

## Sequence Note

# Phylogenetic Analysis of Protease and Transmembrane Region of HIV Type 1 Group O

CHUNFU YANG,<sup>1</sup> FENG GAO,<sup>2</sup> PETER N. FONJUNGO,<sup>1</sup> LEOPOLD ZEKENG,<sup>3</sup>  
GUIDO VAN DER GROEN,<sup>4</sup> DANUTA PIENIAZEK,<sup>1</sup> CHARLES SCHABLE,<sup>1</sup> and RENU B. LAL<sup>1</sup>

### ABSTRACT

The molecular diversity and phylogenetic relationship of 22 HIV-1 group O strains were examined on the basis of the protease gene and the N-terminal region of gp41<sup>env</sup>. Analysis of the newly characterized protease sequences with 12 reference sequences revealed no specific clustering patterns, despite the distinct geographic locations of the specimens. In contrast, analysis of the newly sequenced gp41 sequences with 34 published sequences revealed two distinct clusters, each represented by one full-length sequence (MVP5180 and ANT-70). Further, four of the specimens classified as group O in the protease region clustered with group M in the gp41 region (three subtype A and one subtype G, respectively), suggesting dual and/or recombinant infections with HIV-1 groups M and O. The presence of two distinct clusters in the gp41 region indicates at least two possible subtypes within group O viruses, and this may provide useful information regarding molecular epidemiological studies of group O infections.

**P**HYLOGENETIC ANALYSIS of human immunodeficiency virus type 1 (HIV-1) sequences obtained from infected persons throughout the world has identified three genetically distinct groups designated as major (M), outlier (O), and non-M and non-O (N).<sup>1,2</sup> HIV-1 group M, which includes eight subtypes (A, B, C, D, F, G, H, and J) and four circulating recombinant forms (CRF) (AB, AE, AG, and AGI), is the most prevalent and reveals striking geographic clustering patterns.<sup>1</sup> While subtype B is the most predominant subtype in Europe and North America, non-B subtypes and CRFs are more common in developing countries.<sup>1</sup> In contrast, HIV-1 group O contains a pool of highly divergent, but genetically related viruses with its epicenter restricted to Central Africa (Cameroon, Gabon, and Equatorial Guinea).<sup>2,3</sup> Sporadic infections have been detected in several other parts of the world, including Western Europe and the United States; however, most of these patients had contacts to West Central Africa.<sup>4–7</sup> Although the significance of subtype and group designations on the epidemiological, clinical, and diagnostic characteristics of HIV-1 is not fully under-

stood, the identification of persons infected with HIV-1 group O has been of considerable interest, since early screening assays did not consistently detect serologic responses from persons infected with group O viruses.<sup>8–10</sup>

While previous attempts to define phylogenetic clusters within group O revealed no clustering pattern,<sup>4,11–13</sup> more recent studies have proposed two clusters based on two different gene regions.<sup>14–16</sup> In the present investigation, we undertook a detailed analysis of 22 HIV-1 group O specimens obtained from Cameroon, Spain, and the United States using the protease and gp41 regions to better understand the phylogenetic relationship of group O viruses and their genetic variations. We provide evidence of two distinct phylogenetic clusters of HIV-1 group O isolates. In addition, the results suggest potential mosaic genomes and/or dual infections of groups M and O in four of the specimens.

Plasma samples from 22 persons infected with HIV-1 group O viruses (16 from Cameroon, 4 from Spain, and 2 from the United States) were collected during 1997–1998 (Table 1). The

<sup>1</sup>HIV Immunology and Diagnostics Branch, Division of AIDS, STD, and TB Laboratory Research, NCID, Atlanta, Georgia 30333.

<sup>2</sup>Departments of Medicine and Microbiology, University of Alabama at Birmingham, Birmingham, Alabama 35294.

<sup>3</sup>Laboratoire de Sante Hygiene Mobile, Yaoundé, Cameroon.

<sup>4</sup>Division of Microbiology, Institute of Tropical Medicine, Antwerp, Belgium.

TABLE 1. DEMOGRAPHIC AND PATIENT INFORMATION

Country	Patient ID	Age/sex	Symptom	Phylogenetic analysis results			
				Protease	Accession #	gp41	Accession #
Spain <sup>a</sup>	97ES202	39/M	None	UN <sup>b</sup>	—	O	AF229244
	97ES203	37/F	PCP	O	AF229210	O	AF229233
	97ES204	44/M	EC	O	AF229211	O	AF229232
	97ES205	7/M	PCP	O	AF229212	O	AF229242
United States <sup>c</sup>	97US201	42/F	None	O	AF229206	O	AF229234
	97US265	40/F	None	O	AF229218	O	AF229240
Cameroon	98CM00T	32/F	None	O	AF229222	O	AF229241
	97CM408	38/M	None	O	AF229213	O	AF229238
	97CM830	40/F	None	O	AF229215	O	AF229231
	97CM755	39/M	AIDS	O	AF229214	O	AF229239
	97CM747	36/M	AIDS	O	AF229220	O	AF229243
	98CM601	25/F	None	O	AF229217	O	AF229230
	97CM768	36/F	AIDS	O	AF229209	O	AF229237
	97CM766	43/M	None	O	AF229208	O	AF229236
	97CM798	35/M	AIDS	O	AF229207	O	AF229235
	97CM761	53/F	AIDS/TB	O	AF229216	O	AF229229
	98CM139	35/M	None	O	AF229224	O	AF229227
	98CM226	NA <sup>d</sup>	NA	UN <sup>b</sup>	—	O	AF229228
	97CM773	39/F	AIDS	O	AF229219	G	AF229202
	98CM010	26/F	AIDS	O	AF229221	A	AF229203
	98CM094	31/F	None	O	AF229223	A	AF229204
98CM215	NA	NA	O	AF229225	A	AF229205	

Abbreviation: NA, Not available; PCP, *Pneumocystis carinii* pneumonia; EC, esophageal candidiasis.

<sup>a</sup>The specimens 97ES202 and 97ES203 are from Spain, whereas 97ES204 and 97ES205 are from persons born in Guinea Bissau.<sup>15</sup>

<sup>b</sup>UN, Unamplifiable; amplification in the protease region was not successful.

<sup>c</sup>These cases have previously been described in detail.<sup>6,21</sup>

<sup>d</sup>NA: not available.

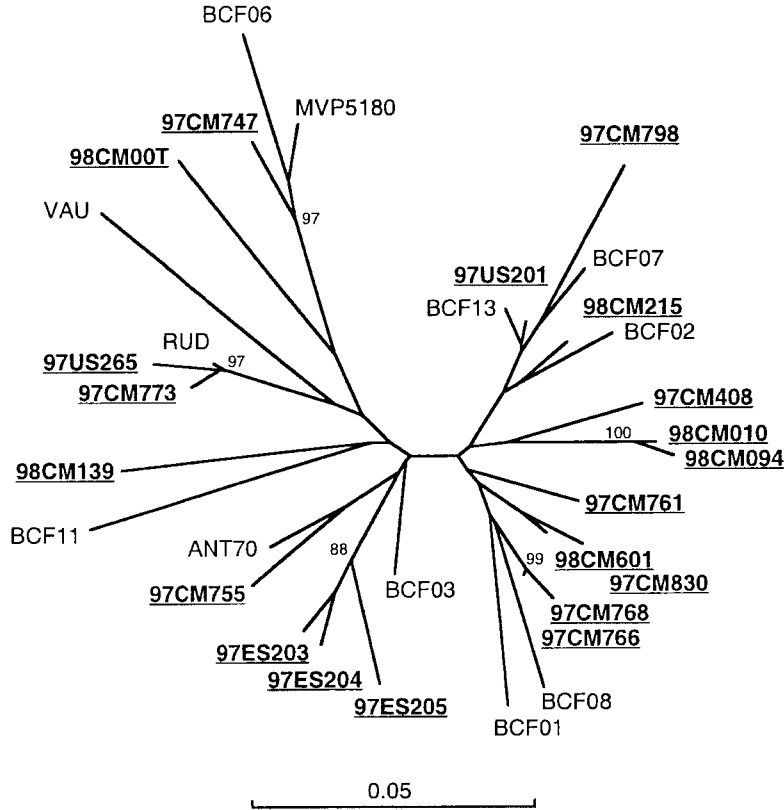
HIV-1 group O infections were confirmed by an anti-HIV-1 gp120 group O-specific peptide enzyme immunoassay (EIA).<sup>10</sup> Eighteen patients had received no antiretroviral therapy, while 4 patients (97ES203, 97ES204, 97ES205, and 97US265) were treated with nucleoside reverse transcriptase inhibitors. Viral RNA was extracted from the plasma samples, and reverse transcriptase-polymerase chain reaction (RT-PCR) and PCR were performed under the conditions and procedures described previously.<sup>17</sup> For protease, the primary PCR primers are CYF1 (5'-TTTTTTAGGGAAATACTGGCCTCC, nucleotides [nt] 2085–2108 based on HXB2, GenBank accession number K03455 [<http://hiv-web.lanl.gov/NUM-HXB2/HXB2.MAIN/html>]) and CYR1 (5'-GGTGTATTATAAGGATTTTCAGG, nt 2703–2725); the nested PCR primers are CYF2 (5'-ACGAGCAGACCCATCAGCCCCACC, nt 2142–2165) and CYR2 (5'-CATTGTTTTACTTTTGGTCCATCCATT, nt 2594–2620). The primary and nested PCR primers for the gp41 region are gp40F1 and gp41R1, and gp46F2 and gp47R2.<sup>17</sup> PCR products were directly used for the sequencing reaction with a BigDye (PE Applied Biosystems, Foster City, CA) terminator cycle sequencing ready reaction kit and the reaction was then run in a DNA sequencer 377 (Applied Biosystems, Foster City, CA). Sequencing was carried out in both direc-

tions. Nucleotide sequences of the protease and gp41 were then aligned by the CLUSTALW multiple sequence alignment program.<sup>18</sup> Phylogenetic trees were constructed by the neighbor-joining method implemented with the PHYLIP 3.2 package<sup>19</sup> after exclusion of gaps. The amino acid sequences of gp41 immunodominant epitopes were also aligned.

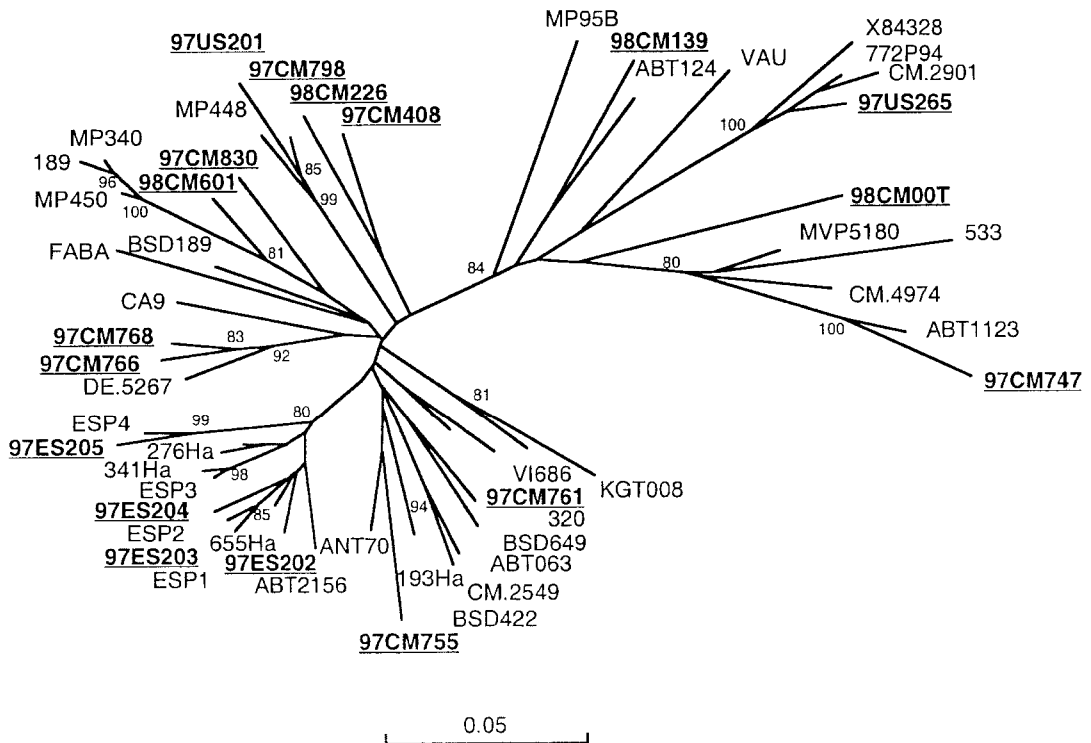
To determine whether HIV-1 group O strains obtained from three different geographic areas can provide distinct phylogenetic clusters, the full-length protease genes of 20 group O strains were phylogenetically analyzed by the neighbor-joining method. Analysis of the protease region failed to identify any well-defined clusters with high bootstrap values, although the overall topology of the tree was suggestive of two distinct clusters (Fig. 1A). Lack of distinct clustering may be due to the length of sequences analyzed, since a larger fragment in the RT region (792 bp) did reveal a distinct cluster, designated subtype A-O.<sup>14</sup> The three sequences from Spanish specimens clustered together with high bootstrap value, and presumably are identical to more recently published sequences.<sup>20</sup> The two U.S. group O specimens were distinct and clustered with the Cameroonian specimens. Taken together, these data support previous observations that each strain of HIV-1 group O represents a distinct strain by itself.<sup>11–13</sup>

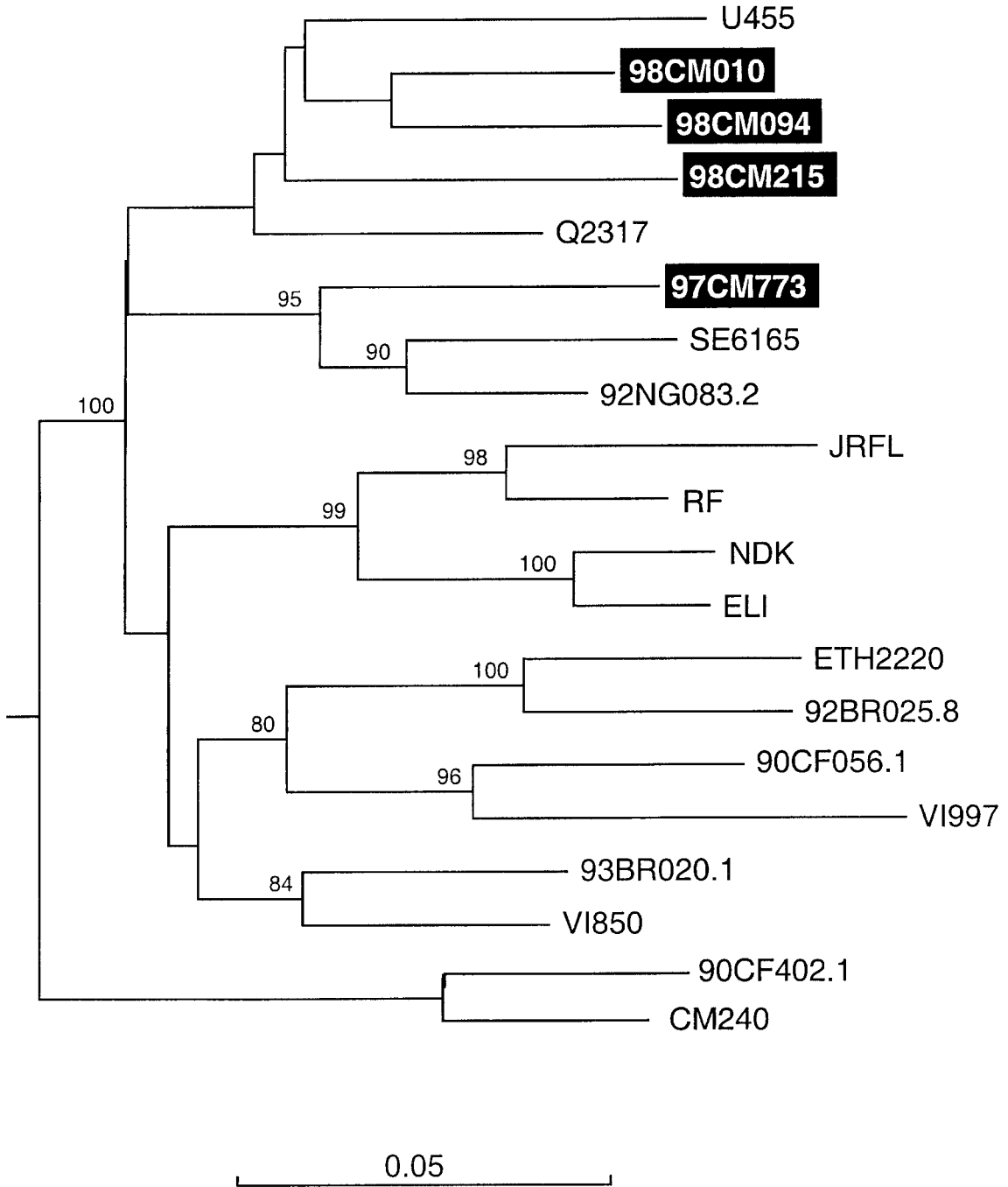
**FIG. 1.** Phylogenetic relationship of the newly sequenced (in boldface, underlined) and published HIV-1 group O sequences in the *pol* (prt; **A**) and *env* (gp41; **B**) regions of group O viruses. Trees were derived from nucleotide sequence alignments of the protease and gp41 regions (consensus lengths of 281 and 261 bp, respectively), using the neighbor-joining method. Numbers at the nodes indicate the percentage of bootstrap values (only values greater than 80% are shown).

**A protease**



**B gp41**





**FIG. 2.** Phylogenetic relationship of newly sequenced (boxes) and previously reported HIV-1 group M sequences in the *env* (gp41) region of group M viruses. The tree was derived by nucleotide sequence alignment of the gp41 region (consensus sequence length, 366 bp), using the neighbor-joining method with SIVcpzGAB as an outgroup. Horizontal branch lengths are drawn to scale, with the bar indicating 0.05 nucleotide substitutions per site. Numbers at the nodes indicate the percentage of bootstrap values of 1000 replicas in which the cluster to the right is supported (only values greater than 80% are shown).

Phylogenetic analysis of the *env* gp41 region of the 22 newly sequenced isolates along with 34 group O sequences from the HIV database revealed that 18 of the 22 isolates clustered with group O isolates (Fig. 1B), whereas the remaining 4 clustered with HIV-1 group M (Fig. 2). Unlike the protease gene analysis, the newly characterized gp41 sequences together with pre-

viously published group O sequences appeared to form two clusters, with one cluster having high bootstrap value (84%) (Fig. 1B). Thus our results corroborate the subcluster designated subtype A-O.<sup>14-16</sup> The gp41 sequences for Spanish and U.S. specimens have previously been published.<sup>15,21</sup> Interestingly, the four remaining isolates from Cameroon that were



identified as group O, based on the protease region, clustered with group M subtypes A (98CM010, 98CM094, and 98CM215) and G (97CM773) in gp41 (Fig. 2). An independent amplification and sequence analysis further confirmed that these specimens are indeed group M in gp41, but group O in protease.

Analysis of the geographic pattern of group O viruses revealed that viruses from persons in Cameroon were scattered in both the protease and gp41 trees. The two U.S. specimens clustered closely with some of the Cameroonian specimens, but were separate from each other, suggesting an independent introduction. In contrast, the four Spanish specimens clustered together tightly with high bootstrap values, and one study has shown that the Spanish patients had a link with Equatorial Guinea, a former Spanish colony.<sup>15</sup> Interestingly, the first known case of AIDS documented in Europe was a Norwegian sailor infected with a group O virus.<sup>20</sup> He developed clinical manifestations of AIDS in 1966, followed by his wife and daughter, all of whom died in the 1970s.<sup>20</sup> Clearly, HIV-1 group O infections are not a new phenomenon; however, group O viruses have not resulted in an epidemic of the magnitude of the current AIDS pandemic caused by HIV-1 group M viruses. Whereas the HIV-1 group M epidemic has resulted in geographically and phylogenetically distinct subtypes, the group O viruses may simply radiate out from an original ancestral virus without forming any well-separated subtypes, possibly because of less efficient replication or transmission.

The amino acid sequence alignment of the immunodominant region of gp41<sup>22</sup> (amino acids 569–688) revealed minor (<20%) substitutions (Fig. 3). Analysis of a cytotoxic T lymphocyte (CTL) epitope within cluster I (ALETLIQNQQLL; amino acids 591–602) revealed substitutions at L<sup>595</sup> → F and I<sup>596</sup> → M. Although the cysteine loop (CKGKLVLC; amino acids 607–613) had greater numbers of substitution, most were conservative (K<sup>610</sup> → R and V<sup>612</sup> → I). Likewise, the ectodomain region (ELDEWA) in B cell cluster II (amino acids 672–677) was highly conserved, with only one substitution in one isolate. However, some regions within gp41 revealed multiple deletions/insertions. For instance, 9 of the 18 group O specimens (2 from the United States, 3 from Spain, and 4 from Cameroon) had a 9-nucleotide deletion that resulted in three amino acid deletions at positions 629–631; the same deletions were also present in the ANT70 strain. Although almost half of Cameroonian specimens published to date also have this deletion,<sup>11,12</sup> the biologic significance of this deletion remains to be determined. Likewise, for most specimens, the number of potential glycosylation sites (NxS/T) was three; however, a few isolates had only two sites and others had four. The amino acid sequences from the four specimens (97CM773, 98CM010, 98CM094, and 98CM215) that were identified as group M, based on the gp41 phylogenetic analysis, indeed were highly conserved and closely related to group M sequences (Fig. 3).

Thus, in addition to providing information that further strengthens a distinct clustering pattern for group O viruses, our results also indicate the presence of group M and O dual and/or recombinant infections. Genotyping analyses revealed that these specimens were group O in the protease region and group M in the gp41 region. Whether such differences in the two regions represent dual infection with group M and group O viruses or the presence of a mosaic genome of groups O and M could not be further examined because of the limited availability of

plasma material. However, it would not be surprising if these four patients were infected with HIV-1 group M/O recombinants, since both mixed infections of groups M and O and intergroup recombinants of groups M and O have been reported in Benin and Cameroon.<sup>23–26</sup> Because recombination between currently circulating lineages of HIV-1 can possibly result in altered biologic properties, continued systematic surveillance is important for monitoring the recombinants.

## ACKNOWLEDGMENTS

We thank Dr. Steve McDougal for critical review of the manuscript and Robin Mosley for editorial assistance.

Use of trade names is for identification only and does not constitute endorsement by the U.S. Department of Health and Human Services, the Public Health Service, or the Centers for Disease Control and Prevention.

## REFERENCES

- 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—1998* (Korber B, Kuiken C, Foley B, Hahn B, McCutchan F, Mellors J, and Sodroski J, eds.). Los Alamos National Laboratory, Los Alamos, New Mexico, 1998, pp. III10–III19.
- Simon F, Maucelere P, Roques P, *et al.*: Identification of a new human immunodeficiency virus type 1 distinct from group M and group O. *Nature Med* 1998;4:1032–1037.
- Peeters M, Gueye A, Mboup S, *et al.*: Geographical distribution of HIV-1 group O viruses in Africa. *AIDS* 1997;11:493–498.
- Loussert-Ajaka I, Chaix M-L, Korber B, *et al.*: Variability of human immunodeficiency virus type 1 group O strains isolated from Cameroonian patients living in France. *J Virol* 1995;69:5640–5649.
- Hampel H, Sawitzky D, Stoffler-Meilicke M, *et al.*: First case of HIV subtype O infection in Germany. *Infection* 1995;23:369–370.
- Rayfield MA, Sullivan P, Bander CI, *et al.*: HIV-1 group O virus identified for the first time in the United States. *Emerg Infect Dis* 1996;2:209–212.
- Soriano V, Gutierrez M, Garcia-Lerma G, *et al.*: First case of HIV-1 group O infection in Spain. *Vox Sang* 1996;71:66.
- Schable C, Zekeng L, Pau C, *et al.*: Sensitivity of United States HIV antibody tests for detection of HIV-1 group O infections. *Lancet* 1994;344:1333–1334.
- Loussert-Ajaka I, Ly TD, Chaix ML, *et al.*: HIV-1/HIV-2 seronegativity in HIV-1 subtype O infected patients. *Lancet* 1994;343:1393–1394.
- Eberle J, Loussert-Ajaka I, Brust S, *et al.*: Diversity of the immunodominant epitope of gp41 of HIV-1 subtype O and its validity for antibody detection. *J Virol Methods* 1997;67:85–91.
- Korber B, Loussert-Ajaka I, Blouin J, and Saragosti S: A comparison of HIV-1 group M and group O: Functional and immunogenic domains in the gag p<sup>24</sup> protein and the C2V3 region of the envelope protein. In: *Human Retroviruses and AIDS—1996* (Myers G, Korber B, Foley B, Jeang K-T, Mellors JW, and Wain-Hobson S, eds.). Los Alamos National Laboratory, Los Alamos, New Mexico, 1996, pp. III41–III56.
- Bibollet-Ruche F, Peeters M, Mboup S, *et al.*: Molecular characterization of the envelope transmembrane glycoprotein of 13 new human immunodeficiency virus type 1 group O strains from six different African countries. *AIDS Res Hum Retroviruses* 1998;14:1281–1285.

13. Bibollet-Ruche F, Loussert-Ajaka I, Simon F, *et al.*: Genetic characterization of accessory genes from human immunodeficiency virus type 1 group O strains. *AIDS Res Hum Retroviruses* 1998;14:951–961.
14. Quinones-Mateu M, Albright JL, Mas A, Soriano V, and Arts EJ: Analysis of *pol* gene heterogeneity, viral quasispecies, and drug resistance in individuals infected with group O strains of human immunodeficiency virus type 1. *J Virol* 1998;72:9002–9015.
15. Mas A, Quinones-Mateu ME, Domingo E, and Soriano V: Phylogeny of HIV type 1 group O isolates based on *env* gene sequences. *AIDS Res Hum Retroviruses* 1999;15:769–773.
16. Janssens W, Heyndrickx L, Van der Auwera G, *et al.*: Interpatient genetic variability of HIV-1 group O. *AIDS* 1999;13:41–48.
17. Yang C, Pieniazek D, Owen SM, *et al.*: Detection of phylogenetically diverse human immunodeficiency virus type 1 groups M and O from plasma by using highly sensitive and specific generic primers. *J Clin Microbiol* 1999;37:2581–2586.
18. Thompson JD, Higgins DG, and Gibson TJ: CLUSTALW: 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.
19. Felsenstein J: *PHYLIP—phylogeny interference package* (version 3.2). *Cladistics* 1989;5:164–166.
20. Jonassen T, Stene-Johansen K, Berg ES, *et al.*: Sequence analysis of HIV-1 group O from Norwegian patients infected in the 1960s. *Virology* 1997;231:43–47.
21. Sullivan PS, Do AN, Ellenberger D, *et al.*: Human immunodeficiency virus (HIV) subtype surveillance of African-born persons at risk for group O and group N HIV infections in the United States. *J Infect Dis* 2000;181:463–469.
22. Dorn J, Masciotra S, Yang C, *et al.*: Analysis of genetic variability within the immunodominant epitopes of envelope gp41 from human immunodeficiency virus type (HIV-1) group M and its impact on HIV-1 antibody detection. *J Clin Microbiol* 2000;38:773–780.
23. Heyndrickx L, Alary M, Janssens W, Davo N, and van der Groen G: HIV-1 group O and group M dual infection in Benin. *Lancet* 1996;347:902–903.
24. Takehisa J, Zekeng L, Miura T, *et al.*: Triple HIV-1 infection with group O and group M of different clades in a single Cameroonian AIDS patient. *J Acquir Immune Defic Syndr Hum Retrovirol* 1997;14:81–82.
25. Takehisa J, Zekeng L, Ido E, *et al.*: Human immunodeficiency virus type 1 intergroup (M/O) recombination in Cameroon. *J Virol* 1999;73:6810–6820.
26. Peeters M, Liegeois F, Torimiro N, *et al.*: Characterization of a highly replicative intergroup M/O human immunodeficiency virus type 1 recombinant isolated from a Cameroonian patient. *J Virol* 1999;73:7368–7375.

Address reprint requests to:

Renu B. Lal  
HIV Immunology and Diagnostics Branch  
DASTLR/NCID  
Centers for Disease Control and Prevention  
Mail Stop D-12, 1600 Clifton Road  
Atlanta, Georgia 30333

E-mail: rbl3@cdc.gov