AIDS RESEARCH AND HUMAN RETROVIRUSES
Volume 17, Number 4, 2001, pp. 361–365
Mary Ann Liebert, Inc.

Sequence Note

Predominance of HIV Type 1 Subtype G among Commercial Sex Workers from Kinshasa, Democratic Republic of Congo

CHUNFU YANG,[1] BIPIN DASH,[1] SHERI L. HANNA,[1] HENRY S. FRANCES,[2] NZILA NZILAMBI,[3] ROBERT C. COLEBUNDERS,[4] MICHAEL ST. LOUIS,[5] THOMAS C. QUINN,[6] THOMAS M. FOLKS,[1] and RENU B. LAL[1]

ABSTRACT

We have investigated the genetic diversity and potential mosaic genomes of HIV-1 during the early part of the HIV-1 epidemic among commercial sex workers (CSWs) in Kinshasa, Democratic Republic of Congo (formerly Zaire). Serologic analysis revealed that 27 (28.7%) of the 94 specimens were seropositive by both peptide and whole-virus lysate EIAs and that 24 were positive by molecular screening assays, using generic primers that can detect all known groups of HIV-1. Phylogenetic analyses of the *gag*^{p24}, C2V3, and gp41 regions of these 24 specimens showed that all were group M; none of them had any evidence of group O, N, or SIVcpz-like sequences. On the basis of *env* sequence analysis, the 24 group M specimens were classified as subtypes G (37.5%), A (21%), F1 (12.5%), CRF01_AE (8%), D (4%), and H (4%); 3 (12.5%) were unclassifiable (U). Similar analysis of the *gag*^{p24} region revealed that the majority of infections were subtype A; however, one-third of the specimens were subtype G. Parallel analysis of *gag*^{p24} and *env* regions revealed discordant subtypes in many specimens that may reflect possible dual and/or recombinant viruses. These data suggest a predominance of subtype G (both pure G and recombinant CRF02_AG) during the early part of the epidemic in Kinshasa. Infections with group N or SIVcpz-like viruses were not present among these CSWs in Kinshasa.

HUMAN IMMUNODEFICIENCY VIRUS TYPE 1 (HIV-1), the etiologic agent for acquired immune deficiency syndrome (AIDS), is characterized by an unusually high degree of genetic variability *in vivo*. Analysis of HIV-1 genome sequences of virus strains from different geographic locales has revealed that HIV-1 can be divided into three groups: M (main), O (outlier), and N (non-M-non-O).[1,2] Group M includes the viruses that dominate the global epidemic, which is further divided into subtypes, subsubtypes, and circulating recombinant forms (CRFs), whereas group O contains a pool of highly divergent but genetically related viruses, with its epicenter restricted to West Central Africa.[3] Group N, the least spreading of the viruses, is represented by only a handful of viruses identified in Cameroon.[4] Thus, genotypic analysis has provided a better understanding of the molecular epidemiology of HIV-1, enabling the detection of emerging HIV-1 variants and improving the tracking of the epidemic worldwide.

The global diversity of the group M viruses is reflected not only by their discrete genetic subtypes and subsubtypes, but by the fact that a large proportion of the virus strains also has CRFs.[1,2] Parallel phylogenetic analysis of different regions of HIV-1 genomes among known group M subtypes, subsubtypes, and unclassified specimens (U) revealed that subtypes, A, B, C, D, G, H, J, and K, and subsubtypes F1 and F2, had nonre-

[1]HIV Immunology and Diagnostics Branch and HIV and Retrovirology Branch, Division of AIDS, STD, and TB Laboratory Research, National Center for Infectious Diseases, Centers for Disease Control and Prevention, Atlanta, Georgia 30333.
[2]Center on AIDS and Other Medical Consequences of Drug Abuse, National Institute on Drug Abuse, National Institutes of Health, Bethesda, Maryland.
[3]Project SIDA, Kinshasa, Democratic Republic of Congo.
[4]Institute of Tropical Medicine, Antwerp 2000, Belgium.
[5]AIDS Program, National Center for HIV, STD, and TB Prevention, CDC, Atlanta, Georgia 30333.
[6]Laboratory of Immunoregulation, National Institute of Allergy and Infectious Diseases, NIH, Bethesda, Maryland 21205.

combinant genomes in some strains, while at least four CRFs have been identified so far: CRF01_AE, CRF02_AG, CRF03_AB, and CRF04_cpx.[1,5–8] It is believed that multiple subtypes cocirculating in the same population will likely yield a higher number of CRF viruses.[5] It is intriguing to note that most of the CRF genomes are highly complex intersubtype recombinants, with some being extremely stable and others having a patchy appearance due to multiple cross-over events. For instance, CRF01_AE is clearly E in *env*, U/A/E in the regulatory regions, and A in *gag* and *pol*, while the CRF04_cpx has an extremely complex mosaic genome structure including subtypes A, G, H, K, and U.[1] While the emergence of these CRF viruses has important implications for both monitoring HIV-1 genetic diversity and developing effective vaccines, little is known about the prevalence of these CRFs in the early part of the HIV-1 epidemic.

In the present study, we studied the possibility that highly divergent CRF viruses existed during the early part of the HIV epidemic in the Democratic Republic of Congo (DRC, formerly Zaire). We used highly sensitive and broadly cross-reactive serologic and molecular detection tools[9–11] to detect and distinguish genetically divergent variants of HIV-1 and SIVcpz alike in commercial sex workers (CSWs) from Kinshasa. We provide evidence indicating the predominance of subtype G (pure G and recombinant CRF02_AG), and potential CRF viruses, during the early part of the HIV epidemic in Kinshasa, DRC.

The 94 blood specimens used in this study were part of the collection amassed by Project SIDA, a multinational AIDS research project conducted in Kinshasa, DRC during the mid-1980s.[12] The Kinshasa Health Department contacted known or suspected female prostitutes working in bars, night clubs, and hotels during September–October 1985. The women were invited to visit a medical clinic established for HIV screening in central Kinshasa. Women were classified as CSWs by the clinic staff on the basis of the criteria of charging money for sexual services, regardless of the number of partners and the length of time in practice. These women were enrolled in the study with informed consent. HIV-1 infection was detected by commercially available enzyme-linked immunosorbent assay (ELISA) and confirmed by Western blot analysis. The detailed demographic and clinical information on these participants has been described elsewhere.[12]

Plasma samples were tested with consensus synthetic peptides derived from the gp41 region of HIV-1 groups M, O, and N, and SIVcpzGAB, as well as with whole-virus lysate-based enzyme immunoassays (EIAs). The 44-mer consensus peptides included WGIKQLQARVLAVERYLKDQQLLGIWGCSGKLICTTAVPWNASW (nucleotides 580–623) for group M, WGIRQLRARLLALETLIQNQQLLNLWGCKGKLVCYTSVKWNRTW for group O, WGIKQLQAKVLAIERYLRDQQILGSLGCSGKTICYTTVPWNETW for group N, and WGVKQLQARLLAVERYLQDQQILGLWGCSGKAVCYTTVPWNNSW for SIVcpzGAB. The procedures for EIAs using the consensus gp41 peptides and whole-virus lysate were essentially the same as described previously.[11]

Viral RNA was extracted from 200 μl of each plasma sample, using a QIAamp viral RNA kit (Qiagen, Valencia, CA). Reverse transcription-polymerase chain reactions (RT-PCRs) were then performed with generic primers gpM-Z and intM-Z, which can amplify all known HIV-1 and SIVcpz viruses,[9,10] as well as with group M-specific primers in the gag^{p24}, protease, and C2V3 regions.[13–15] The primer sequences and protocols have been described in detail elsewhere.[9–11,13–15]

The purified nested PCR products from the gag^{p24}, C2V3, and gp41 regions were used for automated sequencing with a BigDye terminator cycle-sequencing ready reaction kit (PE Applied Biosystems, Foster City, CA) according to the manufacturer's protocol. The sequencing reactions were then run in a 377 DNA sequencer (PE Applied Biosystems). Sequencing was done in both directions. These sequences, along with reference sequences from the HIV sequence database,[1,2] were aligned by the CLUSTAL W (version 1.74) multiple sequence alignment program.[16] The genetic distances between isolates were computed by DNADIST.[17] Phylogenetic trees were constructed after all gaps had been trimmed, using the neighbor-joining method included in the PHYLIP 3.5c package.[17]

To detect possible highly divergent viruses of HIV-1 groups M, O, and N and SIVcpz during the early part of the HIV epidemic, we tested the 94 CSW samples, using a whole-virus lysate and a cocktail of consensus gp41 peptides of groups M, N, and O and SIVcpz as antigens. The results revealed that 27 (28.7%) of the specimens were positive by both EIAs, while 67 (71.3%) were negative (Table 1).

Molecular analysis of the 26 HIV-1-seropositive specimens (RT-PCR amplification was not done on 1 seropositive specimen, 85CD249, because of the limited material available) revealed that 24 (92%) of them were amplified in the gag^{p24}, C2V3, and gp41 regions, with the exception that specimen 85CD319 was not amplified in the C2V3 region (Table 1). The two specimens that could not be amplified in the above-cited three regions also had no amplification by the group M-specific protease or by broadly cross-reactive integrase intM-Z primers.

To gain further insight into possible variants among the seronegative specimens, we tested all 67 HIV-seronegative specimens with highly sensitive group M-specific primers in the protease and gag^{p24} regions, as well as with the broadly cross-reactive primers intM-Z and gpM-Z, which can detect all known HIV-1 groups (M, N, and O) and SIVcpz.[9,10] All 67 specimens were negative in all amplification assays (data not shown).

Phylogenetic analysis of the sequences in the gag^{p24} and *env* (C2V3 and gp41) regions derived from the 24 specimens was performed (Table 1, Fig. 1). Analysis of the env^{gp41} region revealed that 9 (37.5%) were subtype G, 5 (20.8%) were subtype A, 3 (12.5%) were subsubtype F1, 2 (8.3%) were CRF01_AE, and 1 (4%) each were subtypes D and H; 3 (12.5%) specimens remain unclassifiable (Table 1 and Fig. 1A). Likewise, analysis of the C2V3 region also indicated a predominance of subtype G, followed by subtype A. Two specimens (85CD220 and 85CD354) were discordant in subtypes between the C2V3 and gp41 regions and another specimen (85CD319) failed to amplify in the C2V3 region (Table 1, Fig. 1B). The three unclassifiable specimens could not be classified in either region despite the inclusion of newly described subtypes (F2, J, and K) as well as additional unclassifiable sequences from the HIV database (data not shown).[1,2]

We next examined the gag^{p24} region to identify potential mosaic genomes in these specimens. Phylogenetic analysis of this region revealed 11 (45.8%) to be subtype A, 8 (33%) as subtype G, 2 (8%) as subtype F1, and 1 (4%) each as subtypes D, J, and unclassifiable (Table 1, Fig. 1C). Although there was

TABLE 1. SEROLOGY AND SUBTYPE ANALYSIS OF SPECIMENS FROM COMMERCIAL SEX WORKERS, KINSHASA, DRC

		Phylogenetic classification		
			env	
Patient ID	*HIV-1 serology*[a]	gag *(p24)*	*C2V3*	*gp41*
85CD246	+	G	G	G
85CD267	+	G	G	G
85CD273	+	G	G	G
85CD300	+	G	G	G
85CD306	+	G	G	G
85CD349	+	G	G	G
85CD085	+	A	A	A
85CD116	+	A	A	A
85CD355	+	A	A	A
85CD227	+	A	A	A
85CD321	+	A	A	A
85CD241	+	F	F	F
85CD260	+	F	F	F
85CD096	+	D	D	D
Potential dual and/or recombinant forms				
85CD322	+	A	E	E
85CD357	+	A	E	E
85CD350	+	A	G	G
85CD220	+	A	A	G
85CD225	+	A	H	H
85CD354	+	G	A	G
85CD244	+	G	F	F
85CD319	+	J	NA[b]	U
85CD099	+	A	U	U
85CD299	+	U	U	U
85CD266	+	NA	NA	NA
85CD271	+	NA	NA	NA

Abbreviation: U, unclassifiable.

[a]Of the 94 commercial sex worker specimens from Kinshasa, 27 were seropositive and 67 were seronegative. All 67 seronegative specimens were tested by molecular screening assays, using generic primers in the integrase and gp41 regions, as well as group M-specific primers in protease, *gag*^{p24}, and C2V3 regions, and all were negative.

[b]NA, not amplified by RT-PCR.

higher prevalence of subtype A based on *gag*^{p24} region, subtype G still accounted for almost one-third of the infections.

More importantly, we observed discordant subtypes between *gag* and *env* regions (Table 1). Of the 24 specimens, 14 (58.3%) specimens had the same subtype in the *gag*^{p24}, C2V3, and gp41 regions (6 subtype G, 5 subtype A, 2 subtype F1, and 1 subtype D), while the remaining 10 had discordant subtypes. Among the discordant subtypes, we found AE, AG, AH, GF, AAG, and GAG, as well as J and A with unclassifiable C2V3 and gp41 regions. In addition, one specimen could not be classified in all three regions. Whether these discordant subtypes and unclassifiable isolates represent infections with two different subtypes, circulating recombinant forms, or new unidentified viruses remains to be determined.

The detailed molecular analyses of the sequences from the 24 CSWs in Kinshasa revealed a high degree of genetic heterogeneity and potential circulating recombinant forms of group M viruses. We also did not detect infections with HIV-1 groups O and N and SIVcpz-like variants among these specimens, although DRC is known to harbor highly divergent HIV-1 variants.[18] These data are in accordance with findings from studies showing the presence of all known HIV-1 subtypes in DRC and neighboring countries.[18–21]

Analysis of the *gag*^{p24} region revealed that although almost half of infections were subtype A, one-third were due to subtype G. A previous study based on the *gag*^{p17} region also revealed that subtype A was the predominant subtype in DRC.[18] Unlike Cameroon, where most subtype A sequences sequenced so far cluster closely to IbNG-like circulating recombinant forms (CRF02_AG),[1,19,21,22] the subtype A from DRC clustered closely with prototypic subtype A. It is believed that CRF02_AG forms are more common than the prototypic subtype A in Cameroon, Nigeria, and West Africa.[7,22]

Parallel analyses of *gag* and *env* regions suggested that while more than half of the isolates appeared to be noncirculating recombinant forms, the remaining sequences represented potential circulating recombinant forms with discordant classifications of the genomes in these two regions. While this study has

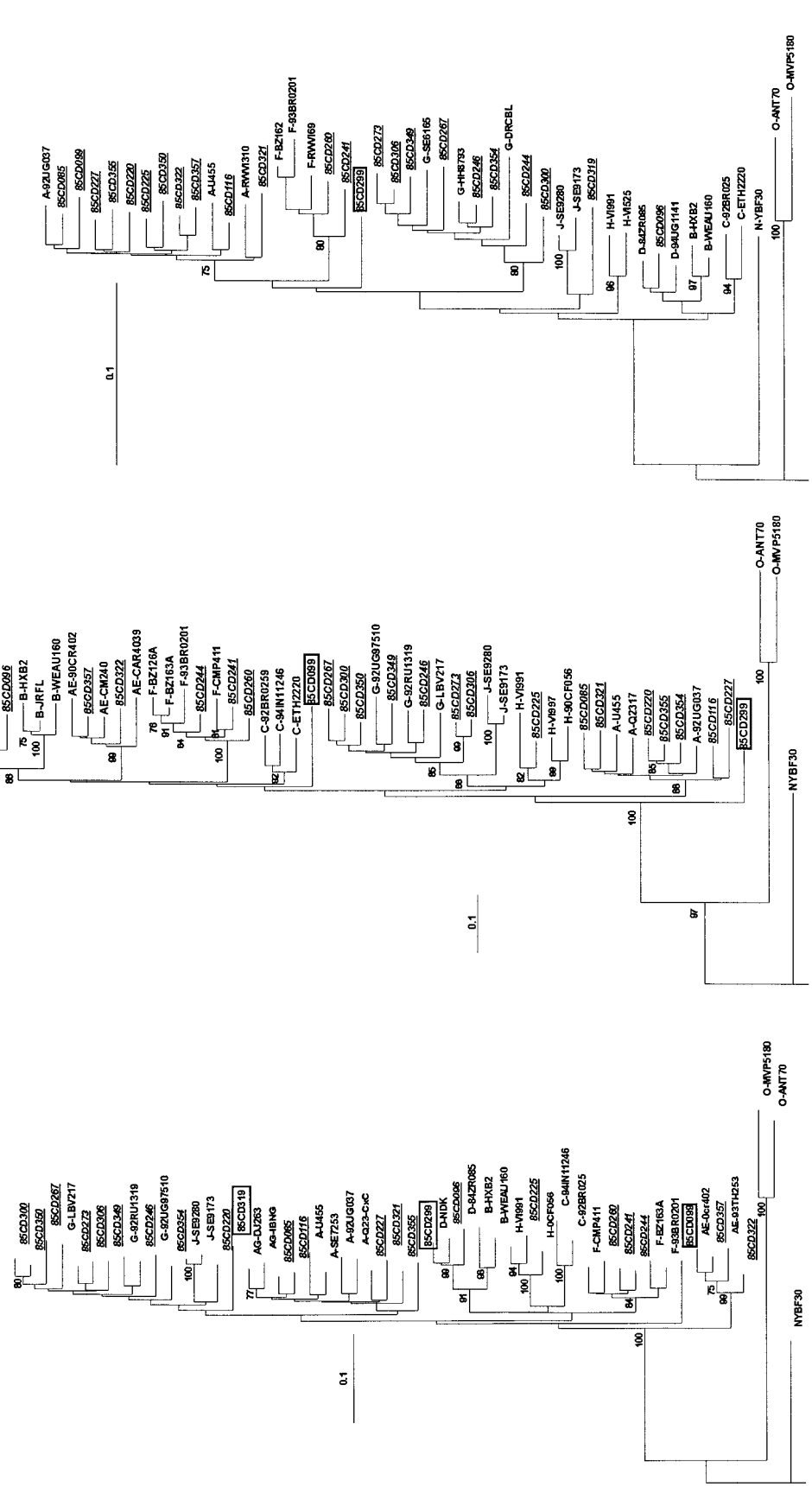

FIG. 1. Phylogenetic classification of the newly characterized HIV-1 group M sequences (85CD#) along with references from the HIV sequence database[2] in *env*^{gp41} (**A**), C2V3 (**B**), and *gag*^{p24} (**C**) regions. The trees were constructed by the neighbor-joining method, using the consensus sequences of gp41 (312 nt), C2V3 (365 nt), and p24 (426 nt). The trees were rooted by SIVcpzGAB. Horizontal branch lengths are drawn to scale, with the bar indicating 0.1 nucleotide substitution per site. Numbers at the nodes indicate the percentage of bootstrap sampling (out of 100) in which the cluster to the right is supported (only values ≥75% are shown). Underlined sequences represent new sequences from the current study, and boxed sequences represent unclassifiable specimens. GenBank accession numbers for the newly characterized sequences are AF260437–AF260507.

not determined whether these discordant genomes represent the true circulating recombinant forms or simply reflect dual infections with two viral subtypes, it is surprising to see such a high prevalence of discordant genomes in the early part of the HIV epidemic. Among these discordant genome structures, some were similar to those previously described (i.e., AE, AG, and AH) while others were unique and/or unclassifiable (G/F and J/U).[1,2] These unique unclassifiable structures were also observed previously in Kimpese, DRC and Cameroon.[18,21] We also observed an unusually high prevalence of subtypes G (potential pure G and circulating recombinant forms A/G, G/F, and G/U). While the dynamic role played by the recombinant G viruses remains to be determined, a study based on the *pol*-integrase region revealed that 7 of 26 infections in the Republic of Congo were subtype G, and all these subtype G specimens represented pure G infections when analyzed in the *env* region.[19] While substantial numbers of subtype G sequences have recombined with subtype A in Africa, data presented here and those by Mboudjeka *et al.*[19] provide strong evidence that pure subtype G continued to spread as an integral structure through Central Africa during the early part of the epidemic. The other recombinant structures such as G/F or G/unclassifiable sequences may not have had a survival advantage and may have resulted in dead-end infections.

In summary, we have used sensitive primers to confirm great genetic diversity and a high proportion of potential recombinant genomes in HIV-1 infections from persons infected in the early part of the HIV pandemic in Kinshasa, DRC, with a predominance of subtype G among the CSWs studied. This study also indicates that infections with group N or SIVcpZ-like viruses were uncommon or did not exist in the early part of the HIV epidemic. Continued studies related to molecular epidemiology, as well as genetic diversity and recombination events within HIV-1, are needed to understand their spread and impact on the epidemic.

REFERENCES

1. Robertson DL, Anderson JP, Bradac JA, *et al.*: HIV-1 nomenclature proposal. In: *Human Retroviruses and AIDS: A Compilation and Analysis of Nucleic Acid and Amino Acid Sequences* (Kuiken CL, Foley B, Hahn B, *et al.*, eds.). Los Alamos National Laboratory, Los Alamos, New Mexico, 1999, pp. 492–505.
2. Carr JK, Foley BT, Leitner T, Salminen M, Korber B, and McCutchan F: Reference sequences representing the principal genetic diversity of HIV-1 in the pandemic. In: *Human Retroviruses and AIDS: A Compilation and Analysis of Nucleic Acid and Amino Acid Sequences* (Korber B, Foley B, McCutchan F, *et al.*, eds.). Los Alamos National Laboratory, Los Alamos, New Mexico, 1998, pp. III10–III19.
3. Peeters M, Gueye A, Mboup S, *et al.*: Geographical distribution of HIV-1 group O viruses in Africa. AIDS 1997;11:493–498.
4. 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;4:1032–1037.
5. Robertson DL, Gao F, Hahn BH, and Sharp PM: Intersubtype recombinant HIV-1 sequences. In: *Human Retroviruses and AIDS: A Compilation and Analysis of Nucleic Acid and Amino Acid Sequences* (Korber B, Hahn B, Foley B, *et al.*, eds.). Theoretical Biology and Biophysics Group, Los Alamos National Laboratory, Los Alamos, New Mexico, 1997, pp. III25–III30.
6. 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;72:5680–5688.
7. Carr J, Salminen KO, Albert J, *et al.*: Full genome sequences of human immunodeficiency virus type 1 subtypes G and A/G intersubtype recombinants. Virology 1998;247:22–31.
8. Gao F, Robertson CL, Carruthers CD, *et al.*: An isolate of human immunodeficiency virus type 1 originally classified as subtype I represents a complex mosaic comprising three different group M subtypes (A, G, and I). J Virol 1998;72:10234–10241.
9. Yang C, Pieniazek D, Owen SM, *et al.*: Detection of phylogenetically diverse human immunodeficiency virus type 1 groups M and O from plasma using highly sensitive and specific generic primers. J Clin Microbiol 1999;37:2581–2586.
10. Yang C, Dash BC, Simon F, *et al.*: Detection of diverse variants of HIV-1 groups M, N, and O and simian immunodeficiency viruses from chimpanzees using generic *pol* and *env* primer pairs. J Infect Dis 2000;181:1791–1795.
11. Masciotra S, Rudolph DL, van der Groen G, Yang C, and Lal RB: Serological detection of infection with diverse HIV-1 and SIVcpz retroviruses using consensus Env peptides. Clin Diagn Lab Immunol 2000;7:706–709.
12. Mann JM, Nzilambi N, Piot P, *et al.*: HIV infection and associated risk factors in female prostitutes in Kinshasa, Zaire. AIDS 1988; 2:249–254.
13. Brennan CA, Laud JK, Golden A, *et al.*: Serologic and phylogenetic characterization of HIV-1 subtypes in Uganda. AIDS 1997; 11:1823–1832.
14. Janini LM, Pieniazek D, Peralta JM, *et al.*: Identification of single and dual infections with distinct subtypes of human immunodeficiency virus type 1 by using restriction fragment length polymorphism analysis. Virus Genes 1996;13:69–81.
15. Schochetman G, Subbarao S, and Kalish ML: Methods for studying genetic variation of the human immunodeficiency virus (HIV). In: *Viral Genome Methods* (Adolph KW, ed.). CRC Press, New York, 1996, pp. 25–41.
16. Thompson JD, Higgins DG, and 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;22:4673–4680.
17. Felsenstein J: *PHYLIP* (*Phylogeny Inference Package*), version 3.5c. Department of Genetics, University of Washington, Seattle, Washington, 1993. [Distributed by the author]
18. Mokili JLK, Wade CM, Burns SM, *et al.*: Genetic heterogeneity of HIV type 1 subtypes in Kimpese, rural Democratic Republic of Congo. AIDS Res Hum Retroviruses 1999;15:655–664.
19. Mboudjeka I, Bikandou B, Zekeng L, *et al.*: Genetic diversity of HIV-1 group M from Cameroon and Republic of Congo. Arch Virol 1999;144:2291–2311.
20. Bikandou B, Takehisa J, Mboudjeka I, *et al.*: Genetic subtypes of HIV type I in Republic of Congo. AIDS Res Hum Retroviruses 2000;16:613–620.
21. Fonjungo P, Dash B, Mpouti M, *et al.*: Presence of diverse human immunodeficiency virus type I viral variants in Cameroon. AIDS Res Hum Retroviruses 2000;6:1319–1324.
22. Car JK, Laukkanen T, Salminen MO, *et al.*: Characterization of subtype A HIV-1 from Africa by full genome sequencing. AIDS 1999;13:1819–1826.

Address reprint requests to:
Renu B. Lal
HIV Immunology and Diagnostics Branch
DASTLR, NCID, CDC, Mail Stop D-12
1600 Clifton Road
Atlanta, Georgia 30333

E-mail: rbl3@cdc.gov