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

*Celine Montavon, †Coumba Toure-Kane, *Florian Liegeois, ‡Eitel Mpoudi, *Anke Bourgeois, *Laurence Vergne, §Jean-Luc Perret, ‖Annie Boumah, ¶Eric Saman, †Souleymane Mboup, *#Eric Delaporte, and *Martine Peeters

*Laboratoire Retrovirus, Institut pour la Recherche en Developement, Montpellier, France; †Laboratoire de Microbiologie, Hopital Le Dantec, Dakar, Senegal; ‡Projet PRESICA, Military Hospital, Yaounde, Cameroon; §Institut de Medecine Tropicale du Service Santé des Armées, Marseilles, France; ‖National AIDS Control Program, Libreville, Gabon; ¶Innogenetics, Ghent, Belgium; and #Service de Maladies Infectieuses et Tropicales, CHU Gui de Chaulhiac, Montpellier, France

Summary: The genetic subtype was identified in *gag* and *env* of 219 HIV-1–positive samples collected in different African countries, 44 from Senegal, 55 from Cameroon, 82 from Gabon, and 38 from Djibouti. In total, 20 (9.1%) samples had discordant subtypes between *gag* and *env*, 6 of 44 (13.9%) in Senegal, 4 of 55 (7.2%) in Cameroon, 1 of 38 (2.6%) in Djibouti, and 10 of 82 (12.1%) in Gabon. Subtypes A and G were predominantly involved in the recombination events. Phylogenetic tree analysis of *gag* showed that an important number of the A sequences form a distinct subcluster with the AG-IBNG prototype strain (a complex A/G mosaic virus): 27 of 32 (84.3%) in Senegal, 12 of 17 (70.6%) in Nigeria, 24 of 39 (61.5%) in Cameroon, and 38 of 70 (54.3%) in Gabon. Full-length genome analysis of 3 and additional sequences in *pol* for 10 such strains confirmed that they have a similar complex A/G mosaic genomic structure. These data suggest that in West Africa, most probably between 60% and 84% of the subtype A viruses are recombinant AG-IBNG viruses. This finding has potential implications on future vaccine, diagnostic, and treatment strategies. The actual and future role of these viruses in the global pandemic must be monitored in all new molecular epidemiologic studies, a discrimination between subtype A and AG-IBNG–like viruses is necessary. **Key Words:** HIV-1 subtypes—*env*—*gag*—Recombination—AG-IBNG—Africa.

Phylogenetic analysis of HIV-1 viruses from different geographic locations has revealed that HIV-1 can be divided into at least three distinctive groups, M, N, and O (1). Group M, which is the major group, comprises most HIV-1 strains that cause AIDS worldwide. Full-length genome sequencing showed that group M HIV-1 viruses can be further subdivided into at least nine different, nonrecombinant, subtypes (A, B, C, D, F, G, H, J, and K) approximately equidistantly related with intrasubtype divergence up to 20% and intersubtype divergence between 25% to 35%, for the *Env* amino acid sequences (2–4). HIV-1 group O viruses have only 50% similarity with the HIV-1 group M isolates in the *env* gene (5,6). There are also many mosaic genomes of HIV-1, some of which are unique or restricted to one transmission cluster, and others that are major circulating forms (3). It was proposed to designate those recombinant viruses as

"circulating recombinant forms (CRF)," by associating CRF with the name of the first full-length sequence of that form. There are currently 4 CRFs of HIV-1: AE-CM240 from southeast Asia, the AG-IBNG from Africa, the AGI-CY032 from Cyprus and Greece, and AB-KAL153-like viruses found in the epidemic among intravenous drugs users (IVDUs) in Kaliningrad in Russia (3).

The various HIV-1 groups and subtypes differ in their geographic dissemination, and so the subtype designations have been critical molecular epidemiologic markers for tracking the course of the global pandemic. HIV-1 group O seems to be endemic in Central West Africa, especially Cameroon, where the frequency of infection is estimated to be 2% and 5% of HIV-1–infected individuals (7–9). Group N viruses were recently reported from only 2 Cameroonian patients (1). Among HIV-1 group M viruses, geographic distribution of the different subtypes is also heterogeneous. In Europe and in North and South America, subtype B is largely dominant, whereas in South and East Africa and in India subtype C is spreading rapidly (10–12). In Southeast Asia, the circulating recombinant AE-CM240 is predominant (12,13). In Africa, all HIV-1 groups and HIV-1 group M subtypes have been documented; however, their distribution on the African continent is not homogeneous (14). Subtype A is predominant in West and Central Africa whereas subtype C is more common in East and South Africa. Subtype D is present in Central and East Africa with prevalences ranging from 5% to more than 40% of the circulating strains (15,16). Subtype G has been documented in many West and Central African countries, whereas subtype H was only found in Central Africa (2,15). In certain populations and regions where multiple HIV-1 subtypes co-circulate, many combinations of intersubtype recombinant viruses have been documented (A/C, A/D, B/F, A/G/I/J) (17–20), and even an intergroup M/O recombinant virus has been recently isolated from a Cameroonian patient (21). Recombinant viruses may have some advantages over the parental strain, including eventual modifications in tropism and replication efficiency (fitness). Therefore, recombination may play a significant role in global HIV evolution, creating novel viral genotypes within populations.

Recombination can be missed if large portions or different regions of the viral genome are not examined. The AG-IBNG recombinant virus has been documented in several African countries (22); however, the real prevalence of this virus in the global epidemic is still unknown. This virus has a complex mosaic genomic structure involving subtype A and G sequences with multiple breakpoints, but given that both *gag* and *env* derive largely from subtype A, these viruses were initially classified as subtype A (23,24). However, they form a significant subcluster within subtype A in the *gag* and *env* regions and can therefore be recognized even by using partial sequencing of familiar regions (22).

In this study, we tried to identify the proportion of recombinant HIV-1 strains by identifying the genetic subtype in *gag* and *env* of more than 200 HIV-1 strains collected in different geographic regions in Africa. We also estimated the proportion of AG-IBNG–like viruses among *gag* or *env* subtype A viruses.

MATERIALS AND METHODS

Study Samples

In total, 219 randomly chosen HIV-1–positive samples of African origin were tested, 55 from patients living in Cameroon, 44 from Senegal, 38 from Djibouti, and 82 from Gabon. All samples from Senegal have been collected among patients seen at one of three major hospitals in Dakar between 1997 and 1998. All Cameroonian samples were collected between 1995 and 1997 from patients visiting the military hospital in Yaounde. The samples from Djibouti were all from citizens of metropolitan France, as previously described (25), who had become infected with HIV-1 after overseas deployment in this country. Most remained asymptomatic but had been infected between 1987 and 1997. The samples from Gabon have been collected in 1997 among different population groups.

DNA was extracted from the primary or short-term cultured peripheral blood mononuclear cells (PBMCs) using Qiagen (Qiagen S.A., Courtabeauf, France) DNA isolation kit.

Genetic Subtyping in the *gag* Region

A 700-bp fragment, corresponding to the p24 region from the *gag* gene, was amplified with previously described primers, G00-G01 and G60-G25 (26). The conditions for polymerase chain reaction (PCR) were as follows: a first denaturation step for 3 minutes at 92°C, followed by 30 cycles of 92°C for 10 seconds, 55°C for 30 seconds, and 1 minute at 72°C, with a final extension for 7 minutes at 72°C, in a final volume of 50 μl. The reaction mixture consists of 50-mmol/l KCl, 10-mmol/l Tris-HCl (pH 9), 0.1% Triton X-100, 1.4-mmol/l $MgCl_2$, 10-pmol each primer, 0.2-mmol/l each dNTP, and 2.5-U Taq polymerase. Moreover, 1 μl of this amplified product was used for the second round using the same reaction mixture and PCR conditions for 40 cycles, in a final volume of 100 μl. The PCR amplification products were detected by electrophoresis on a 1% agarose gel and visualized by ethidium bromide staining.

Nucleotide sequences were obtained by direct sequencing of the PCR products. The amplified DNA was purified using a QiaQuik gel extraction kit (Qiagen S.A., France). Cycle sequencing was performed using fluorescent dye terminator technology (dye terminator cycle sequencing with AmpliTaq DNA polymerase FS (PE Biosystem) according to the manufacturer's instructions. Electrophoresis and data collection were done on an Applied Biosystems 373A automatic DNA sequencer (stretch model, PE Biosystem, Courtaboeuf, France).

TABLE 1. *Distribution of HIV-1 genetic subtypes in gag (p24) and env (V3-V5) genes among samples collected in various African countries*

| Country | Gene | Genetic subtype | | | | | | | | Discordant samples n (%) |
		A n (%)	B n (%)	C n (%)	D n (%)	E n (%)	F n (%)	G n (%)	H n (%)	
Senegal (n = 44)	gag	32 (72.7)	—	5 (11.3)	4 (9.1)	—	—	2 (4.5)	1 (2.2)	5 (11.4)
	env	33 (75.0)	—	6 (13.6)	4 (9.1)	—	—	1 (2.2)	—	
Cameroon (n = 55)	gag	39 (70.9)	—	1 (1.8)	2 (3.6)	—	9 (16.3)	4 (7.2)	—	4 (7.2)
	env	39 (70.9)	—	1 (1.8)	2 (3.6)	—	11 (20)	2 (3.6)	—	
Djibouti (n = 38)	gag	6 (15.7)	4 (10.5)	27 (71)	1 (2.6)	—	—	—	—	1 (2.6)
	env	5 (13.1)	4 (10.5)	27 (71)	1 (2.6)	1 (2.6)	—	—	—	
Gabon (n = 82)	gag	70 (85.3)	—	—	5 (6.1)	—	1 (1.2)	4 (4.8)	2 (2.4)	10 (12.1%)
	env	69 (84.1)	—	5 (6.1)	1 (1.2)	—	1 (1.2)	6 (7.3)	—	
Total (n = 219)	gag	147 (67.1)	4 (1.8)	33 (15.1)	12 (5.5)	—	10 (4.6)	10 (4.6)	3 (1.4)	20 (9.1)
	env	147 (67.1)	4 (1.8)	39 (17.8)	8 (3.7)	1 (0.05)	12 (5.5)	9 (4.1)	—	

Genetic Subtyping in the *env* Region

The genetic subtype in the envelope was determined by heteroduplex mobility assay (HMA) on the V3-V5 region of the envelope (700 bp) as described by Delwart et al. (27) and/or by sequencing of the corresponding fragment. The V3-V5 region was amplified by nested PCR using the previously described primers ED5 through ED12 as outers and ES7 and ES8 as inners. The genetic subtype for samples that remained indeterminate using HMA were determined by sequencing followed by phylogenetic analysis of the corresponding region in the envelope. The amplified fragments were purified with the Qiaquick gel extraction kit (Qiagen) and then directly sequenced on an automatic sequencer (Applied Biosystems 373A stretch model, PE Biosystems) using the ABIPRISM Big Dye Terminator cycle sequencing technology.

Near-Full-Length Genome Sequencing

Three samples clustering with the AG-IBNG virus, in *gag* and *env*, were selected to sequence the near-full-length genome, denoted as 98SE-MP1211 and 98SE-MP1213 from Senegal and 97CA-MP807 from Cameroon. For the samples from Senegal, sequences were done on DNA from primary PBMCs. The near-full-length genomes (covering long-terminal repeat LTR5' to LTR3') were first amplified with LPBS (AAAATCTCTAGCAGTGGCGCCCGAACAGGGAC) and LsiG1 (TCAAGGCAAGCTTTATT*GAG*GCTTAAGCAG) primers, followed by a second round of two overlapping PCRs (overlap of 1500 bp), a 5400-bp fragment covering *gag* to Vpu with G00 (GACTAGC-GGAGGCTAGAG) and Mvpu3 (TACTATRGTCCACACAACTAT-KGCT) and a 5000-bp fragment covering 3' pol to LTR with hpol4235 (CCCTACAATCCCCAAAGTCAAGG) and LsiG1. For the Cameroonian sample, the sequence of the genome was done on DNA from short-term–cultured PBMCs. Provirus was amplified by a series of overlapping nested PCRs amplifying fragments of 2800 bp with an approximate 700-bp overlap.

We used the Boehringer Long Template Expand DNA polymerase (Boehringer Mannheim, Laval, Qc, Canada), according to the instructions of the manufacturer. The overlapping PCR fragments from each patient were either cloned and sequenced or the PCR product was directly sequenced. Sequencing was performed by using cycle sequencing and dye terminator technology on an automated DNA sequencer (ABI 373A stretch, Applied Biosystems). The sequenced fragments were reanalyzed and assembled into contiguous sequences by using the Seqed program (Applied Biosystems).

Sequence Analysis of the *pol* Gene

The protease and RT genes were amplified with seminested PCR using G25rev (GCAAGAGTTTTGGCTGAAGCAATGAG) and IN3 (TCTATVCCATCTAAAAATAGTACTTTCCTGATTCC) as outer primers and AV150 (GTGGAAAGGAAGGACACCAAATGAAAG) and IN3 as inner primers. The PCR parameters for the two rounds were as follows: a first denaturation step for 5 minutes at 92°C, followed by 40 cycles of 92°C for 20 seconds, 50°C for 30 seconds, and 72°C for 2 minutes. PCR reactions were done with the rTth DNA Polymerase from Perkin Elmer (Perkin Elmer-Cetus, Norwalk, CT, U.S.A.) according to the instructions of the manufacturer. The amplified fragments were purified with the Qiaquick gel extraction kit (Qiagen) and then directly sequenced on an automatic sequencer (Applied Biosystems 373A stretch model, PE Biosystems) using the ABIPRISM Big Dye Terminator Cycle sequencing technology.

Phylogenetic Analysis

Nucleotide sequences were aligned using CLUSTAL W (28) with minor manual adjustments, bearing in mind the protein sequences. Regions that could not be aligned unambiguously, due to sequence variability, were omitted from analysis. Phylogenetic trees using the neighbor-joining method and reliability of the branching orders using the bootstrap approach were implemented using CLUSTAL W. Genetic distances were calculated with the Kimura's two-parameter method, (ratio T:t = 2.0) (29).

TABLE 2. *The different discordant gag/env subtype profiles observed in this study and the number of samples seen in each of the countries studied*

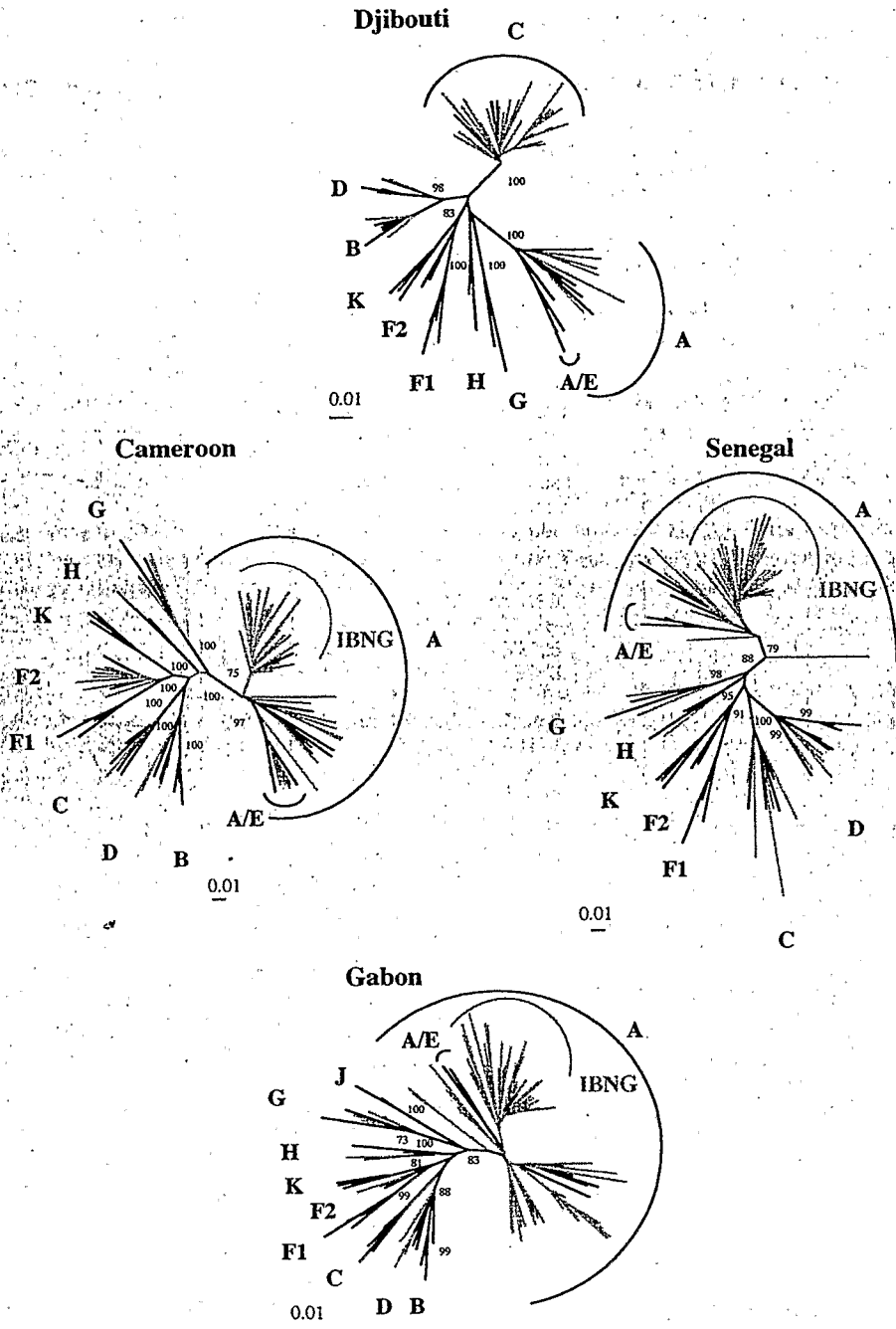| Subtype (gag–env) | Country | | | | Total |
	Senegal	Cameroon	Djibouti	Gabon	
A–G	1	—	—	1	2
A–C	1	—	—	3	4
A–E	—	—	1	—	1
A–F	—	2	—	—	2
D–A	—	—	—	2	2
D–C	—	—	—	2	2
G–A	2	2	—	—	4
H–A	1	—	—	1	2
H–G	—	—	—	1	1

FIG. 1. Phylogenetic tree based on 612 unambiguously aligned nucleotides from the p24 *gag* region of the new HIV-1 isolates (gray) and reference strains (black) representing the different genetic subtypes. The analysis was performed as described in Materials and Methods. The reference sequences were A-U455, A-K89, A-92UG037, A-VI32, B-RF, B-HXB2, B-OYI, B-JRFL, C-92BR025, C-ETH2220, C-ZAM18, C-UG268, D-NDK, D-Z2Z6, D-ELI, D-94UG114, E-90CR402, E-93TH253, E-CM240, F1-93BR020, F1-VI174, F1-VI69, F1-BZ162, F2-MP255, F2-MP257, K-EQTB11, K-MP535, K-VI325, G-92NG003, G-92NG083, H-90CF056, H-VI557, and AG-IBNG. F1 and F2 correspond to representatives of subclades forming each separate clusters within subtype F (4). Bootstrap values <70% are not shown.

Intersubtype Recombinant Analysis

To analyze whether the viruses were recombinant in the sequenced regions, several additional analyses were performed. Diversity plots, using the DIVERT program available online (accessed April 27, 2000; available at http://193.50.234.246/~beaudoin/anrs/Diversity.html), determined the percentage of diversity between selected pairs of sequences by moving a window of 300 bp along the genome alignment in 20-bp increments. Divergence values for each pairwise comparison were plotted at the midpoint of the 300-bp segment.

Simplot 2.5 software (30), kindly provided by Stuart Ray, was used to calculate bootstrap plots. For the bootstrap plots, the Simplot software performed bootscanning on neighbor-joining trees by using SEQBOOT, DNADIST (with Kimura's parameter method and a transition:transversion ratio of 2.0), NEIGHBOR, and CONSENSUS from the Phylip package for a 500-bp window moving along the alignment in increments of 50 bp (this approach is similar to the bootscanning method described by Salminen et al. [31]). We evaluated 100 replicates for each phylogeny. Bootstrap values for the studied sequences were plotted at the midpoint of each window.

TABLE 3. *The number and percentages of the subtype A samples clustering with the AG-IBNG prototype strain in gag (p24), and the genetic distances observed in the different subtype A clusters*

Country	Subtype A		Genetic distances (mean and extreme values)		
	IBNG n (%)	non-IBNG n (%)	IBNG	non-IBNG	IBNG + non-IBNG
Senegal	27/32 (84.3)	5/32 (15.7)	4.3 (1.7–7.7)	7.5 (3.9–10.9)	6.2 (1.7–11.5)
Cameroon	24/39 (61.5)	15/39 (38.4)	4.8 (2.1–7.7)	8.3 (2.9–10.6)	7.3 (1.7–11.2)
Djibouti	—	6/6 (100)	—	7.2 (1.8–9.6)	
Gabon	38/70 (54.3)	32/70 (55.7)	5.0 (1.6–9.9)	7.5 (1.0–11.7)	7.2 (1.0–13.5)

Other subtypes: 5.4 (0.9–10.6).
Genetic distances were calculated with the Kimura two parameter method, using the same alignment as used for the construction of the phylogenetic trees in Figure 1.

Genbank Accession Numbers

The accession numbers for the *gag* sequences are as follows: AJ286366 to AJ286420 for Cameroon; AJ286421 to AJ286500, AJ286539, and AJ286540 for Gabon; and AJ286501 to AJ286538 for France (contaminated with HIV-1 in Djibouti). The *gag* sequences from Senegal are available under the following numbers: AJ274541 to AJ274545, AJ274547, AJ274548, AJ274550 to AJ274552, AJ274555

to AJ274557, AJ274559, AJ274560, AJ274563, AJ274565 to AJ274567, AJ274569 to AJ274574, AJ274576 to AJ274579, AJ274581, AJ274583, AJ274586, AJ274590, AJ274592, AJ274596, AJ274600, AJ274604, AJ274612, AJ274613, AJ274615, AJ286541, and AJ286542.
The *pol* sequences are available in Genbank with the accession numbers AJ286134 to AJ286143. The near-full-length genome sequence accession numbers are: AJ286133 for 97CM-807; AJ251056 for 98SE-MP1211, and AJ251057 for 98SE-MP1213.
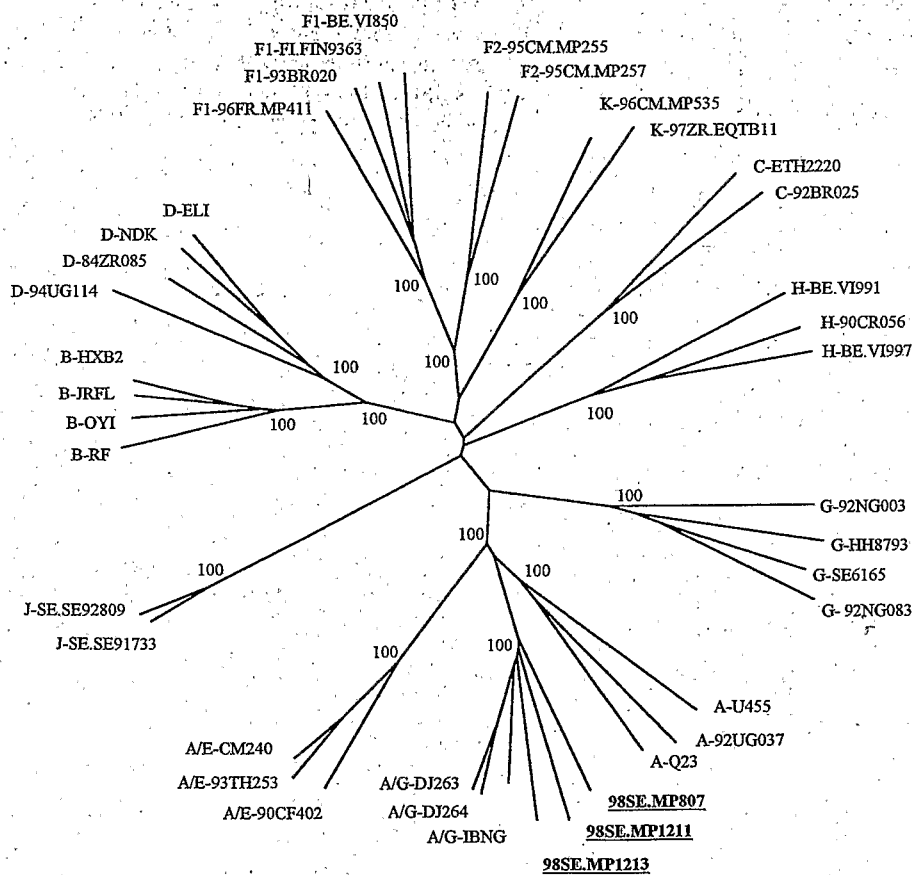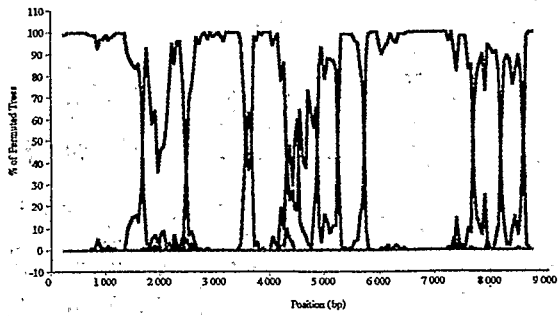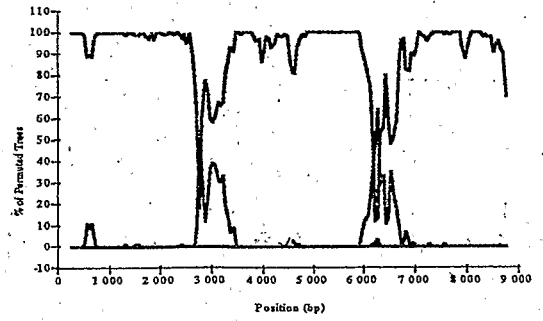

FIG. 2. Phylogenetic tree of the near-full-length sequences of the three new HIV-1 viruses with reference strains from each of the known subtypes. The analysis was performed as described in Materials and Methods and using the same references sequences (except K-VI325) as for the *gag* trees in Figure 1. Bootstrap values <70% are not shown.
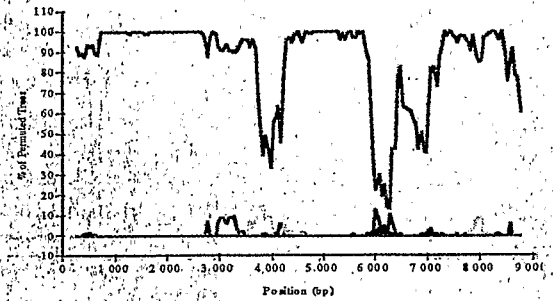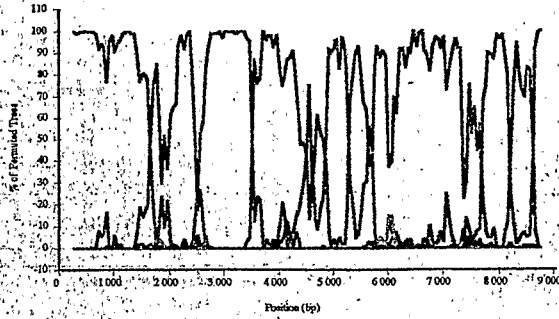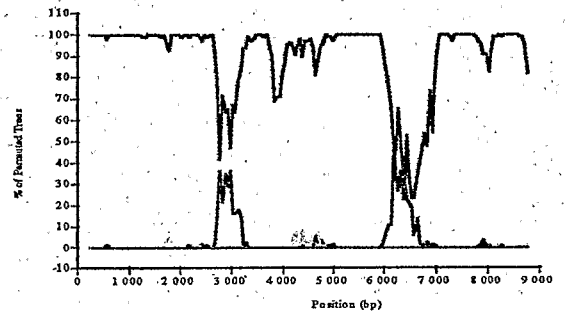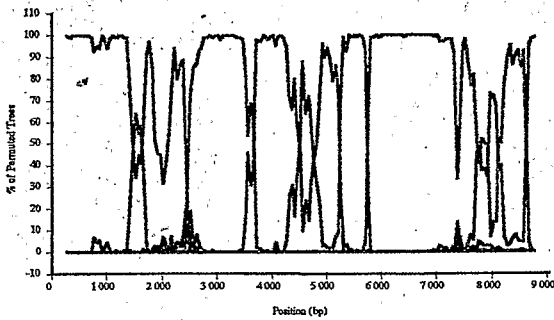
A/G-IBNG

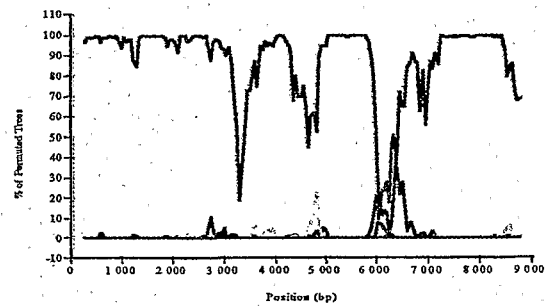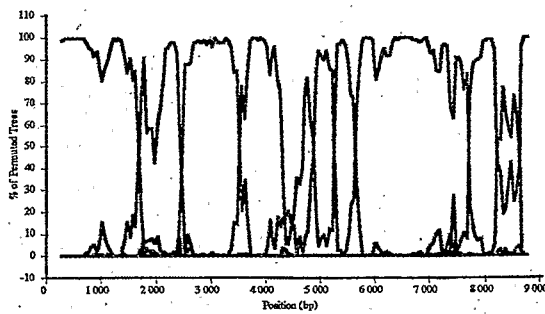## A/G-DJ263

98SE.MP1211

98SE.MP1213

98CM.MP807

—— A
—— B
—— C

—— D
—— F
—— G

—— H
—— J
—— AG-IBNG

RESULTS

Genetic Subtypes in the *env* and *gag* Genes

Table 1 shows the genetic subtypes identified in the V3 to V5 region of the envelope by HMA and the numbers and percentages of the genetic subtypes identified in the p24 region of the *gag* gene by sequencing followed by phylogenetic tree analysis, for the corresponding strains in the different countries. In the West and Central West African countries subtype A was predominant in *gag* and *env*, whereas in the East African country, Djibouti, subtype C was the classification of most samples. As shown in Table 1, the subtype distribution is slightly different between the envelope and the *gag* regions. Only small differences were seen for the percentages of A, C, D, E, F, G, and H subtype samples in the two different genomic regions.

Among the 44 samples from Senegal, 5 (11.4%) had discordant subtype designations between envelope and *gag*. In Cameroon, 4 (7.2%) of 55 samples, and in Gabon 10 (12.9%) of 82 samples were discordant. In Djibouti, only 1 discordant sample was seen and this corresponded to an *env* subtype E sample with a *gag* subtype A sequence, as commonly documented among circulating recombinant AE strains.

Table 2 shows the different discordant subtype profiles observed between *gag* and *env* in the different countries. Overall, among the total 20 discordant samples, 9 different profiles were seen and in 7 of them subtype A sequences, either in *gag* or *env*, were involved. In only 3 of 20 discordant samples, no A sequences were involved and among the other 17 samples, 9 were A in *gag* and 8 were A in the envelope.

Phylogenetic Analysis of the *gag* (p24) Sequences

Figure 1 shows the phylogenetic tree analysis of the *gag* sequences for the different geographic regions. These phylogenetic trees show that among the subtype A sequences, distinct subclusters can be identified, more specifically, an important number of the *gag* A sequences form a distinct subcluster with the AG-IBNG prototype strain (a complex A/G mosaic virus [22]). These AG-IBNG clusters are supported with high bootstrap values of 97% in Cameroon and only 45% and 47%

in Gabon and Senegal, respectively. None of the subtype A samples from Djibouti clustered with the AG-IBNG virus.

Table 3 summarizes the number of subtype A samples studied that may form a separate cluster with the AG-IBNG prototype strain. Overall, 89 of 147 (60.5%) subtype A *gag* sequences are close to AG-IBNG in the phylogenetic tree. In Senegal, 27 of 32 (84.3%) of the *env* subtype A viruses are IBNG-like in *gag*, whereas in Cameroon and Gabon, 24 of 39 (61.5%) and 38 of 70 (54.3%), respectively, of the subtype A strains cluster with AG-IBNG.

Genetic distances have been calculated for the subtype A strains and the subclusters observed within subtype A and the other subtypes (B–H) in the various countries. Data are summarized in Table 3. Genetic distances observed for the AG-IBNG cluster are significantly lower than those calculated for the other subtype A strains: 4.3% in Senegal, 4.8% in Cameroon, and 5% in Gabon versus 7.5%, 8.3%, and 7.2%, respectively, for the non–IBNG-like subtype A strains. The average genetic distances within the other subtypes, B to H, is 5.4% and are generally lower than for subtype A. This was confirmed in all three West African countries studied.
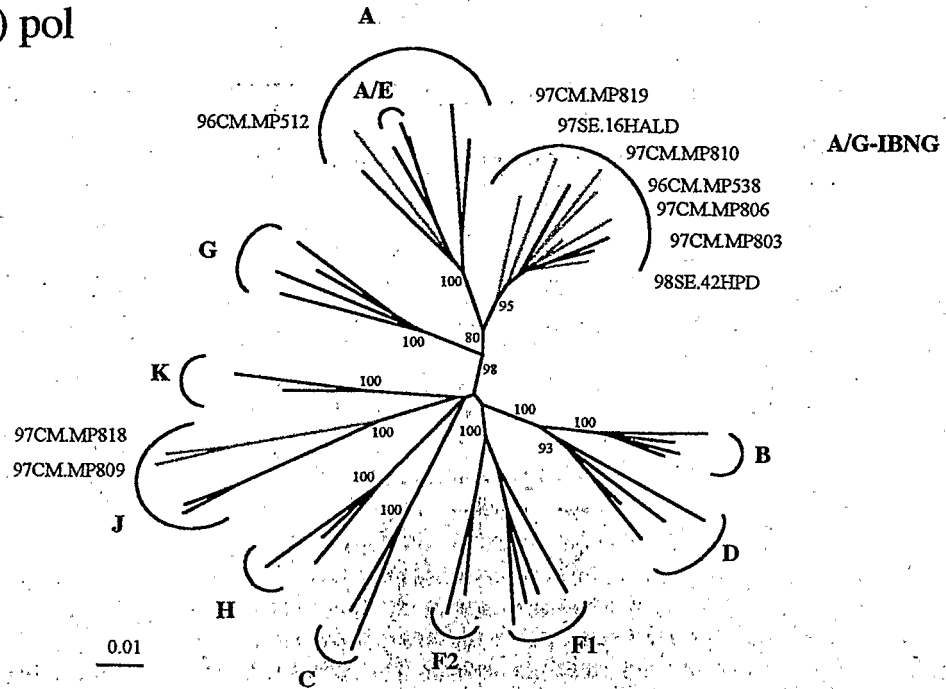
Sequence Analysis of Near-Full-Length Genomes of *gag* AG-IBNG–Like Viruses

To study whether *gag* subtype A viruses that cluster with the AG-IBNG prototype strain virus have a similar complex mosaic genome, the near-full-length genome was sequenced for 3 subtype A viruses clustering with this strain in *gag*: 2 from Senegal (98SE-MP1211 and 98SE-MP1213) and 1 from Cameroon (97CA-MP807). We obtained 8851 bp for MP1211, 8858 bp for MP1213, and 8800 bp for MP807. The sequences start between the first and the tenth amino acid from the *gag* polyprotein, LTR extremities were not sequenced. All other genes were complete without premature stop codons and no major deletion or insertion events were present in the sequences of these strains.
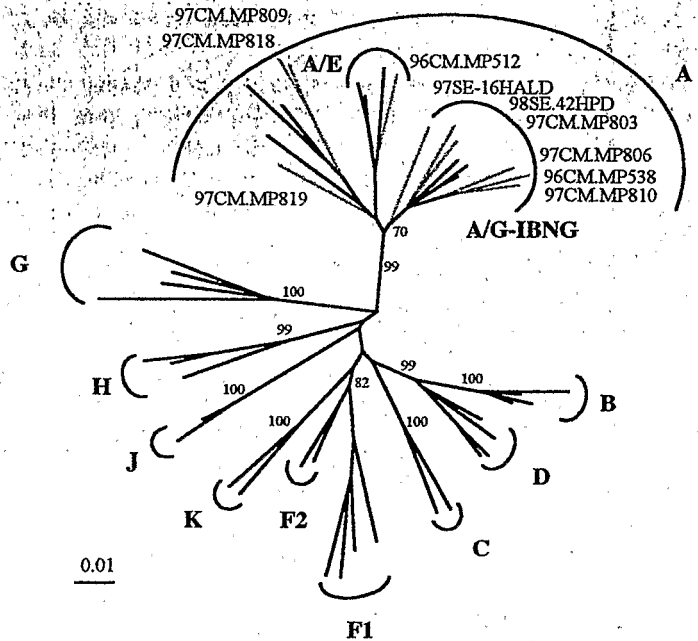
To determine the phylogenetic relationships of the newly characterized viruses, we constructed evolutionary trees from the near-full-length genome. The phylogenetic tree shows that these 3 new strains form a separate clus-

---

FIG. 3. Bootscan plots of the three new HIV-1 near-full-length sequences and the HIV-1$_{\text{IBNG}}$ prototype strain. The new sequences were bootscanned against consensus sequences (calculated on the same sequences used in the phylogenetic tree of the near-full-length sequences in Fig. 2) of the nonrecombinant subtypes A, B, C, D, F, G, H, and J (a, c, e, and g), and against the nonrecombinant subtypes and the AG-IBNG prototype virus (b, d, f, and h). The genome was scanned in segments from 500 bp with an overlap of 50 bp. The x-axis indicates the nucleotide position along the alignment; gaps were removed from the alignment. The y-axis shows bootstrap values (%).
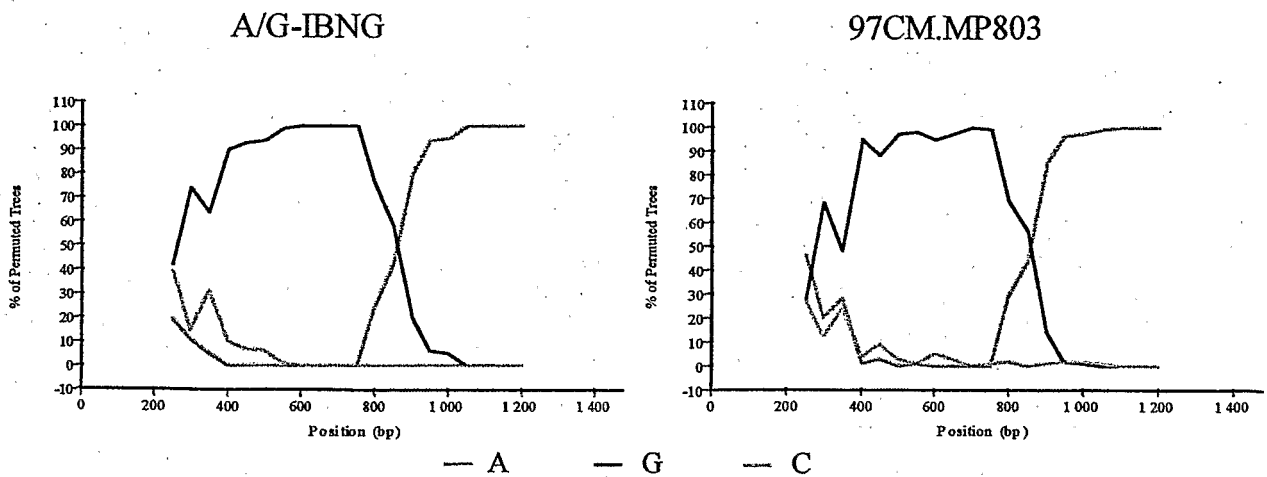
a) pol

A

A/E

96CM.MP512

97CM.MP819
97SE.16HALD
97CM.MP810
96CM.MP538
97CM.MP806
97CM.MP803
98SE.42HPD

A/G-IBNG

G

K

97CM.MP818
97CM.MP809

J

H

C

F2

F1

D

B

100 95 80 98 100 100 100 100 100 100 100 93

0.01

b) gag

97CM.MP809
97CM.MP818

A/E

96CM.MP512

97SE-16HALD
98SE.42HPD
97CM.MP803
97CM.MP806
96CM.MP538
97CM.MP810

A

97CM.MP819

A/G-IBNG

G

H

J

K

F2

F1

C

D

B

70 99 100 99 99 100 100 82 100 100 100

0.01

c)

A/G-IBNG

97CM.MP803

— A   — G   — C

ter with the AG-IBNG virus and with the previously described AG-IBNG–like viruses, DJ263 and DJ264 from Djibouti supported by high bootstrap values (100%; Fig. 2).

To characterize more precisely whether the 98SE-MP1211, 98SE-MP1213 and 97CA-MP807 viruses have the same mosaic structure as the AG-IBNG prototype virus, we performed pairwise sequence comparisons between the near-full-length genome sequence of these isolates and with those available from the database, using the same alignment of nucleotide sequences used for the phylogenetic tree construction. Importantly, distance values were calculated only after removing all sites with a gap in any of the sequences, ensuring that all comparisons were made across the same sites. When the new sequences were plotted against references of all nonrecombinant subtypes, we observed for all 3 sequences similar disproportionate levels of sequence divergence with subtype A and G along their genome, consistent with the pattern observed when the AG-IBNG prototype virus is plotted against the same reference strains (data not shown).

To evaluate systematically the bootstrap values supporting independent monophyletic phylogeny in different parts of the genome, the alignment was bootscanned in regard to the clustering of the putative AG-IBNG–like viruses with a sliding window of 500 bp advancing with 50-bp increments. The same multiple genome alignment generated for the phylogenetic analysis of the near-full-length genomes shown in Figure 2 was used to calculate the consensus reference sequences (50% treshhold) for the 8 groups of nonrecombinant subtypes corresponding to subtypes A, B, C, D, F, G, H, and J. The magnitudes of the bootstrap values supporting the clustering of our strains with consensus sequences from members of each subtype was performed. The results of the bootscan analysis are shown in Figure 3. The complexity of these new strains is readily apparent with alternating segments clustering with subtype A and G, similar to the complexity observed for the AG-IBNG prototype strain. When the 98SE-MP1211, 98SE-MP1213, and 97CA-MP807 strains were bootscanned against an alignment that includes the AG-IBNG strain, bootscanning results of the entire genomes revealed that they all were similar to the

AG-IBNG isolate. The lower bootstrap results found in certain regions of the genome result from the simultaneous presence in the analysis of nonrecombinant subtype A sequences and the A sequences from the AG-IBNG strain.

Sequence Analysis of the *pol* Region from *gag* Subtype A Viruses Including AG-IBNG– Like Viruses

The AG-IBNG HIV-1 virus is mosaic in the *pol* region. We therefore sequenced the *pol* region, covering the protease and reverse transcriptase (RT) genes for 10 subtype A samples in *gag*, among them 6 that had formed a separate cluster with AG-IBNG. Figure 4 shows the phylogenetic tree analysis of the 11 samples from the *gag* and *pol* sequences. Of the 6 samples clustering with AG-IBNG in *gag*, all formed a well-separated cluster with this strain in the *pol* region. For the other 4 subtype A strains, 1 was subtype A in *pol*, 1 clustered with AG-IBNG, and 2 were subtype J in *pol*. Diversity plotting and bootscanning on the *pol* sequences of the AG-IBNG–like sequences showed that all these sequences have a similar mosaic structure in *pol* as the AG-IBNG prototype strain. In Figure 4C, bootscan analysis in the *pol* gene is shown for one isolate from the study (97CM-MP803), as well for the AG-IBNG strain in the same region.

Estimated Prevalence of Recombinant HIV-1 Viruses and AG-IBNG–Like Viruses

Full-length genome analysis showed that the 3 HIV-1 viruses clustering in *gag* (p24) with the prototype AG-IBNG virus have a similar complex A/G mosaic genomic structure. Additional sequence analysis of the *pol* region, showed that 6 viruses clustering in *gag* and *pol* with AG-IBNG had a similar A/G mosaic structure in this gene. These data allow us to extrapolate that in West and Central West Africa, most probably between 60% and 84% of the subtype A viruses are AG-IBNG–like viruses. Moreover, some of the subtype A *gag* sequences involved in recombination (different *env* and *gag* subtypes) cluster with AG-IBNG in the *gag* region. More

---

FIG. 4. Phylogenetic tree based on 612 unambiguously aligned nucleotides from **(A)** the p24 *gag* region and 1536 nucleotides from **(B)** the *pol* region (protease and reverse transcriptase [RT] genes) of 10 HIV-1 isolates and reference strains representing the different genetic subtypes. The analysis was performed as described in Materials and Methods using the same reference sequences as for the trees in Figures 1 and 2. Bootstrap values <70% are not shown. **(C)** Bootscan plots of the one new HIV-1 *pol* sequence from an AG-IBNG–like strain (97CM-MP810) and the AG-IBNG prototype strain. The sequences were bootscanned against consensus sequences (calculated on the same sequences used in the phylogenetic tree of the sequences in Fig. 4A) of the nonrecombinant subyptes A, C, and G. The genome was scanned in segments from 300 bp with an overlap of 20 bp. The x-axis indicates the nucleotide position along the alignment; gaps were removed from the alignment. The y-axis shows the bootstrap values (%).

TABLE 4. *Estimation of the numbers and percentages of recombinant HIV-1 viruses circulating in West Africa*

Country	n tested	Discordant gag/env subtype n (%)	AG-IBNG n (%)	Total n (%)
Senegal	44	5 (13.9%)	27 (61.4%)	33 (75.0%)
Cameroon	55	4 (7.2%)	24 (43.6%)	28 (50.9%)
Gabon	82	10 (12.1%)	38 (46.3%)	48 (58.3%)
Total	181	19 (10.5%)	89 (49.2%)	119 (65.7%)

specifically, 3 of 9 subtype A strains in *gag* and involved in recombination between *gag* and *env* cluster with AG-INBG in the *gag* region, 1 of 4 A/C, and the 2 A/F viruses. This confirms that the AG-IBNG viruses circulate and are implicated in a very large part of global epidemic in this region and that they are also involved in recombination events.

Addition of the AG-IBNG viruses to the numbers of viruses for which discordant *gag* and envelope subtypes were observed make the overall minimal percentages of recombinant virus involved in the HIV-1 epidemic in Africa very high (Table 4): 75.0% in Senegal (6 discordant with 27 AG-IBNG), 51.0% in Cameroon (4 discordant with 24 AG-IBNG), and 54.7% in Gabon (10 discordant with 38 AG-IBNG).

DISCUSSION

Until recently, based on partial sequences of different genes or full-length genome sequencing, it was estimated that up to 10% of the HIV-1 strains in the global pandemic are intersubtype recombinants (32–34). All these recombinant viruses originated from geographic regions where multiple subtypes cocirculated and most probably result from coinfections with divergent viruses. Based on the discordant subtype results between partial *gag* and *env* sequences, 20 (9.1%) of the 219 strains in our study were potential recombinant viruses, with percentages ranging from 2.6% in Djibouti to 12.1% in Gabon. These data correspond to the global estimations on recombinant viruses that are circulating.

However, in some regions the number of recombinant viruses that circulate may even be higher. In a previous study on the genetic subtype distribution in Nigeria, we documented 12 (32.4%) of 37 samples from that country with discordant *gag* and *env* subtypes (35). These high values were also reported in another study on Nigerian samples (36).

In 1994, AG-IBNG, a new subtype A isolate from Ibadan, Nigeria, was reported on the basis of the gp120 sequence (23). Based on *gag* and *env* sequences, it was

later described as a distinct subcluster within A subtype (24). With the recent availability of full-length subtype A and G sequences, it is clear that this strain is a complex mosaic virus with alternating subtype A and subtype G sequences (22). Geographic dispersion of the AG-IBNG subtype was already suggested by the presence of this isolate in Nigeria (in West Africa), and two viruses with a similar mosaic genome from Djibouti (DJ263 and DJ264) in East Africa (22). Additional full-length genome sequencing in our study from strains obtained in 2 other African countries (two from Senegal and two from Cameroon) confirm the geographic spread of this AG-IBNG–like viruses. From our data and the previously described data on the Djibouti strains, we can extrapolate with a relative high probability that the viruses clustering with AG-IBNG in *gag* are predominantly AG-IBNG–like viruses. Our study shows on a large number of samples that in West and Central West Africa, AG-IBNG viruses comprise most of the *env* and *gag* subtype A strains currently circulating in these countries; 84.0% in Senegal, 61.0% in Cameroon, and 54.3% in Gabon. In a previous study on Nigerian samples, we have shown that most subtype A viruses from this country clustered with AG-IBNG: 71% based on p24 *gag* sequences and 64% based on *env* sequences (35). Genetic distances observed among the different AG-IBNG viruses indicate that they have been spreading for some time. The high prevalences of these particular A/G recombinant in Senegal, Gabon, Cameroon, and Nigeria are an additional factor in considering these viruses as a separate and distinct group in the global epidemic.

Virtually any isolate, including recombinant viruses, may compete more efficiently in a given country than the original genotype if it acquired some selective advantage. However, it remains striking to see how efficiently recombinant viruses have spread among different population groups in different geographic regions (suggesting that they could have a better viral fit than the parental nonrecombinant strains). In Thailand, where subtypes E (an A/E recombinant) and B were initially introduced, subtype E became predominant (13,37). In China, subtype B and C viruses have been introduced, and a recombinant B/C virus is spreading now through various parts of the country. Several reports documented the introduction of subtypes A and B among intravenous drug users in Kalingrad, Russia, and a recombinant A/B virus is rapidly spreading there (39,40). In West Africa, subtype A is predominant with prevalences ranging from 70% to >85% of the strains circulating (15) and, as shown in this study, AG-IBNG viruses comprise most envelope subtype A viruses in different countries.

Recombinant viruses are already contributing substan-

tially to the global pandemic, and the likelihood of generating recombinant viruses will only continue to increase as the different HIV-1 subtypes spread to all continents (25,41,42). In a previous study, we documented that even the complex AG-IBNG–like viruses are implicated in an intergroup (M/O) recombination event (21) and several viruses with discordant *gag* and *env* subtypes had *gag* sequences that clustered with the AG-IBNG virus. It becomes more and more evident that the genetic subtype distribution is a dynamic and unpredictable process.

Addition of the AG-IBNG–like viruses to the number of viruses for which discordant *gag* and *env* subtypes were observed made the overall minimal percentages of recombinant viruses in West and Central West Africa high at 51% to 75%. Moreover, the predominance may even be higher in that we had considered only recombination between *gag* and *env*, and not all parts of the genome. We should also indicate the difference between recent recombinants and older recombinants such as the IBNG-like viruses.

Further studies are required to find out which factors (epidemiologic, demographic, biologic) are responsible for the predominance of AG-IBNG viruses in this part of Africa. Our finding that AG-IBNG–like viruses predominate in this part of the world can have implications for future vaccines, as well as diagnostic, and treatment strategies. Especially with the recent introduction of antiretroviral treatment in Africa, one must be aware that all the AG-IBNG–like viruses are subtype G in the protease gene, and recently a naturally occurring lower susceptibility for HIV-1 subtype to protease inhibitors has been shown (43). The current and future role of these viruses in the global pandemic must be monitored in all new molecular epidemiologic studies and discrimination between subtype A and AG-IBNG–like viruses is necessary.

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. Janssens W, Buve A, Nkengasong JN. The puzzle of HIV-1 subtypes in Africa. *AIDS* 1997;11:705–12.
3. Carr J, Foley B, Leitner T, Salminen M, Korber B, McCutchan F. Reference sequences representing the principal genetic diversity of HIV-1 in the pandemic. In: *Human retrovirus and AIDS*, pt 3. Los Alamos NM: Los Alamos National Laboratory, 1998.
4. Triques K, Bourgeois A, Vidal N, et al. Near-full-length genome sequencing of divergent African HIV-1 subtype F viruses leads to the identification of a new HIV-1 subtype designated K. *AIDS Res Hum Retroviruses* 2000;16:139–51.
5. Vanden Haesevelde M, Delcourt JL, De Leys RJ, et al. Genomic cloning and complete sequence analysis of a highly divergent African human immunodeficiency virus isolate. *J Virol* 1994;68:1586–96.
6. Janssens W, Heyndrickx L, Van der Auwera G, et al. Interpatient genetic variability of HIV-1 group O. *AIDS* 1999;13:41–8.
7. Peeters M, Gueye A, Mboup S, et al. Geographical distribution of HIV-1 group O viruses in Africa. *AIDS* 1997;11:493–8.
8. Mauclere P, Loussert-Ajaka I, Damond F, et al. Serological and virological characterization of HIV-1 group O infection in Cameroon. *AIDS* 1997;11:445–53.
9. Zekeng L, Gurtler L, Afane Ze E, et al. Prevalence of HIV-1 subtype O infection in Cameroon: preliminary results. *AIDS* 1994;8:1626–28.
10. Novitski VA, Montano MA, McLane MF, et al. Molecular cloning and phylogenetic analysis of human immunodeficiency virus type 1 subtype C: a set of 23 full-length clones from Botswana. *J Virol* 1999;73:4427–32.
11. Jameel S, Zafrullah M, Ahmad M, Kapoor GS, Sehgal SA. Genetic analysis of HIV-1 from Punjab, India reveals the presence of multiple variants. *AIDS* 1995;9:685–90.
12. Workshop report from the European Commission and the Joint United Nations Programme on HIV/AIDS. 1997. HIV-1 subtypes: implications for epidemiology, pathogenicity, vaccines and diagnostics. *AIDS* 1997;11:17–36.
13. Chen YM, Lee CM, Lin RY, Chang HJ. Molecular epidemiology and trends of HIV-1 subtypes in Taiwan. *J Acquir Immune Defic Syndr Hum Retrovirol* 1998;19:393–402.
14. Kanki PJ, Peeters M, Gueye-Ndiaye A. Virology of HIV-1 and HIV-2: implications for Africa. *AIDS* 1997;11(Suppl B):S33–42.
15. Peeters M, Esu-Williams E, Nzila N, et al. Molecular epidemiology of HIV-1 genetic subtypes in West and Central Africa [abstract 430/11163]. XII World Conference on AIDS. Geneva, Switzerland, 1998.
16. Rayfield MA, Downing RG, Baggs J, et al. A molecular epidemiologic survey of HIV in Uganda. *AIDS* 1998;12:521–7.
17. Renjifo B, Chaplin B, Mwakagile D, et al. Epidemic expansion of HIV type 1 subtype C and recombinant genotypes in Tanzania. *AIDS Res Hum Retroviruses* 1998;14:635–8.
18. Leitner T, Escanilla D, Marquina S, et al. Biological and molecular characterization of subtype D, G, and A/D recombinant HIV-1 transmissions in Sweden. *Virology* 1995;209:136–46.
19. Marquina S, Leitner T, Rabinovich RD, Benetucci J, Libonatti O, Albert J. Coexistence of subtypes B, F, and as B/F *env* recombinant of HIV type 1 in Buenos Aires Argentina. *AIDS Res Hum Retroviruses* 1996;12:1651–4.
20. Montavon C, Bibollet-Ruche F, Robertson D, et al. The identification of a complex A/G/I/J recombinant HIV-1 virus in different West African countries. *AIDS Res Hum Retroviruses* 1999;15:1707–12.
21. Peeters M, Liegeois F, Torimiro N, et al. Characterization of a highly replicative intergroup M/O recombinant HIV-1 virus isolated from a Cameroonian patient. *J Virol* 1999;73:7368–75.
22. 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.
23. Howard TM, Olaylele DO, Rasheed S. Sequence analysis of the glycoprotein 120 coding region of a new HIV type 1 subtype A strain (HIV-IIbNg) from Nigeria. *AIDS Res Hum Retroviruses* 1994;10:1755–7.
24. 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.
25. Lasky M, Perret JL, Peeters M, et al. Presence of multiple non-B

subtypes and divergent subtype B strains of HIV-1 in individuals infected after overseas deployment. *AIDS* 1997;11:43–51.

26. Sander-Buell E, Salminen MO, McCutchan FE. Sequencing primers for HIV-1. In: *Human retroviruses and AIDS,* pt 3. Los Alamos, NM: Los Alamos National Laboratory, 1995: 15–21.

27. Delwart EL, Shpaer EG, Louwagie J, et al. Genetic relationships determined by a DNA heteroduplex mobility assay: analysis of HIV-1 *env* genes. Science 1993;262:1257–61.

28. 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.

29. 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.

30. Ray SC. *Software Simplot for Windows, version 2.5,* 1999. Available at http://www.med.jvh.edu/deptmed/srav

31. Salminen MO, Carr JK, Burke DS, McCutchan FE. Identification of breakpoints in intergenotypic recombinants of HIV type 1 by bootscanning. *AIDS Res Hum Retroviruses* 1995;11:1423–5.

32. Gao F, Robertson DL, Carruthers CD, et al. A comprehensive panel of near-full-length clones and reference sequences for non-subtype B isolates of human immunodeficiency virus type 1. *J Virol* 1998;77:5680–98.

33. Robertson DL, Gao F, Hahn BH, Sharp PM. Intersubtype recombinant HIV-1 sequences. In: *Human retrovirus and AIDS,* pt 3. Los Alamos, NM: Los Alamos National Laboratory, 1997:25–30.

34. McCutchan FE, Salminen MO, Carr JK, Burke DS. HIV-1 genetic diversity. *AIDS* 1996;10(Suppl 3):S13–20.

35. Peeters M, Esu-Williams E, Vergne L, et al. Predominance of A and G Human Immunodeficiency Viruses type 1 in Nigeria with geographical differences in their distribution. *AIDS Res Hum Retroviruses* 2000;16:315–25.

36. McCutchan FE, Carr JK, Bajani M, et al. Subtype G and multiple forms of A/G intersubtype recombinant human immunodeficiency virus type 1 in Nigeria. *Virology* 1999;254:226–34.

37. Kalish ML, Baldwin A, Raktham S, et al. The evolving molecular epidemiology of HIV-1 envelope subtypes in injecting drug users in Bangkok, Thailand: implications for HIV vaccine trials. *AIDS* 1995;9:851–7.

38. Shao Y, Su L, Zhao F, et al. Genetic recombination of HIV-1 strains identified in China [abstract 429/11179]. Presented at the XII World Conference on AIDS, Geneva, 1998.

39. 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.

40. Liitsola K, Tashkinova I, Laukkanen T, et al. HIV-1 genetic subtype A/B recombinant strain causing an explosive epidemic in injecting drug users in Kaliningrad. *AIDS* 1998;12:1907–19.

41. Brodine SK, Mascola JR, Weiss PJ, et al. Detection of diverse HIV-1 genetic subtypes in the USA. *Lancet* 1995;346:1198–9.

42. Heyndrickx L, Janssens W, Coppens S, et al. HIV type 1 C2V3 *env* diversity among Belgian individuals. *AIDS Res Hum Retroviruses* 1998;14:1291–6.

43. Descamps D, Apetrei C, Collin G, Damond F, Simon F, Brun-Vézinet F. Naturally occurring decreased susceptibility of HIV-1 subtype G to protease inhibitors. *AIDS* 1998;12:1109–11.

Rapid Communications

Articles accepted as Rapid Communications will normally be published within 8 weeks of acceptance. When submitting a paper for consideration as a Rapid Communication, please adhere to the following guidelines:

- Submit your paper by fax or air express service to the appropriate editor with an accompanying letter stating why the paper merits special attention.
- All figures should be submitted by overnight express.
- The paper should not exceed 4 printed pages (approximately 12 double-spaced type-written manuscript pages, including illustrations, tables, and references).
- The paper should include an abstract, key words, methods, results, discussion, and reference sections.
- The title page should include the corresponding author's telephone and fax numbers.
- Authors will receive proofs of their article for review by fax and will be expected to return corrections by fax within **48 hours** of receipt. Changes received after this deadline will not be accepted.

Papers that are not accepted as Rapid Communications may be resubmitted as full-length articles.

JAIDS

JOURNAL OF ACQUIRED IMMUNE DEFICIENCY SYNDROMES

ISSN = 1525-4135