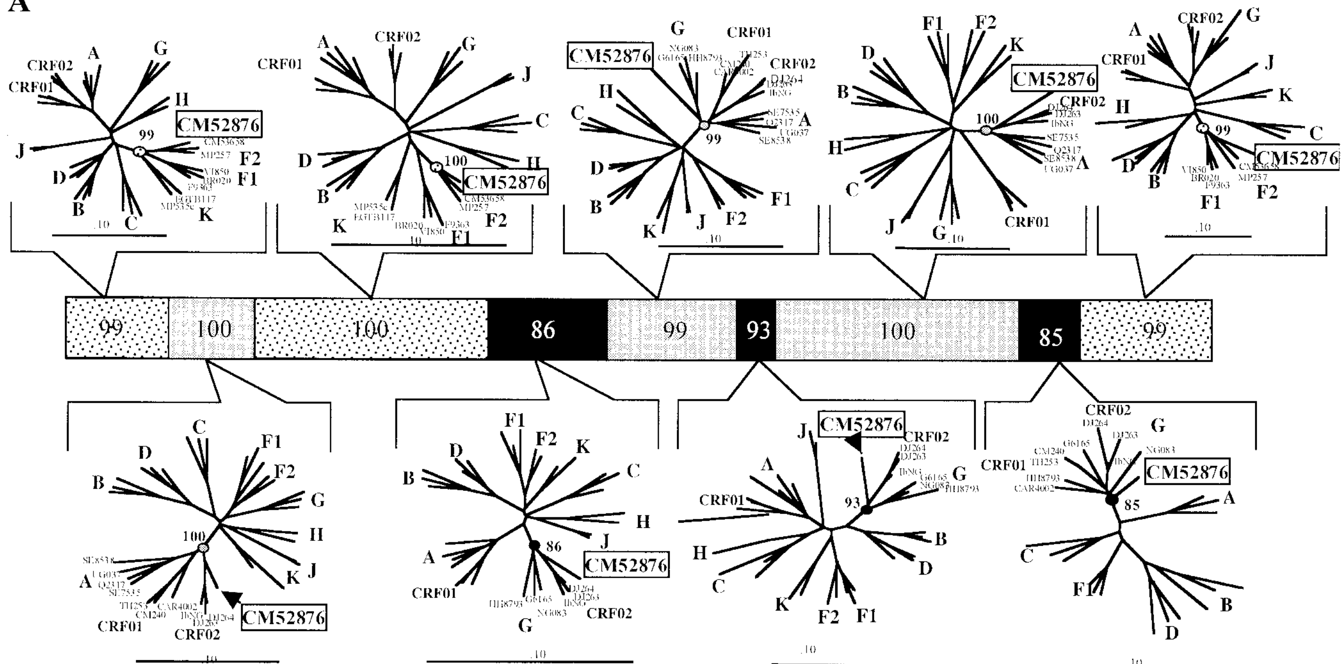


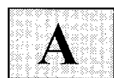
FIG. 4. Subtype analysis of CM53392. Neighbor joining phylogenetic trees with parsimony bootstrap values are shown for each segment of the genome. The bootstrap value joining CM53392 to the subtype of the fragment is indicated. There are 10 segments of subtype A and G; with respect to HXB-2 (Genbank Accession No. K03455), the segments are A, 796–2198; G, 2199–3078; A, 3079–4753; G, 4754–4981; A, 4982–5997; G, 5998–6278; A, 6279–8148; G, 8149–8620; A, 8621–9085; G, 9086–9606. For comparison, the segments of subtypes A and G in CRF02_AG (IbNG) are shown at the bottom of the figure. Subtype A, gray; subtype G, black.



B

CM53379
100
77
96
CAM53379
93
100 86 77 86 96 79 93
86
86
79
.10

gag
pol
vif
vpr
vpu
rev
tat
env
nef
LTR



positive samples from Cameroon has revealed the presence of at least seven new genotypes (18%). All of these are recombinants between subtype A and other subtypes, and most are recombinants involving subtypes A and G, but they are all clearly distinct from the CRF02 viruses (Figs. 3–7). In fact, it is striking that there were no recombinants observed between CRF02 and any other subtype, despite the fact that CRF02 is the most common subtype. This may be due only to sampling because full genome sequencing was done with only two of the samples that were CRF02 in C2-V5 of envelope. There is only one report of recombinants between CRF02 and other subtypes (Janssens *et al.,* 2000) and it may be that they are not common. It would be valuable to examine this issue in future studies to see if there are biological features of the virus or the immune response that tend to suppress recombination between this CRF and other strains.

Most of the recombinants found in this study are AG-containing recombinants. They have a common pattern of break points across the genome and often also have a "preference" for specific subtypes in specific genes. The pattern is one that is shared by the most common genetic form in the country, CRF02_AG. This commonality might be due to independent, evolutionary selection of specific subtypes for specific genetic functions or may be due simply to the presence of an ancestral lineage that is detectable to a greater or lesser extent in all of the progeny. In a location of rich viral diversity such as Cameroon there is still much to be learned about group M HIV-1 evolution and its distribution in populations.

METHODS

Subjects

Military personnel in Cameroon were offered free, anonymous HIV testing and self-selected volunteers were tested. Four army bases were selected for study: Douala, Yaounde, Bertoua, and Koutaba (Table 1). Douala, which harbors a major seaport on the Atlantic Ocean, is situated on the estuary of the Wouri River. Yaounde is the capitol of the country; Bertoua is a large town on the road connecting Yaounde to the capitol of the neighboring Central African Republic. Koutaba is a small village close to Bafoussam in the western part of the country. Blood was drawn into EDTA tubes and plasma was separated from peripheral blood mononuclear cells (PBMC) on Ficoll gradients. The plasma was tested for HIV antibodies by ELISA (Bioelisa HIV1 + 2, BIOKIT S.A., Spain). Reactive samples were confirmed by two different Western blots (HIV Blot 2.2, Genelabs Diagnostics, Singapore and Calypte, Cambridge Biotech, Cambridge, MA) and those confirmed positive on both were used for viral load determination and DNA extraction. Viral load was measured using the Roche Amplicor HIV-1 monitor test, v. 1.5. High-molecular-weight DNA was extracted from the PBMC using QIAmp DNA extraction kits (Qiagen, Valencia, CA).

PCR

The DNA from PBMC was amplified by nested PCR in the C2-V5 region of env, generating an amplicon of approximately 600 bp. The primers used were AA750 5′-GGATCAAAGYCTAAARCCATGTGT-3′ and AA1570 5′-GGAGCAGCAGGAAGCACTAGTGGC-3′, followed by AAV3 5′-GCCAGTRGTATCAACTCAAYTGC-3′ and AA1400 5′-TTTCCTCCTCCAGGTCTGAAG-3′ (Artenstein *et al.,* 1995). Cycling conditions were 94°C for 45 s, 55°C for 30 s, and 72°C for 2.5 min for 27 cycles, followed by a 5-min extension at 72°C. Alternatively, primers ED5 and ED12, followed by ES7 and ES8, were used as described elsewhere (Bachmann *et al.,* 1994).

Some samples were selected for full genome sequencing, including several that appeared to be unique by partial sequencing. Limiting template dilution into the first round was performed to decrease the complexity of the sample and allow for direct sequencing of the second round PCR product. The virtually full-length genome was amplified using MSF12b 5′-AAATCTCTAGCAGTGGCGCCCGAACAG-3′ and OFMR1, 5′-TGAGGGATCTCTAGTTACCAGAGTC 3′, followed by F2nst 5′-GCGGAGGCTAGAAGGAGAGAGATGG-3′ and ofm19 5′-GCACTCAAGGCAAGCTTTATTGAGGCTTA-3′. PCR was performed as described (Salminen *et al.,* 1995b), using the Expand Long Template kit (Boehringer Mannheim) and a hot-start method with a melting wax barrier (Dynawax). Cycling conditions were 94°C for 2 min and then 10 cycles of 94°C for 10 s, 60°C for 30 s, and 68°C for 8 min. This was followed by 20 cycles where the annealing temperature was 55°C. The final extension step was 68°C for 10

FIG. 5. (A) Subtype analysis of CM52876. Neighbor joining phylogenetic trees with parsimony bootstrap values are shown for each of the segments of the genome. The bootstrap value joining CM52876 to the subtype of the fragment is indicated. There are nine segments of subtypes A, F2, and G; with respect to HXB-2, the segments are F2, 796–1606; A, 1607–2198; F2, 2199–4096; G, 4097–4981; A, 4982–5997; G, 5998–6278; A, 6279–8148; G, 8149–8525; F2, 8526–9606. For comparison the genes of HIV-1 are shown at the bottom of the figure. (B) Subtype analysis of CM53379. Neighbor joining phylogenetic trees with parsimony bootstrap values are shown for each of the segments of the genome. The bootstrap value joining CM53379 to the subtype of the fragment is indicated. There are at least seven segments of subtypes A, F, G, H, and J; with respect to HXB-2, the segments are A, 796–2068; G, 2069–2586; F, 2587–3078; unclassified (U), 3079–5997; J, 5998–6278; G, 6279–6778; H, 6779–7758; A, 7759–9606. For comparison, the genes of HIV-1 are shown at the bottom of the figure.

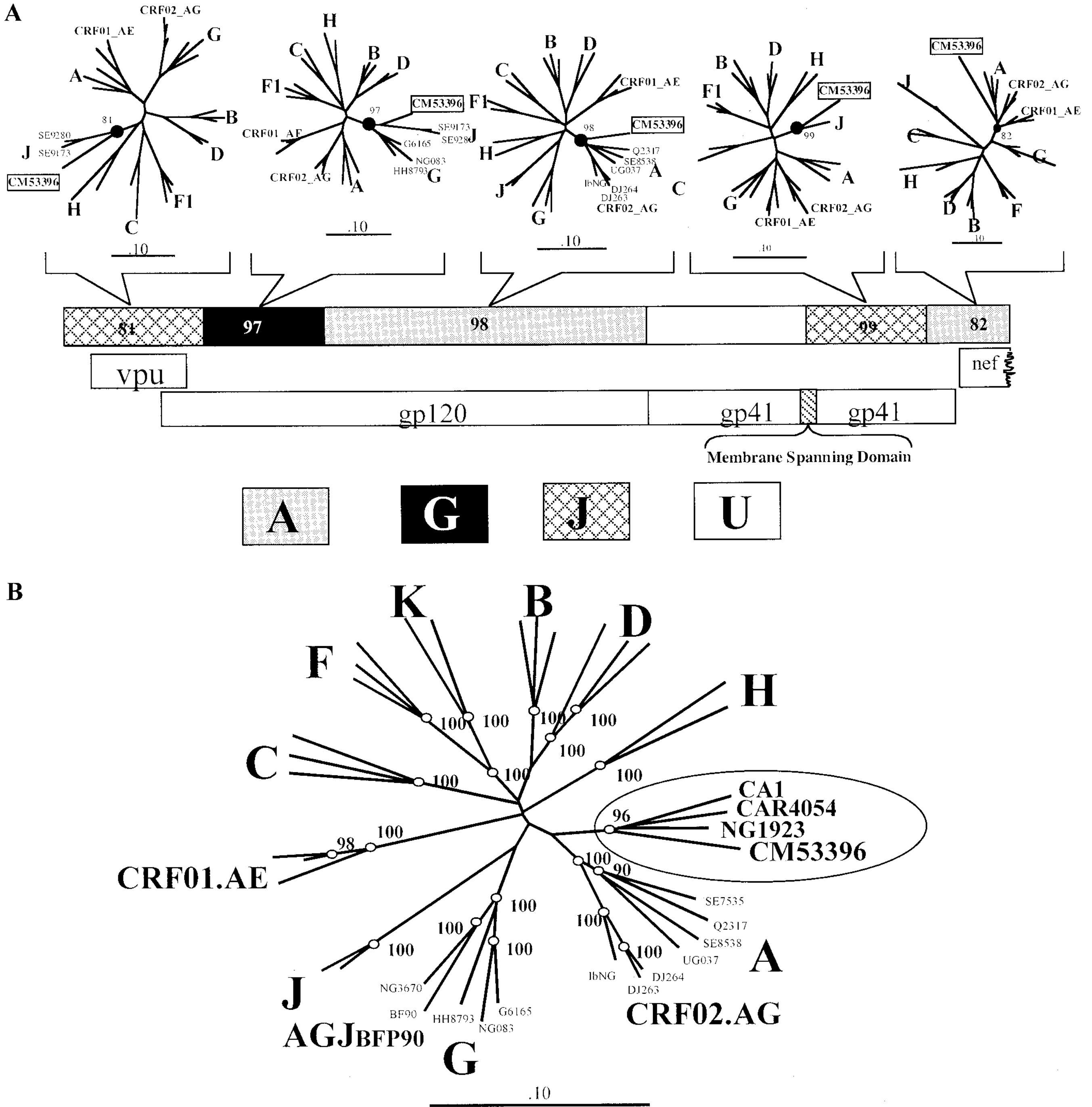


FIG. 6. Subtype analysis of a partial genome segment of CM53396. (A) Neighbor joining phylogenetic trees with parsimony bootstrap values are shown for each of the segments of the genome. The bootstrap value joining CM53379 to the subtype of the fragment is indicated. There are at least six segments of subtypes A, G, and J; with respect to HXB-2 (Accession No. K03455), the segments are J, 5972–6278; G, 6279–6778; A, 6779–7758; unclassified (U), 7759–8327; J, 8328–8772; A, 8773–9005. For comparison the genes of HIV-1 are shown below. (B) Phylogenetic analysis of gp120 of envelope. Neighbor joining tree with parsimony bootstrap values showing the relationship between CM53396, CA1 (Nkengasong *et al.,* 1994), CAR4054 (Murphy *et al.,* 1993), and NG1923 (McCutchan *et al.,* 1999) with a significant bootstrap value (96%).

min. Multiple second round PCR amplifications were combined to provide sufficient template for sequencing.

DNA sequencing

DNA template for automated sequencing was prepared by purification in an Amicon-100 column. The DNA was fully sequenced on both strands using fluorescent dye terminators and an Applied BioSystems (Applied Biosystems Inc, Foster City, CA) Model 373A or 377 DNA sequencer. DNA sequences were assembled using Sequencher software (Genecodes Inc., Ann Arbor MI) on Macintosh computers. All sequence ambiguities were resolved.

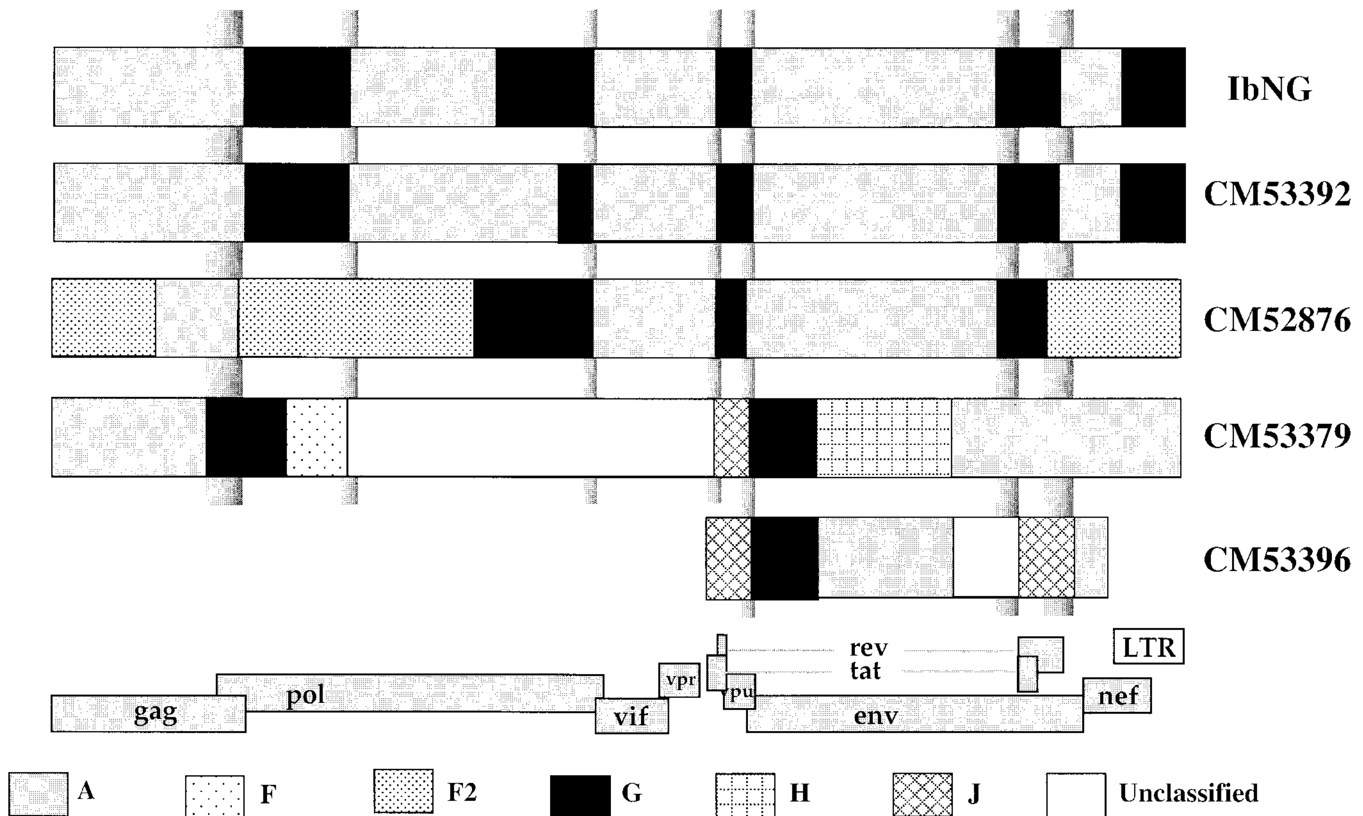


FIG. 7. Comparative structure of novel AG-containing recombinants from Cameroon. The four new AG-containing recombinants described in this report are diagrammed below the structure of CRF02_AG, IbNG. Break points common to at least three of the recombinants are marked with vertical bars and the corresponding locations of the genes of HIV-1 are at the bottom of the figure.

Analysis

A multiple alignment of the C2-V5 env sequences with selected reference sequences was constructed. Reference sequences of subtype A (U455 and 92UG037 from Uganda, Q2317 from Kenya), subtype B (MN and WR27 from the United States and OYI from Gabon), subtype C (ETH2220 from Ethiopia and BR025 from Brazil), subtype D (ELI, NDK from the Democratic Republic of Congo (DRC) and UG114 from Uganda), subtype F1 (FIN9363 from Finland and VI850 from DRC), subtype F2 (MP257, MP255, and CA20 from Cameroon), subtype G (SE6165 and HH8793 from the Congo and Kenya, respectively), subtype H (VI991 and CF056 from DRC and the Central African Republic, respectively), subtype J (SE9280 and SE9173 from DRC) (Carr *et al.,* 1998a), and subtype K (EGTB117 and MP535 from Cameroon) (Triques *et al.,* 2000) were used. In addition to the subtypes, CRFs were used as references, including CRF01.AE (CM240, 93TH253 from Thailand, and 90CR402 from the Central African Republic) and CRF02_AG (DJ263 from Djibouti and IbNG from Nigeria) (Carr *et al.,* 1998a).

Genetic relationships can be obscured by the presence of recombinant and novel forms in the analysis of HIV-1 strains. To address this, phylogenetic trees were constructed that included only one unknown viral sequence at a time. The genetic subtype of each new strain was assessed using bootstrap analysis (Felsenstein, 1985).

Distance scanning and bootscanning were used to determine the presence of recombination and to locate break points (Carr *et al.,* 1996; Salminen *et al.,* 1995a). In these techniques, the multiple alignment was broken into overlapping segments of equal length (300 nt) and each segment was phylogenetically analyzed. The distance scan was performed by computing, for each segment, the genetic distances between the known subtypes or CRFs and the unknown sequence using the maximum likelihood method (Felsenstein, 1981). This distance was then plotted at the corresponding position on the genome. “Bootscanning,” an analytic approach which tracks the bootstrap value of the node joining an unknown sequence with sequences of known subtypes progressively across the genome, was also used to identify putative recombination break points (Salminen *et al.,* 1995a). A bootstrap value equal to or greater than 70% was considered definitive (Hillis and Bull, 1993). Using these two scanning techniques, sequences were identified as recombinant and recombinant break points were identified. Once the break points were identified, the segments were then examined in separate phylogenetic

analyses and the phylogenetic trees were used to test the subtype assignment. The locations of the break points are indicated with respect to their location in the reference strain HXB2 (Genbank Accession No. K03455).

Phylogenetic trees were constructed and the consistency of branching order was evaluated using the NEIGHBOR, CONSENSE, and DNAPARS modules of the Phylip Package (V3.52c) (Felsenstein, 1989) and TREE-TOOL (Maciukenas, 1994). Each segment was analyzed by building a phylogenetic tree using neighbor joining (Saitou and Nei, 1987) or maximum parsimony (Felsenstein, 1996) and the stability of the nodes assessed by bootstrap analysis (Felsenstein, 1985).

Nucleotide sequence Accession numbers

The sequences obtained here were submitted to Genbank and are available under the following Accession numbers: AF372366–AF372404 and AF377954–AF377959.

ACKNOWLEDGMENT

The authors thank Hassan Zahwa for the Western blots.

REFERENCES

Andersson, S., Norrgren, H., Dias, F., Biberfeld, G., and Albert, J. (1999). Molecular characterization of human immunodeficiency virus (HIV)-1 and -2 in individuals from Guinea-Bissau with single or dual infections: Predominance of a distinct HIV-1 subtype A/G recombinant in West Africa. *Virology* **262**(2), 312–320.

Artenstein, A. W., VanCott, T. C., Mascola, J. R., Carr, J. K., Hegerich, P. A., Gaywee, J., Sanders-Buell, E., Robb, M. L., Dayhoff, D. E., Thitivichianlert, S., Nitayaphan, S., McNeil, J. G., Birx, D. L., Michael, R. A., Burke, D. S., and McCutchan, F. E. (1995). Dual infection with human immunodeficiency virus type 1 of distinct envelope subtypes in humans. *J. Infect. Dis.* **171,** 805–810.

Bachmann, M. H., Delwart, E. L., Shpaer, E. G., Lingenfelter, P., Singal, R., Mullins, J. I., and Character, W. N. H. I. (1994). Rapid genetic characterization of HIV type 1 strains from four World Health Organization-sponsored vaccine evaluation sites using a heteroduplex mobility assay. *AIDS Res. Hum. Retroviruses* **10,** 1345–1353.

Carr, J. K., Foley, B. T., Leitner, T., Salminen, M., Korber, B., and McCutchan, F. (1998a). Reference sequences representing the principal genetic diversity of HIV-1 in the pandemic. *In* "Human Retroviruses and AIDS" (B. Korber, C. L. Kuiken, B., Foley, B., Hahn, F. McCutchan, J. W. Mellors, and J. Sodroski, Eds.). Theoretical Biology and Biophysics Group, Los Alamos NM.

Carr, J. K., Laukkanen, T., Salminen, M. O., Albert, J., Alaeus, A., Kim, B., Sanders-Buell, E., Birx, D. L., and McCutchan, F. E. (1999). Characterization of subtype A HIV-1 from Africa by full genome sequencing. *AIDS* **13**(14), 1819–1826.

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

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.

Ellenberger, D. L., Pieniazek, D., Nkengasong, J., Luo, C. C., Devare, S., Maurice, C., Janini, M., Ramos, A., Fridlund, C., Hu, D. J., Coulibaly, I. M., Ekpini, E., Wiktor, S. Z., Greenberg, A. E., Schochetman, G., and Rayfield, M. A. (1999). Genetic analysis of human immunodeficiency virus in Abidjan, Ivory Coast reveals predominance of HIV type 1 subtype A and introduction of subtype G. *AIDS Res. Hum. Retroviruses* **15,** 3–9.

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

Felsenstein, J. (1985). Confidence limits of phylogenies: An approach using the bootstrap. *Evolution* **39,** 783–791.

Felsenstein, J. (1989). PHYLIP-phylogenetic inference package (version 3.2). *Cladistics* **5,** 164–166.

Felsenstein, J. (1996). Inferring phylogenies from protein sequences by parsimony, distance, and likelihood methods. *Methods Enzymol.* **266,** 418–427.

Fonjungo, P., Mpoudi, E., Torimiro, J., Alemnji, G., Eno, L., Nkengasong, J., Gao, F., Rayfield, M., Folks, T., Pieniazek, D., and Lal, R. (2000). Presence of diverse Human Immunodeficiency Virus Type 1 viral variants in Cameroon. *AIDS Res. Hum. Retroviruses* **16**(13), 1319–1324.

Heyndrickx, L., Janssens, W., Gürtler, L., Zekeng, L., Loussert-Ajaka, I., Vereecken, K., Willems, B., Coppens, S., Ndumbe, P., Fransen, K., Saman, E., Alary, M., and van der Groen, G. (1998). Differential diagnosis of HIV type 1 group O and M infection by polymerase chain reaction and PstI restriction analysis of the pol gene fragment. *AIDS Res. Hum. Retroviruses* **14,** 973–977.

Hillis, D. M., and Bull, J. J. (1993). An emperical test of bootstrapping as a method for assessing confidence in phylogenetic trees. *Syst. Biol.* **42,** 182–192.

Howard, T. M., and Rasheed, S. (1996). Genomic structure and nucleotide sequence analysis of a new HIV type 1 subtype A strain from Nigeria. *AIDS Res. Hum. Retroviruses* **12,** 1413–1425.

Hussein, M., Abebe, A., Pollakis, G., Brouwer, M., Petros, B., Fontanet, A. L., and de Wit, T. F. (2000). HIV-1 subtype C in commercial sex workers in Addis Ababa, Ethiopia. *J. Acquired Immune Defic. Syndr.* **23**(2), 120–127.

Janssens, W., Buve, A., and Nkengason, J. N. (1997). The puzzle of HIV-1 subtypes in Africa [editorial]. *AIDS* **11,** 705–712.

Janssens, W., Salminen, M., Laukkanen, T., Heyndrickx, L., Auwere, G. v. d., Colebunders, R., McCutchan, F. E., and Groen, G. v. d. (2000). Near full-length genome analysis of HIV Type 1 CRF02_AG Subtype C and CRF02.AG subtype G recombinants. *AIDS Res. Hum. Retroviruses* **16**(12), 1183–1189.

Kaleebu, P., Whitworth, J., Hamilton, L., Rutebemberwa, A., Lyagoba, F., Morgan, D., Duffield, M., Biryahwaho, B., Magambo, B., and Oram, J. (2000). Molecular epidemiology of HIV type 1 in a rural community in southwest Uganda. *AIDS Res. Hum. Retroviruses* **16**(5), 393–401.

Karim, Q. A., and Karim, S. S. A. (1999). Epidemiology of HIV infection in South Africa. *AIDS* **13**(6), 4–7.

Maciukenas, S. (1994). "Treetool" v.2.0.1: ed. Ribosomal RNA Database Project, University of Illinois Board of Trustees.

Mboudjeka, I., Zekeng, L., Takehisa, J., Miura, T., Ido, E., Yamashita, M., Kaptue, L., and Hayami, M. (1999). HIV type 1 genetic variability in the northern part of Cameroon. *AIDS Res. Hum. Retroviruses* **15**(11), 951–956.

McCutchan, F. E., Carr, J. K., Bajani, M., Sanders-Buell, E., Harry, T. O., Stoeckli, T. C., Robbins, K. E., Gashau, W., Nasidi, A., Janssens, W., and Kalish, M. L. (1999). Subtype G and multiple forms of A/G intersubtype recombinant human immunodeficiency virus type 1 in Nigeria. *Virology* **254,** 226–234.

Michael, N. L., Herman, S. A., Kwok, S., Dreyer, K., Wang, J., Christopherson, C., Spadoro, J. P., Young, K. K., Polonis, V., McCutchan, F. E., Carr, J., Mascola, J. R., Jagodzinski, L. L., and Robb, M. L. (1999). Development of calibrated viral load standards for group M subtypes of human immunodeficiency virus type 1 and performance of an improved AMPLICOR HIV-1 MONITOR test with isolates of diverse subtypes. *J. Clin. Microbiol.* **37**(8), 2557–2563.

Murphy, E., Korber, B., Georges-Courbot, M.-C., You, B., Pinter, A., Cook,

D., Kieny, M.-P., Georges, A., Mathiot, C., Barre-Sinoussi, F., and Girard, M. (1993). Diversity of V3 region sequences of human immunodeficiency viruses type 1 from the Central African Republic. *AIDS Res. Hum. Retroviruses* **9,** 997–1006.

Neilson, J. R., John, G. C., Carr, J. K., Lewis, P., Kreiss, J. K., Jackson, S., Nduati, R. W., Mbori-Ngacha, D., Panteleeff, D. D., Bodrug, S., Giachetti, C., Bott, M. A., Richardson, B. A., Bwayo, J., Ndinya-Achola, J., and Overbaugh, J. (1999). Subtypes of human immunodeficiency virus type 1 and disease stage among women in Nairobi, Kenya. *J. Virol.* **73**(5), 4393–4403.

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.

Robertson, D. L., Anderson, J. P., Bradac, J. A., Carr, J. K., Foley, B., Funkhouser, R. K., Gao, F., Hahn, B. H., Kalish, M. L., Kuiken, C., Learn, G. H., Leitner, T., McCutchan, F., Osmanov, S., Peeters, M., Pieniazek, D., Salminen, M., Sharp, P. M., Wolinsky, S., and Korber, B. (2000). HIV-1 nomenclature proposal [letter]. *Science* **288**(5463), 55–56.

Saitou, N., and Nei, M. (1987). The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4,** 406–425.

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

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

Takehisa, J., Zekeng, L., Ido, E., Mboudjeka, I., Moriyama, H., Miura, T., Yamashita, M., Gürtler, L. G., Hayami, M., and Kaptué, L. (1998). Various types of HIV mixed infections in Cameroon. *Virology* **245,** 1–10.

Triques, K., Bourgeois, A., Vidal, N., Mpoudi-Ngole, E., Mulanga-Kabeya, C., Nzilambi, N., Torimiro, N., Saman, E., Delaporte, E., and Peeters, M. (2000). Near-full-length genome sequencing of divergent African HIV type 1 subtype F viruses leads to the identification of a new HIV type 1 subtype designated K. *AIDS Res. Hum. Retroviruses* **16**(2), 139–151.