



# Unprecedented Degree of Human Immunodeficiency Virus Type 1 (HIV-1) Group M Genetic Diversity in the Democratic Republic of Congo Suggests that the HIV-1 Pandemic Originated in Central Africa

NICOLE VIDAL,<sup>1</sup> MARTINE PEETERS,<sup>1\*</sup> CLAIRE MULANGA-KABEYA,<sup>1</sup> NZILA NZILAMBI,<sup>2</sup>  
DAVID ROBERTSON,<sup>3</sup> WANTABALA ILUNGA,<sup>4</sup> HUROGO SEMA,<sup>5</sup> KAZADI TSHIMANGA,<sup>6</sup>  
BENI BONGO,<sup>7</sup> AND ERIC DELAPORTE<sup>1,8</sup>

Laboratoire Retrovirus, IRD,<sup>1</sup> and CHU, Gui de Chauliac,<sup>8</sup> Montpellier, France; Projet SIDA<sup>2</sup> and BCC/SIDA,<sup>7</sup> Kinshasa, BRC/SIDA<sup>4</sup> and Hôpital Bonzola,<sup>6</sup> Mbuyi-Maya, and CDI, Bwamanda,<sup>5</sup> Democratic Republic of Congo; and Department of Zoology, University of Oxford, Oxford, United Kingdom<sup>3</sup>

Received 13 June 2000/Accepted 1 August 2000

The purpose of this study was to document the genetic diversity of human immunodeficiency virus type 1 (HIV-1) in the Democratic Republic of Congo (DRC; formerly Zaire). A total of 247 HIV-1-positive samples, collected during an epidemiologic survey conducted in 1997 in three regions (Kinshasa [the capital], Bwamanda [in the north], and Mbuyi-Maya [in the south]), were genetically characterized in the *env* V3-V5 region. All known subtypes were found to cocirculate, and for 6% of the samples the subtype could not be identified. Subtype A is predominant, with prevalences decreasing from north to south (69% in the north, 53% in the capital city, and 46% in the south). Subtype C, D, G, and H prevalences range from 7 to 9%, whereas subtype F, J, K, and CRF01-AE strains represent 2 to 4% of the samples; only one subtype B strain was identified. The highest prevalence (25%) of subtype C was in the south, and CRF01-AE was seen mainly in the north. The high intersubtype variability among the V3-V5 sequences is the most probable reason for the low (45%) efficiency of subtype A-specific PCR and HMA (heteroduplex mobility assay). Eighteen (29%) of 62 samples had discordant subtype designations between *env* and *gag*. Sequence analysis of the entire envelope from 13 samples confirmed the high degree of diversity and complexity of HIV-1 strains in the DRC; 9 had a complex recombinant structure in gp160, involving fragments of known and unknown subtypes. Interestingly, the unknown fragments from the different strains did not cluster together. Overall, the high number of HIV-1 subtypes cocirculating, the high intrasubtype diversity, and the high numbers of possible recombinant viruses as well as different unclassified strains are all in agreement with an old and mature epidemic in the DRC, suggesting that this region is the epicenter of HIV-1 group M.

Phylogenetic analysis of many isolates of human immunodeficiency virus type 1 (HIV-1) from Africa and from other regions of the world revealed three major lineages of HIV-1: group M (for main), group N (for non-M/non-O), and group O (for outlier) (7, 35, 53). Within group M, subtypes and circulating recombinant forms (CRFs) have been proposed (5). To be considered a subtype, in phylogenetic analysis the strains should resemble each other, and no other particular lineage, across the entire genome, and CRFs should have similar mosaic genomes with the same intersubtype breakpoints (48). By these criteria, there are nine subtypes of HIV-1 group M: A, B, C, D, F, G, H, J, and K. All known representatives of what were initially described as subtype E appear in fact to be recombinants of subtypes A and E (4, 15) and are now designated CRF01\_AE (48). CRF02\_AG corresponds to the IbNg strain from Ibadan, Nigeria, a complex mosaic virus of alternating subtype A and subtype G sequences (6, 20) which is highly prevalent in western Africa (32, 43). CRF03\_AB strains are responsible for the explosive HIV outbreak among intravenous drug users in Kaliningrad (27). CRF04\_cpx viruses correspond to the previously described *env* subtype I viruses (19, 25, 36).

Subtype designations have been powerful molecular epidemiological markers to track the course of the HIV-1 pandemic. Preliminary data indicate a very heterogeneous distribution and dominance of different genetic subtypes depending on the country analyzed. In Africa, all known HIV-1 genetic subtypes and groups, including groups N and O, are present (12a, 22, 44). Whether the various groups, subtypes, and recombinant forms of HIV-1 have biological differences (for example, with respect to transmissibility and the course of disease progression) is not known (21, 44). A relationship between genetic subtype and natural resistance against antiretroviral drugs (3, 10, 11), as well as between subtypes and the efficiency of serological and molecular test for HIV diagnosis (2, 28, 41), has been observed. The degree to which vaccines based on one subtype will elicit cross-protection against other subtypes is still poorly understood (60). For the above-mentioned reasons, it is important to study the geographic distribution of the different HIV-1 genetic subtypes.

The Democratic Republic of Congo (DRC) is located in Central Africa and is bordered by nine countries: Congo-Brazzaville on the west, the Central African Republic and Sudan on the north, Uganda, Rwanda, Burundi, and Tanzania on the east, and Zambia and Angola on the south. It is the second-largest country on the continent (2,344,885 km<sup>2</sup>) and has a total population of 30 million inhabitants, 60% of whom live in rural areas. In 1984, the government of the DRC became one

\* Corresponding author. Mailing address: Laboratoire Retrovirus, IRD, BP 5045, 34032 Montpellier Cdx 1, France. Phone: 33-4 67 41 62 97. Fax: 33-4 67 61 94 50. E-mail: martine.peeters@mpl.ird.fr.

Fonds Documentaire IRD



010025698

10498

Fonds Documentaire IRD

Cote: B x 25698 Ex: 1

of the first in Africa to endorse a national policy for HIV/AIDS prevention and control, and it committed support to basic research and epidemiological studies through *Projet SIDA*. Many reference data regarding HIV prevalence, data on behavioral, biological, and demographic factors, as well as reports on preventive interventions were generated until 1991 (30, 38, 39, 49). The country has been relatively peaceful, but mismanagement and corruption have led to a severe social and economic crisis resulting in a deterioration of the socioeconomic situation which accelerated in the early 1990s. In September 1991 and January 1993 there was a political crisis, and the resultant rioting led to a highly disorganized political and social environment, disrupting prevention activities, which resulted in all foreign staff being evacuated. With the withdrawal of foreign funding since 1991, many government facilities are no longer functional and several have been closed. As a result of this breakdown of the health system, aggravated in some areas by the massive exodus of Rwandan (Kivu region) and internal (Kasai region) refugees, new disease outbreaks (e.g., cholera and Ebola virus infection) have been documented. Against this background, we conducted in 1997 a seroepidemiologic survey; surprisingly, the results showed that HIV prevalence rates had remained relatively unchanged in selected populations over a 10-year period (33).

Several HIV subtypes have been associated with the DRC through patients living in Europe, but only a limited number of samples collected in the DRC have been genetically characterized: 14 from Kinshasa and 66 from Kimpese, a rural town situated at 200 km west of Kinshasa (1, 26, 31, 45, 59). These results suggest a high genetic diversity of HIV-1, but the precise distribution of HIV-1 subtypes in the DRC remains poorly documented. The purpose of this large serosurvey conducted in 1997 on selected population groups from different geographic locations in the DRC was to obtain current data on HIV infection, especially on the relative prevalences of the HIV-1 subtypes circulating in the DRC.

#### MATERIALS AND METHODS

**Specimen and DNA isolation.** A total of 247 HIV-1-positive samples were genetically characterized. The samples were collected during an epidemiologic survey conducted in April 1997 (33) in three regions of the DRC: Equateur Province (Bwamanda) in the north, bordering the Central African Republic; Kasai Province (Mbuji-Mayi) in the south, bordering Zambia; and the capital city, Kinshasa, in the west. Bwamanda and Mbuji-Mayi are fairly rural compared with Kinshasa. Participants were recruited among female sex workers, tuberculosis patients, patients clinically suspect for HIV infection, pregnant women, and blood donors. After giving informed consent, the participants were interviewed according to a standardized questionnaire developed for this study. The questionnaire included demographic characteristics such as age, nationality, and travel out of the DRC.

Blood samples were collected in EDTA anticoagulant tubes. From all HIV-positive samples, plasma and peripheral blood mononuclear cells were separated by Ficoll gradient centrifugation. Plasma and cell pellets were stored at  $-20^{\circ}\text{C}$  and shipped on dry ice for further genetic characterization.

DNA was extracted from the dry cell pellets by using an IsoQuick isolation kit (Microprobe Corp., Garden Cove, Calif.) or a Qiagen (Courtabeuf, France) DNA isolation kit.

**Subtype A-specific PCR.** Subtype A-specific PCR was done as previously described (42). Briefly, a nested PCR was performed to obtain a 350-bp fragment using outer primer pair ED5-ED12 (the same as used for heteroduplex mobility assay [HMA]) and inner primers LRM1B and LRM3B (specific for subtype A). An initial denaturation step for 5 min at  $94^{\circ}\text{C}$  was followed by 30 cycles of  $94^{\circ}\text{C}$  for 15 s,  $55^{\circ}\text{C}$  for 30 s, and  $72^{\circ}\text{C}$  for 2 min, with a final extension for 7 min at  $72^{\circ}\text{C}$  for the first round. Five microliters from this first round was used for the second round with the inner primers, using the following conditions for 35 cycles:  $94^{\circ}\text{C}$  for 20 s,  $55^{\circ}\text{C}$  for 30 s, and  $72^{\circ}\text{C}$  for 1 min, with a final extension of 5 min at  $72^{\circ}\text{C}$ . The PCR amplification products were detected by electrophoresis on a 1% agarose gel and visualized by ethidium bromide staining.

**HMA.** The V3-V5 region from the envelope gene was amplified by a nested PCR as previously described (9) with ED5 and ED12 as outer primers and with ES7 and ES8 as inner primers. The PCR conditions were as follows: an initial denaturation step for 2 min at  $92^{\circ}\text{C}$ , followed by 30 cycles of  $92^{\circ}\text{C}$  for 20 s,  $55^{\circ}\text{C}$  for 30 s, and  $72^{\circ}\text{C}$  for 2 min, with a final extension for 7 min at  $72^{\circ}\text{C}$  for the first round. One to five microliters from this amplification was used for the second round with the inner primers, using the following cycling conditions for 40 cycles:  $92^{\circ}\text{C}$  for 20 s,  $55^{\circ}\text{C}$  for 30 s, and  $72^{\circ}\text{C}$  for 2 min. The reaction mixture consisted of 50 mM KCl, 10 mM Tris-HCl (pH 9), 0.1% Triton X-100, 1.25 (for the first round) or 1.8 (for the second round) mM  $\text{MgCl}_2$ , 0.2 mM each deoxynucleoside triphosphate, 2.5 U of *Taq* polymerase, and 10 pmol of each primer for the first round and 20 pmol for the second round. The PCR amplification products were detected by electrophoresis on a 1% agarose gel and visualized by ethidium bromide staining. To avoid PCR product cross-contamination, pre-PCR and post-PCR manipulations were performed in separate rooms.

Heteroduplex molecules were obtained by mixing 4  $\mu\text{l}$  of two divergent PCR-amplified DNA fragments (the unknown patient strain with a plasmid from typed reference strains) denatured at  $94^{\circ}\text{C}$  for 2 min and renatured by rapid cooling on wet ice. The reference strains used in this study, and countries of isolation, were A1 (RW20, Rwanda), A2 (IC144, Ivory Coast), A3 (SF170, Rwanda), B1 (BR20, Brazil), B2 (TH14, Thailand), B3 (SF162, United States), C1 (MA959, Malawi), C2 (ZM18, Zambia), C3 (IN868, India), C4 (BR25, Brazil), D1 (UG21, Uganda), D2 (UG38, Uganda), D3 (UG46, Uganda), E1 (TH22, Thailand), E2 (TH06, Thailand), E3 (CAR7, Central African Republic), F1 (BZ162, Brazil), F2 (BZ163, Brazil), G1 (RU131, Russia), G2 (LBV21-7, Gabon), G3 (VI525, Gabon), H1 (CA13, Cameroon), H2 (VI557, Zaire), and H3 (VI997, Belgium). The reaction was performed in 100 mM NaCl-10 mM Tris-HCl (pH 7.8)-2 mM EDTA in a final volume of 9  $\mu\text{l}$ . Heteroduplex formation was resolved by electrophoresis at 250 V for 3 h on a nondenaturing 5% polyacrylamide gel in TBE buffer (88 mM Tris borate, 89 mM boric acid, 2 mM EDTA) and was detected after ethidium bromide staining. The electrophoretic mobility of the heteroduplexes was inversely proportional to the sequence divergence of the two annealed strands.

**Genetic subtyping by sequence and phylogenetic tree analysis (i) Sequencing of the envelope V3-V5 region.** The genetic subtype in the envelope was determined by direct sequencing of the V3-V5 region of the envelope (700 bp). The samples were amplified with the HMA primers (ED5 and ED12 as outer primers; ES7 and ES8 as inner primers) under the same conditions as described above. The amplified DNA was purified using a QiaQuick gel extraction kit (Qiagen). The amplified products were directly sequenced using fluorescent dye terminator technology (dye terminator cycle sequencing with AmpliTaq DNA polymerase FS; Perkin-Elmer, Roissy, France) or Big-Dye chemistry (Perkin-Elmer) as instructed by the manufacturer. Electrophoresis and data collection were done on an Applied Biosystems 373A automatic DNA sequencer (Stretch model).

**(ii) Sequencing of the gag p24 region.** A 700-bp fragment corresponding to the p24 region from the gag gene was amplified with previously described primer pairs G00-G01 and G60-G25 (52). The PCR conditions were as follows: an initial denaturation step for 3 min at  $92^{\circ}\text{C}$ , followed by 30 cycles of  $92^{\circ}\text{C}$  for 10 s,  $55^{\circ}\text{C}$  for 30 s, and 1 min at  $72^{\circ}\text{C}$ , with a final extension for 7 min at  $72^{\circ}\text{C}$ , in a final volume of 50  $\mu\text{l}$ . The reaction mixture consists of 50 mmol of KCl/liter, 10 mmol of Tris-HCl (pH 9)/liter, 0.1% Triton X-100, 1.4 mmol of  $\text{MgCl}_2$ /liter, 10 pmol of each primer, 0.2 mmol of each deoxynucleoside triphosphate/liter, and 2.5 U of *Taq* polymerase. One microliter from 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  $\mu\text{l}$ . 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 QiaQuick gel extraction kit (Qiagen) and directly sequenced using Big-Dye chemistry (Perkin-Elmer). Electrophoresis and data collection were done on an Applied Biosystems 373A automatic DNA sequencer (Stretch model).

**(iii) Sequencing of the entire envelope (gp160).** The entire envelope was amplified by a nested PCR as previously described by Gao and coworkers (14), with outer primers A and N and inner primers B and M. The amplification reaction was performed with the Expand Long Template PCR system (Boehringer Mannheim, Indianapolis, Ind.) as instructed by the manufacturer. Briefly, the conditions for both PCR rounds were as follows: a denaturation step of 4 min at  $94^{\circ}\text{C}$ ; 10 cycles of  $92^{\circ}\text{C}$  for 20 s,  $50^{\circ}\text{C}$  for 30 s, and  $68^{\circ}\text{C}$  for 4 min; 20 cycles with 20-s increments at the elongation step. PCR fragments were purified with a QiaQuick gel extraction kit (Qiagen). Direct sequencing of the envelope was done with *env* B, *env* M, and other primers encompassing the envelope.

**(iv) Phylogenetic analysis.** Nucleotide sequences were aligned using CLUSTAL W (57), with minor manual adjustments as appropriate for the protein sequences. Regions that could not be aligned unambiguously, due to length or sequence variability, were omitted from the analysis. Phylogenetic trees generated by the neighbor-joining method (50) and reliability of the branching orders determined by the bootstrap approach (13) were implemented with CLUSTAL W. Genetic distances were calculated with the Kimura's two-parameter method (23).

**Intersubtype recombinant analysis.** To analyze whether the viruses were recombinant in the sequenced regions, several additional analyses were performed. Diversity plots, using the online program DIVERT (<http://igs-server.cnr-mrs.fr/ans/phylogenetics>), determined the percent diversity between selected pairs of sequences by moving a window of 300 bp along the genome alignment in 20-bp

increments. The divergence values for each pairwise comparison were plotted at the midpoint of each 300-bp segment. Simplot for Windows version 2.5 (distributed by author, S. C. Ray [http://www.med.jhu.edu/deptmed/sray/download/]) 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 CONSENSE from the PHYLIP package for a 500-bp window moved along the alignment in increments of 10 bp. We evaluated 100 replicates for each phylogeny. The bootstrap values for the studied sequences were plotted at the midpoint of each window.

Nucleotide sequence accession numbers. The sequences reported have been submitted to GenBank under the following accession numbers: AJ404008 to AJ404203 and AJ404487 for the *env* V3-V5 sequences, AJ404232 to AJ404293 for the *gag* p24 sequences, and AJ401034 to AJ401046 for the gp160 sequences.

## RESULTS

**Study population.** A total of 247 HIV-1-positive samples, collected during the epidemiologic survey conducted in April 1997 and previously published (33), were genetically characterized. Overall, 142 samples from Kinshasa, the capital city in the west; 60 from Mbuyi-Mayi, Kasai Province, in the south; and 45 samples from Bwamanda, Equateur Province, in the north were studied. HIV-1-positive samples were obtained from 84 tuberculosis patients, 47 from Kinshasa, 26 from Bwamanda, and 11 from Mbuyi-Mayi. Samples from 66 patients suspected to have AIDS, 27 from Kinshasa, 21 from Mbuyi-Mayi, and 18 from Bwamanda, were studied. Samples from pregnant women ( $n = 34$ ) were obtained mainly from Kinshasa ( $n = 13$ ) and Mbuyi-Mayi ( $n = 20$ ). An additional 63 samples were included from miscellaneous population groups: 20 sexually transmitted disease (STD) patients from Kinshasa, 8 blood donors (6 from Mbuyi-Mayi and 2 from Kinshasa), 25 female sex workers (24 from Kinshasa and 1 from Mbuyi-Mayi), and 10 asymptomatic adults. The population groups were predominantly young adults aged 20 to 30 years; less than 2% of them had traveled outside the DRC.

**Efficiency of rapid and simple tools for genetic subtyping.** A total of 109 samples were analyzed at random by a subtype A-specific PCR as previously described. Only 27 samples (24.7%) were reactive in this assay; these were confirmed by HMA and/or sequencing as being subtype A. Among the 82 samples identified as non-A by this method, 33 were classified as A by HMA and/or sequencing and 49 were classified as non-A. The subtype A-specific PCR allowed the rapid identification of only a quarter of the randomly chosen samples from the DRC; moreover, only 45% (27 of 60) of the subtype A samples could be detected.

The HMA is an easily used tool for identifying the genetic subtype of HIV-1 and requires equipment less sophisticated than that used for sequencing. A total of 88 samples, randomly chosen from our study, were assessed with this technique. For only 40 (45%) could the genetic subtype be identified; subtypes of the remaining 48 samples were undetermined by HMA. The samples identified by HMA were subtypes A ( $n = 34$ ), D ( $n = 3$ ), G ( $n = 1$ ), and H ( $n = 2$ ). Sequence analysis of the same region in the envelope followed by phylogenetic tree analysis classified the HMA-undetermined samples as follows: 15 A, 1 B, 8 C, 5 D, 3 CRF01-AE, 1 F1, 3 G, 6 H, 1 J, 4 K, and 1 that did not cluster with any of the known subtypes.

The low number of subtype A samples recognized by the subtype A-specific primers and the high number of samples remaining undetermined by HMA suggest a much higher genetic variability among the HIV-1 samples from the DRC than in western or eastern Africa. Due to the low efficiency of these rapid and/or simple subtyping techniques (less than 25% for subtype A-specific PCR and only 45% for HMA), genetic subtyping was continued by direct sequencing of the V3-V5 region of the envelope for the remaining samples. Overall, for

197 of the 247 samples tested, the final genetic subtype was identified by sequence and phylogenetic tree analysis of the *env* V3-V5 region.

**Geographic distribution of HIV-1 *env* genetic subtypes in the DRC.** Figure 1 and Table 1 show the regional distribution of HIV-1 subtypes in the DRC. The genetic heterogeneity in the V3-V5 region from the envelope is remarkable in all three regions. In particular, multiple HIV-1 subtypes are cocirculating, there being at least 6 subtypes in Bwamanda, 7 in Mbuyi-Mayi, and 10 in Kinshasa. Moreover, the genetic subtype distribution differs between the three regions. Subtype A is predominant in the three cities; the lowest prevalences were observed in Kinshasa (43.7%) and Mbuyi-Mayi (56.7%), and the highest prevalence was seen in Bwamanda (68.9%). One subtype B strain was documented in the north. Subtype C was largely present in the south, representing 25% of the samples tested in Mbuyi-Mayi. Subtype C was absent in the north and represented only 2.2% of the samples in Kinshasa. Subtype D was documented in Kinshasa (19 of 142 [13.4%]), and Mbuyi-Mayi (4 of 60 [6.6%]). CRF01-AE was seen in the north, representing 3 (6.7%) of the 45 samples tested in that region and in Kinshasa (0.7%). Subtype F was seen in low prevalences in the three regions; 9 of the 10 samples were F1, and only one F2 sample was identified (in Kinshasa). Subtype G represented 10.5% (15 of 142) of the HIV-1 strains in Kinshasa and 3.3% (2 of 60) of those in Mbuyi-Mayi. Subtype H was observed in Kinshasa and Bwamanda, where it represents 9.8% (14 of 142) and 8.8% (4 of 45) of the circulating strains. Subtype J was seen at lower prevalences, 3.5 and 3.3% in Kinshasa and the south, respectively, and prevalences of subtype K were between 1.7 and 4.4%. A total of 15 (6%) strains did not cluster with any of the known subtypes, and the majority of them represented unique lineages; 11 were from Kinshasa, 3 were from Bwamanda, and 1 was from Mbuyi-Mayi. None of the strains from our study clustered with the subtype I fragment in the V3-V5 region of the CRF04\_cpx viruses.

The predominant population groups tested were tuberculosis and AIDS patients, and in some regions they represented the majority of the samples tested. The difference in subtype distribution according to geographic region is confirmed if tuberculosis and AIDS patients from each region are compared to each other, especially for subtype A. Overall, 42.3% of the 85 tuberculosis patients were infected with subtype A, which is lower than the overall subtype A prevalence of 51.4%. But despite the low sample numbers, the prevalence of subtype A was higher in Bwamanda (17 of 26 [65.4%]) than in Kinshasa (16 of 47 [34%]) and Mbuyi-Mayi (3 of 12 [25%]). Six of 12 tuberculosis patients in Mbuyi-Mayi were infected with subtype C, versus only 1 of 47 in Kinshasa. Of the 64 AIDS patients tested, 56.2% were infected with subtype A, which is slightly higher than the overall prevalence of subtype A when all population groups are taken together. Again, the difference in geographic distribution of subtype A was confirmed, with 13 of 18 (72%) in Bwamanda, 12 of 20 (60%) in Mbuyi-Mayi, and 11 of 27 (40.7%) in Kinshasa. As found for the tuberculosis patients, subtype C AIDS patients were seen only in Mbuyi-Mayi and subtype E was found only in the north.

**Phylogenetic and distance analysis of the HIV-1 group M sequences in the envelope.** Figure 2 shows the phylogenetic tree of envelope sequences covering the V3-V5 region for the sequences obtained from Kinshasa (Fig. 2a) and from the two other cities (Bwamanda and Mbuyi-Mayi) in the DRC (Fig. 2b). The phylogenetic analysis shows a very high degree of divergence within each subtype. To measure the intrasubtype distances in the V3-V5 region among strains circulating in the DRC, we calculated the distances based on a phylogenetic tree

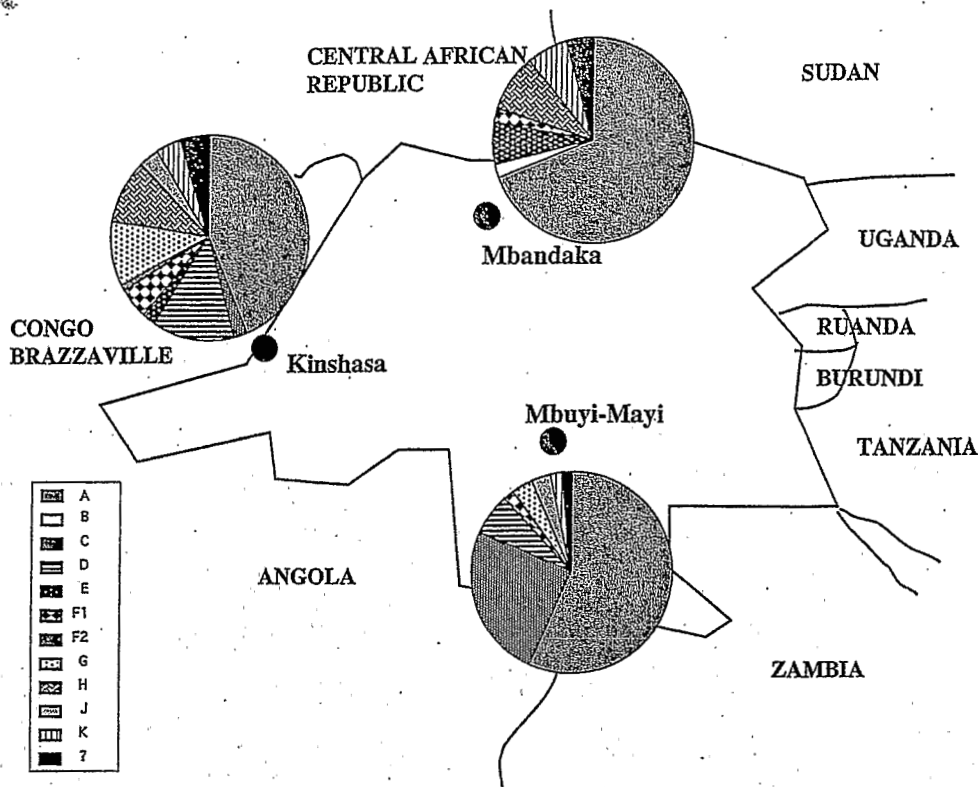


FIG. 1. Distribution of HIV-1 *env* genetic subtypes in different geographic locales from the DRC.

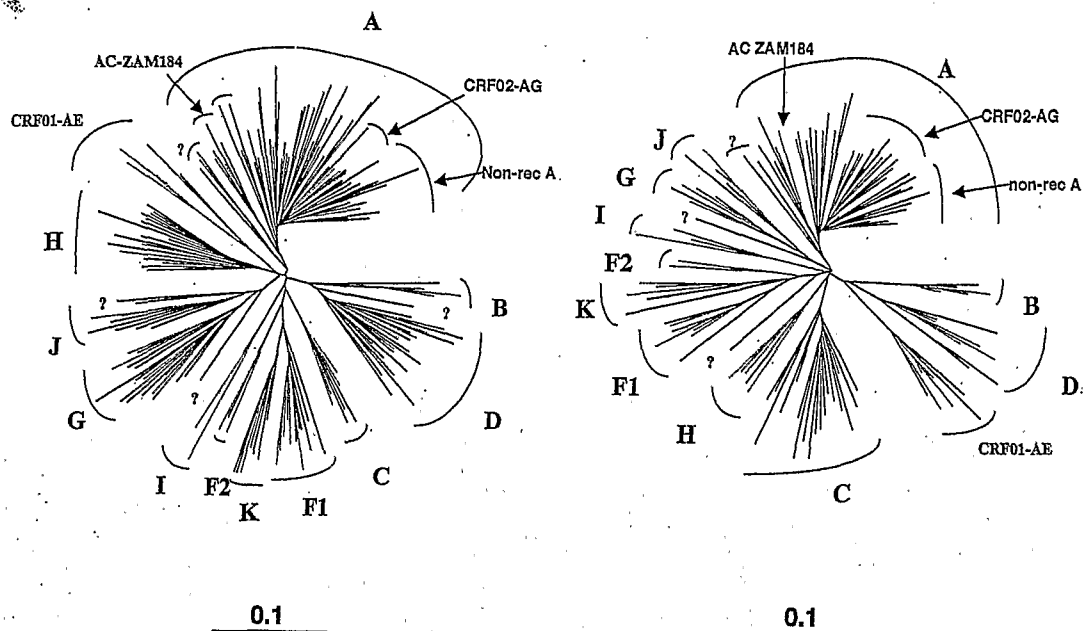
which included all the V3-V5 sequences from the three regions studied. Pairwise nucleotide sequence distances were estimated using the Kimura two-parameter model. The reference strains used in this analysis were from strains representing each known subtype, and when possible for each subtype representatives obtained from different geographic locales were chosen: subtype A references were from Kenya (A-KE.Q2317), Uganda (A-UG.92UG307) and Somalia (A-SE.SOSE7253); subtype B strains were from the United States (US.JRFL, US.RF, US.WEAU160) and France (B-FR.HXB2R); subtype C strains were from Ethiopia (C-ET.ET2220), Brazil (C-BR.92BR025), India (C-IN.21068), and Botswana (C-BW.96BW0502); subtype D strains were from the DRC (D-ZR.NDK, D-ZR.ELI, and D-ZR.84ZR085) and Uganda (D-UG.94UG114); F1 vi-

ruses were from Brazil (F1-BR.93BRO20, F1-BR.BZ163, and F1-BR.BZ126), a Belgian infected in the DRC (F1-BE.VI850), and from Finland (F1-FI.FIN6393); F2 references were available only from Cameroon (F2-CM95MP255 and F2-CM95MP257); subtype G isolates were from a Finnish patient infected in Kenya (G-FI.HH8793), a Swedish patient infected in the DRC (G-SE.SE6165), and from DRC (G-BE.DRCBL); subtype H was from Belgian patients infected in Central Africa (H-BE.VI991 and H-BE.VI997) and the Central African Republic (H-CF.90CF056); subtype J samples were from Africans living in Sweden (J-SE.SE91733 and J-SE.SE92809); subtype K samples were from Cameroon (K-CM96-MP535) and the DRC (K-ZR97-EQTB11); and *env* subtype E or CRF01\_AE viruses were from Thailand (AE-TH.TN235, AE-TH.93TH253, and AE-TH.CM240) and the Central African Republic (AE-CF.90CF402).

Figure 3 summarizes the mean intra- and intersubtype distances. The genetic distances observed among the DRC strains were relatively high and were in general higher than the distances observed between reference strains from different geographic locales. For example, the mean intrasubtype distances observed among subtype A, C, F1, and AE reference strains were much lower than among the DRC strains from each of these subtypes: 9.1% versus 16.6% for subtype A, 11.7% versus 15.9% for subtype C, 8.8% versus 12.8% for subtype F1, and 8.6% versus 16.8% for CRF01\_AE strains. The reference strains used for subtypes A, C, F1, and CRF01\_AE were all from different geographic locales. Phylogenetic analysis suggests also that different subclusters can be defined within subtype A, although due to the short sequence fragment examined, these tentative subclusters are not always supported by high bootstrap values. Four strains formed a distinct subclus-

TABLE 1. Genetic subtypes in the *env* V3-V5 region of HIV-1 isolates from three cities in the DRC

Subtype	No. (%)			Total
	Kinshasa (n = 142)	Bwamanda (n = 45)	Mbuyi-Mayi (n = 60)	
A	62 (43.7)	31 (68.9)	34 (56.7)	127 (51.4)
B		1 (2.2)		1 (0.4)
C	3 (2.2)		15 (25.0)	18 (7.3)
D	19 (13.4)		4 (6.6)	23 (9.3)
E	1 (0.7)	3 (6.7)		4 (1.6)
F	8 (5.6)	1 (2.2)	1 (1.7)	10 (4.1)
G	15 (10.5)		2 (3.3)	17 (6.9)
H	14 (9.8)	4 (8.9)		18 (7.3)
J	5 (3.5)		2 (3.3)	7 (2.8)
K	4 (2.8)	2 (4.4)	1 (1.7)	8 (3.2)
?	11 (7.7)	3 (6.6)	1 (1.7)	15 (6.0)



(a) Kinshasa

(b) Bwamanda and Mbuyi-Mayi

FIG. 2. Phylogenetic tree based on 441 unambiguously aligned nucleotides from the *env* V3-V5 region of the 197 new HIV-1 isolates, from Kinshasa (a) and from Bwamanda and Mbuyi-Mayi (b), and reference strains representing the different genetic subtypes: A-KE.Q2317, A-SE.SOSE7253, A-92UG037, CRF02-AG-IBNG, CRF02-AG-DJ263, CRF02-AG-DJ264, B-RF, B-WEAU160, B-JRFL, B-HBX2, C-ETH2220, C-92BR025, C-IN.21068, C-BW.96BW0502, D-NDK, D-ZR.84ZR085, D-94UG114, D-ELI, CRF01-AE-90CR402, CRF01-AE-93TH253, CRF01-AE-CM240, F1-93BR020, F1-BZ163, F1-BZ126, F1-BE.VI850, F1-FL.FIN6393, F2-95CMMP255, F2-95CMMP257, G-BE.DRCBL, G-HH8793, G-SE6165, H-90CF056, H-BE.VI991, H-BE.VI997, J-SE.SE91733, J-SE.SE92809, K-96CMMP535, and K-97ZREQTB11. The analysis was performed as described in Materials and Methods. Reference strains are in grey, and strains from the DRC are indicated in black. Non-rec, nonrecombinant.

ter, supported by 98% of the bootstrap values with the previously described AC-ZAM184 strain from Zambia, which was classified as subtype A in the studied region (51). Another cluster, supported by 95% of the bootstraps, is formed by three DRC strains from our study, KCC2, KCC3, and KTB13. The known nonrecombinant subtype A reference strains together with the CRF02-AG (IbNg) viruses and some DRC strains form a subcluster, and a fourth subcluster of exclusively DRC subtype A strains is also present. These latter subclusters were not supported by high bootstrap values.

**Genetic subtypes in *gag* (p24) and *env* (V3-V5).** To determine the proportion of recombinant viruses that circulate in the DRC, 62 samples (24 *env* A, 2 *env* C, 3 *env* D, 2 *env* CRF01\_AE, 4 *env* F1, 1 *env* F2, 3 *env* G, 2 *env* H, 4 *env* J, 4 *env* K, and 13 unclassified) were also sequenced in the p24 region from the *gag* gene. Overall, 18 (29%) of the 60 samples, including 2 CRF01\_AE strains, had discordant subtype designations between *env* and *gag*. Among the remaining 16 discordant samples, 11 different profiles were seen: one G/A, two ?/D, one D/F1, one G/F2, one G/H, one A/J, one D/J, two ?/K, three G/? , one A/? , and one K/? for *gag* and *env* subtypes, respectively. Seven samples which could not be classified in *env* did not cluster with any of the known subtypes in *gag* either. More importantly, the majority of these unclassified samples also did not cluster together. Almost all of the circulating subtypes are involved in recombination events.

Figure 4 shows the phylogenetic tree of the *gag* sequences covering the p24 region. As found for the *env* sequences, phylogenetic analysis of the *gag* sequences shows a very high diversity within each subtype, and we can also distinguish different subclusters within subtype A. As observed in the *env* tree,

we see in *gag* a separate subcluster with the AC-ZAM184 strain, which is classified as A in *gag*, and we can also identify subclusters of the nonrecombinant subtype A viruses, the CRF02-AG (IbNg) viruses, and a separate group of DRC *gag* sequences. As for the partial *env* subtype A sequences, the identification of subclusters is not always supported by high bootstrap values.

**Genetic characterization of the entire envelope.** The entire envelope was sequenced for 13 samples. For seven of them (three J, two A, one D, and one F1) the subtype was identified in the V3-V5 region; for six the genetic subtype could not be identified in this region of the envelope. The phylogenetic tree analysis of the gp160 sequences is shown in Fig. 5a. The genetic subtype could be clearly identified in the entire envelope for only 5 of the 13 samples: the 3 subtype J viruses, 1 subtype D, and 1 subtype G. The eight other viruses either did not cluster with any of the known subtypes or were only distantly related to them, suggesting an intersubtype recombinant structure in gp160.

Figure 5b shows structures of the different sequences after complementary analysis using the diversity plots, bootscanning, and confirmation of the results by phylogenetic trees. The three samples (MBTB 4, MBS 41, and KTB 147) identified as subtype J in the V3-V5 region were also classified as subtype J in the gp160 and appeared to be nonrecombinant after diversity and bootscan analysis. The EQS18 strain, which was easily classified as subtype A in the V3-V5 region, did not cluster with any of the known subtypes when phylogenetic tree analysis was performed on the entire envelope sequence; more detailed analysis revealed a complex mosaic structure involving subtypes A, G, and J, and for some *env* regions the virus clustered

	A	B	C	D	A/E	F1	F2	G	H	J	K
A	16.7(9.8-24.3)*										
B	21.9	-									
C	22.1	22.6	15.9(10.2-23.2)								
D	22.5	19.5	21.2	15.7(10.9-21.4)							
A/E	22.8	23.7	22.6	22.3	16.9(11.1-23.6)						
F1	23.2	21.7	20.3	21.0	23.1	12.8(9.8-15.7)					
F2	23.2	24.3	22.7	22.2	24.5	18.7	-				
G	22.8	24.6	22.9	24.0	25.3	23.8	24.1	16.4(12.3-20.9)			
H	21.6	23.1	21.9	22.2	22.8	22.8	23.7	23.6	17.3(11.6-22.5)		
J	21.9	23.3	21.5	22.4	24.1	22.1	23.2	20.6	22.1	17.1(13-23.6)	
K	22.6	21.1	21.7	22.1	24.2	19.7	21.3	23.7	22.4	22.4	16.3(13-20)

FIG. 3. Intra- and intersubtype genetic distances in the *env* V3-V5 region of HIV-1 isolates from the DRC. Range of lowest to highest distance within a subtype is in parentheses. A/E, CRF01-AE.

always between subtypes G and J, and no clear subtype designation could be made. The KCC2 sample was a divergent subtype A virus in the V3-V5 region; the entire envelope sequence was predominantly subtype A, but a 400-bp fragment in

the middle of the envelope did not cluster with any of the known subtypes. The sample identified as subtype D in V3-V5, KCD 4, was confirmed to be a nonrecombinant subtype D in the entire envelope. The EQS16 strain, classified as a divergent F1 virus based on V3-V5 sequences, appeared to be a complex recombinant involving subtypes F1 and K and unknown subtypes. The divergent KTB 22 sample appeared to be almost entirely a nonrecombinant subtype G virus, but for a small fragment in the envelope the subtype could not be identified since it was difficult to discriminate between G and J in that region. For the five other samples that could not be classified based on partial or entire envelope sequences, KFE 267 and MBFE 250 did not cluster with any of the known subtypes in the majority of the entire envelope gene, but they formed a separate and well-supported cluster in these unknown regions. KMST 91 was a recombinant involving divergent subtype A fragments and unknown fragments, KMST 120 was a recombinant with subtype J and predominantly unknown fragments, and for KTB 54 none of the envelope gene could be classified except for the final fragment, which was subtype G. Interestingly, except for regions between the KFE 267 and MBFE 250 sequences, the unknown fragments did not cluster together.

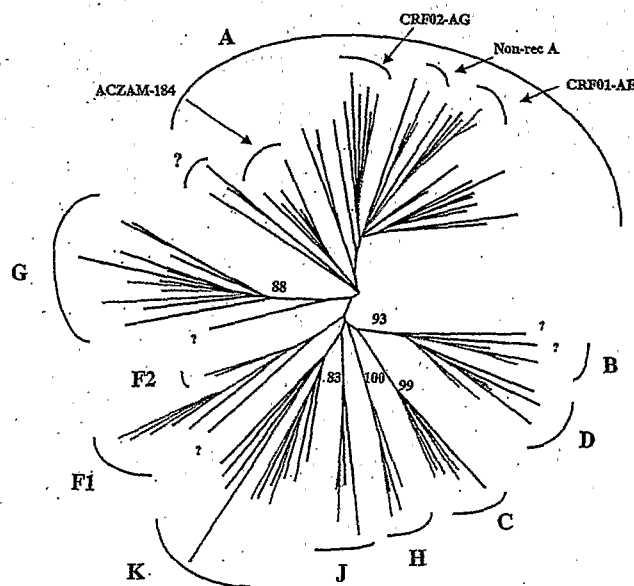


FIG. 4. Phylogenetic tree based on 621 unambiguously aligned nucleotides from the *gag* p24 region of the 62 new HIV-1 isolates and reference strains representing different genetic subtypes. The analysis was performed as described in Materials and Methods, and the same reference strains as in Fig. 2 were used. Reference strains are in grey, and strains from DRC are indicated in black.

DISCUSSION

The major goal of this study was to determine the prevalence and the geographic distribution of the genetic subtypes of HIV-1 in selected populations in the DRC. A very high genetic diversity was seen with all known subtypes found to be cocirculating. Regional differences were seen in subtype distribution, but in each region at least 6 to 10 different clades cocirculated. Overall, subtype A is predominant, with decreasing prevalence from north to south: 69% in Equateur Province, 53% in the capital city, and 46% in the south. Prevalences of subtypes C, D, G, and H range between 7 and 9%, whereas

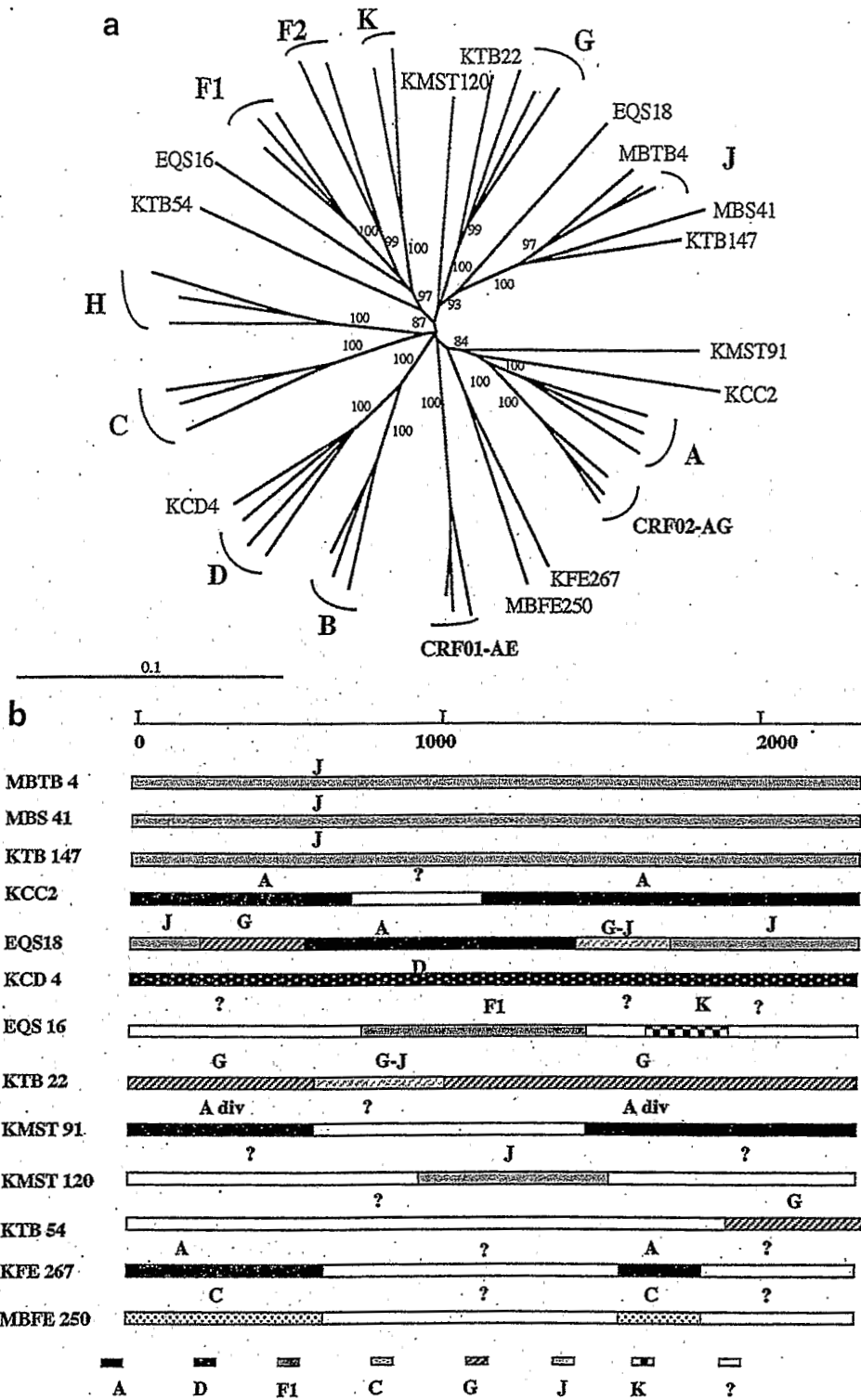


FIG. 5. (a) Unrooted phylogenetic tree based on 2,278 unambiguously aligned nucleotides from the entire gp160 gene from 13 new HIV-1 isolates and reference strains representing different genetic subtypes. Analysis was performed as described in Materials and Methods; the reference strains used in the phylogenetic tree analysis were A-KE.Q2317, A-SE.SOSE7253, A-92UG037, CRF02-AG-IBNG, CRF02-AG-DJ263, CRF02-AG-DJ264, B-WEAU160, B-JRFL, B-HBX2, C-ETH2220, C-IN.21068, C-BW.96BW0502, D-NDK, D-ZR.84ZR085, D-94UG114, CRF01-AE-90CR402, CRF01-AE-93TH253, CRF01-AE-CM240, F1-93BR020, F1-BE.VI850, F1-FLFIN6393, F2-95CMMP255, F2-95CMMP257, G-BE.DRCBL, G-HH8793, G-SE6165, H-90CF056, H-BE.VI991, H-BE.VI997, J-SE.SE91733, J-SE.SE92809, K-96CMMP535, and K-97ZRBQTB11. (b) Putative recombination breakpoints within the gp160 gene were localized based on DIVERT and bootscan analysis (see Materials and Methods) using the reference strains listed above. The results were confirmed by phylogenetic tree analysis with 1,000 bootstrap replicates. Subtypes were designated when the bootstrap values were above 800, divergent subtypes were designated when bootstrap values were between 650 and 800, and fragments were identified as unclassified when they did not cluster at all with any of the known subtypes.

subtypes F, J, and K and the CRF01\_AE strains represent 2 to 4% of the samples; only one subtype B strain was identified. The highest prevalences of subtype C were seen in Mbuyi-Mayi, Kasai Province, well known for its diamond mines. Strikingly, subtype C is predominant in all southern African countries and also in Zambia, which borders the DRC on the south (18). In our study the highest prevalences of CRF01\_AE were seen in the north, which is concordant with previous studies reporting CRF01\_AE in the Central African Republic (34), which borders the DRC on the north.

Despite the predominance of subtype A, the overall prevalence of 51.4% is generally lower than in the surrounding countries. Only in the north are subtype A prevalences documented for the Central African Republic comparable to those observed in northern Equateur Province, where comparable subtypes also cocirculate (34). In countries bordering on the east, subtype A prevalences range from 37% in Tanzania to 57% in Uganda and 70% in Kenya (37, 46, 47). But the important difference is that in these latter countries only three subtypes cocirculate: mainly A and D in Uganda and Kenya; and equal proportions of A, C, and D in Tanzania. As previously mentioned, in the southern African countries the predominant strains responsible for the AIDS epidemic belong to subtype C, representing more than 90% of HIV-1 infections (18, 61). No data are available for Angola, and only sporadic strains of HIV-1 from Congo-Brazzaville have been genetically characterized. But in other west central African countries, known for the presence of many HIV variants, the subtype A prevalences are significantly higher, 70 to 80% in Cameroon and Gabon (32, 56). We recently documented that in west and west central Africa, the majority of the *env* and *gag* subtype A viruses are in fact CRF02\_AG (IbNg)-like viruses, which is apparently not the case in the DRC, where only a minority of the *env* or *gag* subtype A viruses cluster with the IbNg strain (32).

In addition to the high numbers of subtypes that cocirculate, the intrasubtype variability is relatively high among the V3-V5 sequences. This was also observed by Mokili and coworkers, who genetically characterized HIV-1-positive samples from Kimpese (western part of the DRC, 225 km away from Kinshasa) in the *gag* p17 region (31). The *env* distances within subtypes are greater in the DRC than those observed in a similar study conducted by our group in Nigeria; for subtype A, mean intrasubtype distances are 16.6% in the DRC and 10.6% in Nigeria. Similarly, subtype G intrasubtype distances were 14.3% in the DRC and 8.3% in Nigeria (43). The history of the AIDS epidemic is different for the two countries; rapidly increasing seroprevalences over time in Nigeria, versus stable and low prevalences in the DRC (12, 33).

This high intrasubtype diversity is the most probable reason for the low efficiency of simple and more rapid techniques for genetic subtyping. Indeed, in western Africa, 80% of the subtype A samples can be detected with the specific subtype A primers (42), and HMA was able to identify the genetic subtype for more than 90% of the samples (43, 58). In the DRC, only 45% of the subtype A samples could be detected with the subtype A-specific primers, and less than 50% could be identified by HMA.

For 29% of the samples, discordant subtypes between *env* and *gag* were observed, with 11 different discordant profiles involving almost all cocirculating subtypes. This number of discordant samples is higher than what we observed in Senegal, Cameroon, and Gabon, where about 10% of the samples studied had different subtype designations between the two genomic regions (32). The exception was Nigeria, where we documented more than 35% discordance (43). Another major

difference is that in these latter countries, subtypes A and G are predominantly involved in the discordant *gag* and *env* samples, representing 85% of the cases, versus 56% in the DRC (32, 43).

In addition, sequence analysis of the entire envelope from 13 samples confirmed the high degree of diversity and complexity of HIV-1 isolates in the DRC. Of these 13 samples, 6 did not cluster with any of the known subtypes in the V3-V5 region, 4 had a complex recombinant structure in gp160, involving fragments of known and unknown subtypes, and 2 had almost entirely unknown envelope sequences. For 7 of the 13 samples, the genetic subtype could be identified in the V3-V5 region, but 3 of them appeared to have a mosaic gp160 genome. Interestingly, the unknown fragments from the different strains did not cluster together, suggesting the presence of even more subtypes.

Previous reports on genetic subtypes of strains originating in the DRC but isolated from individuals living in Europe suggested a high diversity of HIV-1 in the DRC. For instance, some of the first African HIV-1 isolates to be characterized, MAL and Z321 (obtained from a stored plasma sample obtained in 1976 in a rural area in the northern part of the DRC), have been identified as complex recombinants (8, 16, 55). The presence of recombinant viruses early in the AIDS epidemic suggests that HIV had been present for a while in this region of Africa. Moreover, the first HIV-1 group M sequence from the DRC was documented for a plasma sample from 1959 (62), indicating that group M viruses had been present for more than 40 years in this country. Reports on early AIDS cases confirm also that AIDS is an old disease in Central Africa (54).

The genetic diversity observed in the DRC represents a real challenge for future vaccine development, as well as for efficiency of antiretroviral treatment and diagnostic tests. At least 10 different subtypes cocirculate, and the high intrasubtype distances suggest the presence of subclades. Full-length sequencing will be necessary to determine the extent to which these viruses, especially the tentative subclusters within subtype A, form a subclade or another circulating recombinant form. Full-length sequencing is also required to identify the extent to which unclassified samples represent new subtypes or recombinants.

Interestingly, together with this high genetic diversity in the DRC, the HIV seroprevalence is low and stable. Very soon after the recognition of the global pandemic in the world, a stable seroprevalence was observed in a rural area in Equateur Province between 1976 and 1986 (40). Similarly, the prevalences reported in Lubumbashi, in southern DRC, were in strong contrast with the rapidly increasing and much higher rates in neighboring Zambia and other East African countries (28). Our seroepidemiologic survey conducted in 1997 confirmed also that there were no substantial changes in HIV infection rates in Kinshasa (33). The low and stable HIV prevalence cannot be attributed to any preventive activities, because the health care system and health education programs have declined rapidly since the early 1990s due to decrease in funding and the withdrawal of international cooperation. Importantly, our survey was performed in 1997, before the outbreak of the civil war and thus before the arrival in the DRC of military troops from at least 10 different African countries. It will be important to study the impact of these events on HIV prevalences and on the genetic subtype distribution in the DRC.

Overall, the high number of HIV-1 subtypes cocirculating, the high intrasubtype diversity, and the high numbers of possible recombinant viruses as well as different unclassified strains observed in our study are all consistent with an old and



mature epidemic in Central Africa, more particularly in the DRC, suggesting that this region is the epicenter for HIV-1 group M viruses. Recently computer analysis of various HIV-1 isolates dated the origin of HIV-1 group M viruses to between 1914 and 1941 (24). Together, these virological and epidemiological data suggest that the iatrogenic introduction of HIV into humans through SIVcpz contamination of oral polio vaccines is unlikely (19). The HIV-1 group M subtypes were established before the vaccination programs, which would imply at least as many introductions by oral polio vaccines as there are different group M subtypes.

#### ACKNOWLEDGMENTS

We express our gratitude to the Ministry of Health, National AIDS/STD program, for permission to perform this survey, and especially to B. Kapita, head of the National Ethical Committee. We also thank the following individuals for assistance with field work, for logistical support, and for kind cooperation: B. Edidi, M. Minlangu, T. Tschimpaka, L. Kambembo, L. Atibu, N. Mama, M. Mbete, M. Uwondwa, M. Kazadi, M. Kity, M. Mundele, and the directors and staff of St. Josephs Hospital in Kinshasa, the Sanatorium in Kinshasa, Projet SIDA Laboratory in Kinshasa, the Kabila hospital and STD clinic in Kinshasa, Mama Bobi Ladawa and Bonzola hospital in Mbuyi-Maya, St. Joseph clinic in Mbuyi-Maya, and the infectious disease hospital CDI in Bwamanda.

This study was sponsored by a grant from the Agence National de Recherche sur le SIDA, France (Projet SIDAK).

#### REFERENCES

- Alizon, M., S. Wain-Hobson, L. Montagnier, and P. Sonigo. 1986. Genetic variability of the AIDS virus: nucleotide sequence analysis of two isolates from African patients. *Cell* 46:63-74.
- Apetrei, C., I. Loussert-Ajaka, D. Descamps, F. Damond, S. Saragosti, F. Brun-Vezinet, and F. Simon. 1996. Lack of screening test sensitivity during HIV-1 non-subtype B seroconversions. *AIDS* 10:F57-F60.
- Apetrei, C., D. Descamps, G. Collin, I. Loussert-Ajaka, F. Damond, M. Duca, F. Simon, and F. Brun-Vezinet. 1998. Human immunodeficiency virus type 1 subtype F reverse transcriptase sequence and drug susceptibility. *J. Virol.* 72:3534-3538.
- Carr, J. K., M. O. Salminen, C. Koch, D. Gotte, A. W. Arntstein, P. A. Hegerich, D. St. Louis, D. S. Burke, and F. E. McCutchan. 1996. Full-length sequence and mosaic structure of a human immunodeficiency virus type 1 isolate from Thailand. *J. Virol.* 70:5935-5943.
- Carr, J. K., B. T. Foley, T. Leitner, M. O. Salminen, B. Korber, and F. E. McCutchan. 1998. Reference sequences representing the principal genetic diversity of HIV-1 in the pandemic, p. 10-19. *In* B. Korber, C. L. Kuiken, B. Foley, B. Hahn, F. McCutchan, J. W. Mellors, and J. Sodroski (ed.), *Human retroviruses and AIDS: a compilation and analysis of nucleic acid and amino acid sequences*. Los Alamos National Laboratory, Los Alamos, N.Mex.
- Carr, J. K., M. O. Salminen, J. Albert, E. Sanders-Buell, D. Gotte, D. L. Birx, and F. E. McCutchan. 1998. Full genome sequences of human immunodeficiency virus type 1 subtypes G and A/G recombinants. *Virology* 247:22-31.
- Charneau, P., A. M. Borman, D. Quillent, S. Guetard, S. Chamaret, J. Cohen, G. Remy, L. Montagnier, and F. Clavel. 1994. Isolation and envelope sequence of a highly divergent HIV-1 isolate: definition of a new HIV-1 group. *Virology* 205:247-253.
- Choi, D. J., S. Dube, T. P. Spicer, H. B. Slade, F. C. Jensen, and B. J. Polesz. 1997. HIV type 1 isolate Z321, the strain used to make a therapeutic HIV type 1 immunogen, is intersubtype recombinant. *AIDS Res. Hum. Retroviruses* 13:357-361.
- Delwart, E. L., E. G. Shpaer, J. Louwagie, F. E. McCutchan, M. Grez, H. Rubsamen-Waigmann, and J. I. Mullins. 1993. Genetic relationships determined by a DNA heteroduplex mobility assay: analysis of HIV-1 env genes. *Science* 262:1257-1261.
- Descamps, D., G. Collin, F. Letourneur, C. Apetrei, F. Damond, I. Loussert-Ajaka, F. Simon, S. Saragosti, and F. Brun-Vezinet. 1997. Susceptibility of human immunodeficiency virus type 1 group O isolates to antiretroviral agents: in vitro phenotypic and genotypic analyses. *J. Virol.* 71:8893-8898.
- Descamps, D., C. Apetrei, G. Collin, F. Damond, F. Simon, and F. Brun-Vezinet. 1998. Naturally occurring decreased susceptibility of HIV-1 subtype G to protease inhibitors. *AIDS* 12:1109-1101.
- Esu-Williams, E., C. Mulanga-Kabeya, H. Takema, A. Zwandor, K. Aminu, I. Adamu, O. Yetunde, I. Akinsete, D. Patrel, M. Peeters, and E. Delaporte. 1997. Seroprevalence of HIV-1, HIV-2, and HIV-1 group O in Nigeria: evidence for a growing increase of HIV infection. *J. Acquir. Immune Defic. Syndr. Hum. Retrovirol.* 16:204-210.
- European Commission and the Joint United Nations Programme on HIV/AIDS. 1997. HIV-1 subtypes: implications for epidemiology, pathogenicity, vaccines and diagnostics. *AIDS* 11:UNAIDS17-UNAIDS36.
- Felsenstein, J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 36:783-791.
- Gao, F., L. Yue, S. Craig, C. L. Thornton, D. L. Robertson, F. E. McCutchan, J. A. Bradac, P. M. Sharp, and B. H. Hahn. 1994. Genetic variation of HIV type 1 in four World Health Organization-sponsored vaccine evaluation sites: generation of functional envelope (glycoprotein 160) clones representative of sequence subtypes A, B, C, and E. WHO Network for HIV Isolation and Characterization. *AIDS Res. Hum. Retroviruses* 10:1359-1368.
- Gao, F., D. L. Robertson, S. G. Morrison, H. Hui, S. Craig, J. Decker, P. N. Fultz, M. Girard, G. M. Shaw, B. H. Hahn, and P. M. Sharp. 1996. The heterosexual human immunodeficiency virus type 1 epidemic in Thailand is caused by an intersubtype (A/E) recombinant of African origin. *J. Virol.* 70:7013-7029.
- Gao, F., D. L. Robertson, C. D. Carruthers, S. G. Morrison, B. Jian, Y. Chen, F. Barre-Sinoussi, M. Girard, A. Srinivasan, A. G. Abimiku, G. M. Shaw, P. M. Sharp, and B. H. Hahn. 1998. A comprehensive panel of near-full-length clones and reference sequences for non-subtype B isolates of human immunodeficiency virus type 1. *J. Virol.* 72:5680-5698.
- Gao, F., D. L. Robertson, C. D. Carruthers, Y. Li, E. Balles, L. G. Kostrikis, M. O. Salminen, F. Bibollet-Ruche, M. Peeters, D. D. Ho, G. M. Shaw, P. M. Sharp, and B. H. Hahn. 1998. An isolate of human immunodeficiency virus type 1 originally classified as subtype I represents a complex mosaic comprising three different group M subtypes (A, G, and I). *J. Virol.* 72:10234-10241.
- Heyndrickx, L., W. Janssens, L. Zekeng, R. Musonda, S. Anagonou, G. Van der Auwera, S. Coppens, K. Vereecken, K. De Witte, R. Van Rempelbergh, M. Kahindo, L. Morison, F. E. McCutchan, J. K. Carr, J. Albert, M. Essex, J. Goudsmit, B. Asjo, M. Salminen, A. Buve, and G. van der Groen. 2000. Simplified strategy for detection of recombinant human immunodeficiency virus type 1 group M isolates by *gag/env* heteroduplex mobility assay. *J. Virol.* 74:363-370.
- Hooper, E. 1999. *The river. A journey to the origins of AIDS and HIV*. Little, Brown & Co., Boston, Mass.
- Howard, T. M., and S. Rasheed. 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.
- Hu, D. J., A. Buve, J. Baggs, G. van der Groen, and T. J. Dondero. 1999. What role does HIV-1 subtype play in transmission and pathogenesis? An epidemiological perspective. *AIDS* 13:873-881.
- Janssens, W., A. Buve, and J. N. Nkengasong. 1997. The puzzle of HIV-1 subtypes in Africa. *AIDS* 11:705-712.
- Kimura, M. 1980. A simple method for estimating evolutionary rate of base substitution through comparative studies of nucleotide sequences. *J. Mol. Evol.* 16:111-120.
- Korber, B., M. Muldoon, J. Theiler, F. Gao, R. Gupta, A. Lapedes, B. H. Hahn, S. Wolinsky, and T. Battacharya. 2000. Timing the ancestor of the HIV-1 pandemic strains. *Science* 288:1789-1796.
- Kostrikis, L., E. Bagdades, Y. Cao, L. Zhang, D. Dimitriou, and D. Ho. 1995. Genetic analysis of human immunodeficiency virus type 1 strains from patients in Cyprus: identification of a new subtype designated subtype I. *J. Virol.* 69:6122-6130.
- Laukkanen, T., J. Albert, K. Liitsola, S. D. Green, J. K. Carr, T. Leitner, F. E. McCutchan, and M. O. Salminen. 1999. Virtually full-length sequences of HIV type 1 subtype J reference strains. *AIDS Res. Hum. Retroviruses* 15:293-297.
- Liitsola, K., I. Tashkinova, T. Laukkanen, G. Korovina, T. Smolskaja, O. Momot, N. Mashkilevson, S. Chaplinskias, H. Brummer-Korvenkontio, J. Vanhatalo, P. Leinikki, and M. O. Salminen. 1998. HIV-1 genetic subtype A/B recombinant strain causing an explosive epidemic in injecting drug users in Kaliningrad. *AIDS* 12:1907-1919.
- Loussert-Ajaka, I., T. D. Ly, M. L. Chaux, D. Ingrand, S. Saragosti, A. M. Courroue, F. Brun-Vezinet, and F. Simon. 1994. HIV-1/HIV-2 seronegativity in HIV-1 subtype O infected patients. *Lancet* 343:1393-1394.
- Magazani, K., G. Laleman, J. H. Perriens, K. Kizonde, K. Mukendi, M. Mpungu, N. Badinga, and P. Piot. 1993. Low and stable HIV seroprevalence in pregnant women in Shaba province, Zaire. *J. Acquir. Immune Defic. Syndr.* 6:419-423.
- Mann, J. M., N. Nzilambi, P. Piot, N. Bosenge, M. Kalala, H. Francis, R. C. Colebunders, P. K. Azila, J. W. Curran, and T. C. Quinn. 1998. HIV infection and associated risk factors in female prostitutes in Kinshasa, Zaire. *AIDS* 2:249-254.
- Mokili, J. L., C. M. Wade, S. M. Burns, W. A. Cutting, J. M. Bopopi, S. D. Green, J. E. Peutherer, and P. Simmonds. 1999. Genetic heterogeneity of HIV type 1 subtypes in Kimpese, rural Democratic Republic of Congo. *AIDS Res. Hum. Retroviruses* 15:655-664.
- Montavon, C., C. Toure-Kane, F. Liegeois, E. Mpoudi, A. Bourgeois, L. Vergue, J. L. Perret, A. Boumah, E. Saman, S. Mboup, E. Delaporte, and M. Peeters. 2000. 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. Acquir. Immune Defic. Syndr. 23:363-374.
33. Mulanga-Kabeya, C., N. Nzilambi, B. Edidi, M. Minlangu, T. Tshimpaka, L. Kambembo, L. Atibu, N. Mama, W. Ilunga, H. Sema, K. Tshimanga, B. Bongo, M. Peeters, and E. Delaporte. 1998. Evidence of stable HIV prevalences in selected populations in the Democratic Republic of the Congo. *AIDS* 12:905-910.
  34. Muller-Trutwin, M. C., M. L. Chaix, F. Letourneur, E. Begaud, D. Beaumont, A. Deslandres, B. You, J. Morvan, C. Mathiot, F. Barre-Sinoussi, and S. Saragosti. 1999. Increase of HIV-1 subtype A in Central African Republic. *J. Acquir. Immune Defic. Syndr.* 21:164-171.
  35. Myers, G., B. Korber, B. Hahn, K. T. Jeang, J. Mellors, F. E. McCutchan, L. Henderson, and G. Pavlakis. 1995. Human retroviruses and AIDS: a compilation and analysis of nucleic acid and amino acid sequences. Los Alamos National Laboratory, Los Alamos, N.Mex.
  36. Nasioulas, G., D. Paraskevsi, E. Magiorkinis, M. Theodoridou, and A. Hatzakis. 1999. Molecular analysis of the full-length genome of HIV type 1 subtype I: evidence of A/G/I recombination. *AIDS Res. Hum. Retroviruses* 15:745-758.
  37. Neilson, J. R., G. C. John, J. K. Carr, P. Lewis, J. K. Kreiss, S. Jackson, R. W. Nduati, D. Mbori-Ngacha, D. D. Panteleeff, S. Bodrug, C. Giachetti, M. A. Bott, B. A. Richardson, J. Bwayo, J. Ndinya-Achola, and J. Overbaugh. 1999. Subtypes of human immunodeficiency virus type 1 and disease stage among women in Nairobi, Kenya. *J. Virol.* 73:4393-4403.
  38. Nelson, A. M., S. E. Hassig, M. Kayembe, L. Okonda, C. Mulanga, C. Brown, K. Kayembe, M. M. Kalengayi, and F. G. Mullick. 1991. HIV seropositivity and mortality at the University Hospital, Kinshasa, Zaire. *AIDS* 5:583-586.
  39. Ngaly, B., S. Bertozzi, and R. Ryder. 1990. Obstacles in the optimal management of HIV infection/AIDS in Africa. *J. Acquir. Immune Defic. Syndr.* 3:430-437.
  40. Nzilambi, N., K. M. De Cock, D. N. Forthal, H. Francis, R. W. Ryder, I. Malebe, J. Getchell, M. Laga, P. Plot, and J. B. McCormick. 1988. The prevalence of infection with human immunodeficiency virus over a 10-year period in rural Zaire. *N. Engl. J. Med.* 318:276-279.
  41. Parekh, B., S. Philipps, T. Granade, J. Baggs, D. Hu, and R. Respass. 1999. Impact of HIV type 1 subtype variation on viral RNA quantitation. *AIDS Res. Hum. Retroviruses* 15:133-142.
  42. Peeters, M., F. Liegeois, F. Bibollet-Ruche, D. Patrel, J. L. Perret, N. Vidal, E. Esu-Williams, S. Mboup, E. Mpoudi, N. Nzila, and E. Delaporte. 1998. Subtype-specific polymerase chain reaction for the identification of HIV-1 genetic subtypes circulating in Africa. *AIDS* 12:671-686.
  43. Peeters, M., E. Esu-Williams, L. Vergne, C. Montavon, C. Mulanga-Kabeya, T. Harry, A. Ibrionke, D. Lesage, D. Patrel, and E. Delaporte. 2000. Prevalence of A and G human immunodeficiency viruses type 1 in Nigeria with geographical differences in their distribution. *AIDS Res. Hum. Retroviruses* 16:315-325.
  44. Peeters, M., and P. M. Sharp. Genetic diversity of HIV: the moving target. *AIDS 2000*, a year in review. *AIDS Suppl.*, in press.
  45. Potts, K. E., M. L. Kalish, C. I. Banea, G. M. Orloff, M. St. Louis, C. Brown, N. Malanda, M. Kavuka, G. Schochetman, and C. Y. Ou. 1993. Genetic diversity of human immunodeficiency virus type 1 strains in Kinshasa, Zaire. *AIDS Res. Hum. Retroviruses* 9:613-618.
  46. Rayfield, M. A., R. G. Downing, J. Baggs, D. J. Hu, D. Pieniazek, C. C. Luo, B. Biryahwaho, R. A. Otjen, S. D. Sempala, and T. J. Dondero. 1998. A molecular epidemiologic survey of HIV in Uganda. *AIDS* 12:521-527.
  47. Renjifo, B., P. Gilbert, B. Chaplin, F. Vannberg, D. Mwakagile, G. Msamanga, D. Hunter, W. Fawzi, and M. Essex. 1999. Emerging recombinant human immunodeficiency viruses: uneven representation of the envelope V3 region. *AIDS* 13:1613-1621.
  48. Robertson, D., J. Anderson, J. Bradac, J. K. Carr, B. Foley, R. K. Funkhouser, F. Gao, B. H. Hahn, M. L. Kalish, C. Kuiken, G. H. Learn, T. Leitner, F. E. McCutchan, S. Osmanov, M. Peeters, D. Pieniazek, M. Salminen, P. M. Sharp, S. Wolinsky, and B. Korber. 1999. HIV-1 nomenclature proposal: a reference guide to HIV-1 classification, p. 492-505. In B. Korber et al. (ed.), *Human retroviruses and AIDS 1999: a compilation and analysis of nucleic acid and amino acid sequences*. Los Alamos National Laboratory, Los Alamos, N.Mex.
  49. Ryder, R. W., W. Nsa, S. E. Hassig, F. Behets, M. Rayfield, B. Ekungola, A. M. Nelson, U. Mulenda, H. Francis, and K. Mwandagalirwa. 1989. Perinatal transmission of HIV-1 to infants of seropositive women in Zaire. *N. Engl. J. Med.* 320:1637-1642.
  50. Saitou, N., and M. Nei. 1987. The neighbor joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4:406-425.
  51. Salminen, M. O., J. K. Carr, D. L. Robertson, P. Hegerich, D. Gotte, C. Koch, E. Sanders-Buell, F. Gao, P. M. Sharp, B. H. Hahn, D. S. Burke, and F. E. McCutchan. 1997. Evolution and probable transmission of intersubtype recombinant human immunodeficiency virus type 1 in a Zambian couple. *J. Virol.* 71:2647-2655.
  52. Sanders-Buell, E., M. O. Salminen, and F. E. McCutchan. 1995. Sequencing primers for HIV-1, p. 15-21. In G. Myers (ed.), *Human retroviruses and AIDS*, part III. Los Alamos National Laboratory, Los Alamos, N.Mex.
  53. Simon, F., P. Maucere, P. Roques, I. Lousserat-Ajaka, M. C. Muller-Trutwin, S. Saragosti, M. C. Georges-Courbot, F. Barre-Sinoussi, and F. Brun-Vezinet. 1998. Identification of a new human immunodeficiency virus type 1 distinct from group M and group O. *Nat. Med.* 4:1032-1037.
  54. Sonnet, J., J. L. Michaux, F. Zech, J. M. Brucher, M. de Bruyere, and G. Burton. 1987. Early AIDS cases originating from Zaire and Burundi (1962-1976). *Scand. J. Infect. Dis.* 19:511-517.
  55. Srinivasan, A., D. York, D. Butler, R. Jannoun-Nasr, J. Getchell, J. McCormick, C. Y. Ou, G. Myers, T. Smith, and E. Chen. 1989. Molecular characterization of HIV-1 isolated from a serum collected in 1976: nucleotide sequence comparison to recent isolates and generation of hybrid HIV. *AIDS Res. Hum. Retroviruses* 5:121-129.
  56. Takehisa, J., L. Zekeng, E. Ido, I. Mboudjeka, H. Moriyama, T. Miura, M. Yamashita, L. G. Gurtler, M. Hayami, and L. Kaptue. 1998. Various types of HIV mixed infections in Cameroon. *Virology* 245:1-10.
  57. Thompson, J. D., D. G. Higgins, and T. J. Gibson. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22:4673-4680.
  58. Toure-Kane, C., C. Montavon, M. A. Faye, P. M. Gueye, P. S. Sow, I. Ndoye, A. Gaye-Diallo, E. Delaporte, M. Peeters, and S. Mboup. 2000. Identification of all HIV-1 group M subtypes in Senegal, a country with low and stable seroprevalence. *AIDS Res. Hum. Retroviruses* 16:603-609.
  59. Triques, K., A. Bourgeois, N. Vidal, E. Mpoudi-Ngole, C. Mulanga-Kabeya, N. Nzilambi, J. Torimiro, E. Saman, E. Delaporte, and M. Peeters. 2000. Near full-length genome sequencing of divergent African HIV-1 subtype F viruses leads to the identification of a new subtype designated K. *AIDS Res. Hum. Retroviruses* 16:139-151.
  60. van der Groen, G., P. N. Nyambi, E. Beirnaert, D. Davis, K. Fransea, L. Heyndrickx, P. Onda, G. Van der Auwera, and W. Janssens. 1998. Genetic variation of HIV type 1: relevance of interclade variation to vaccine development. *AIDS Res. Hum. Retroviruses* 14:S211-S221.
  61. Van Harmelen, J. H., E. Van der Ryst, A. S. Loubser, D. York, S. Madurai, S. Lyons, R. Wood, and C. Williamson. 1999. A predominantly HIV type 1 subtype C restricted epidemic in South African urban populations. *AIDS Res. Hum. Retroviruses* 15:395-398.
  62. Zhu, T., B. T. Korber, A. J. Nahmias, E. Hooper, P. M. Sharp, and D. D. Ho. 1998. An African HIV-1 sequence from 1959 and implications for the origin of the epidemic. *Nature* 391:594-597.