Sequence Note

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

CHUNFU YANG,1 FENG GAO,2 PETER N. FONJUNGO,1 LEOPOLD ZEKENG,3 GUIDO VAN DER GROEN,4 DANUTA PIENIAZEK,1 CHARLES SCHABLE,1 and RENU B. LAL1

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 gp41env. 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.

Phylogenetic 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).1,2 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.1 While subtype B is the most predominant subtype in Europe and North America, non-B subtypes and CRFs are more common in developing countries.1 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).2,3 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.4–7 Although the significance of subtype and group designations on the epidemiological, clinical, and diagnostic characteristics of HIV-1 is not fully understood, 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.8–10

While previous attempts to define phylogenetic clusters within group O revealed no clustering pattern,4,11–13 more recent studies have proposed two clusters based on two different gene regions.14–16 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

1HIV Immunology and Diagnostics Branch, Division of AIDS, STD, and TB Laboratory Research, NCID, Atlanta, Georgia 30333.
2Departments of Medicine and Microbiology, University of Alabama at Birmingham, Birmingham, Alabama 35294.
3Laboratoire de Sante Hygiene Mobile, Yaoundé, Cameroon.
4Division of Microbiology, Institute of Tropical Medicine, Antwerp, Belgium.
HIV-1 group O infections were confirmed by an anti-HIV-1 gp120 group O-specific peptide enzyme immunoassay (EIA). 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. For protease, the primary PCR primers are CYF1 (5′-TTTTTTAGGGAAATACTGGCCTCC, nucleotides [nt] 2085–2108 based on HXB2, GenBank accession number K03455[^10^]) and CYR1 (5′-GGTGTATTATAAGGATTTTCAGG, nt 2703–2725); the nested PCR primers are CYF2 (5′-ACGAGCAGACCCATGCAGCCATT, nt 2142–2165) and CYR2 (5′-CATTTTTATCTTTTGTTTCATT, nt 2594–2620). The primary and nested PCR primers for the gp41 region are gp40F1 and gp41R1, and gp46F2 and gp47R2. 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 directions. Nucleotide sequences of the protease and gp41 were then aligned by the CLUSTALW multiple sequence alignment program. Phylogenetic trees were constructed by the neighbor-joining method implemented with the PHYLIP 3.2 package after exclusion of gaps. The amino acid sequences of gp41 immunodominant epitopes were also aligned.

Phylogenetic analysis results

<table>
<thead>
<tr>
<th>Country</th>
<th>Patient ID</th>
<th>Age/sex</th>
<th>Symptom</th>
<th>Protease Accession #</th>
<th>gp41 Accession #</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spain[^a^]</td>
<td>97ES202</td>
<td>39/M</td>
<td>None</td>
<td>UN[^b^]</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>97ES203</td>
<td>37/F</td>
<td>PCP</td>
<td>O</td>
<td>AF229210</td>
</tr>
<tr>
<td></td>
<td>97ES204</td>
<td>44/M</td>
<td>EC</td>
<td>O</td>
<td>AF229211</td>
</tr>
<tr>
<td></td>
<td>97ES205</td>
<td>7/M</td>
<td>PCP</td>
<td>O</td>
<td>AF229212</td>
</tr>
<tr>
<td>United States[^c^]</td>
<td>97US201</td>
<td>42/F</td>
<td>None</td>
<td>O</td>
<td>AF229206</td>
</tr>
<tr>
<td></td>
<td>97US265</td>
<td>40/F</td>
<td>None</td>
<td>O</td>
<td>AF229218</td>
</tr>
<tr>
<td>Cameroon</td>
<td>98CM00T</td>
<td>32/F</td>
<td>None</td>
<td>O</td>
<td>AF229222</td>
</tr>
<tr>
<td></td>
<td>97CM408</td>
<td>38/M</td>
<td>None</td>
<td>O</td>
<td>AF229213</td>
</tr>
<tr>
<td></td>
<td>97CM830</td>
<td>40/F</td>
<td>None</td>
<td>O</td>
<td>AF229215</td>
</tr>
<tr>
<td></td>
<td>97CM755</td>
<td>39/M</td>
<td>AIDS</td>
<td>O</td>
<td>AF229214</td>
</tr>
<tr>
<td></td>
<td>97CM747</td>
<td>36/M</td>
<td>AIDS</td>
<td>O</td>
<td>AF229220</td>
</tr>
<tr>
<td></td>
<td>98CM601</td>
<td>25/F</td>
<td>None</td>
<td>O</td>
<td>AF229217</td>
</tr>
<tr>
<td></td>
<td>97CM768</td>
<td>36/F</td>
<td>AIDS</td>
<td>O</td>
<td>AF229209</td>
</tr>
<tr>
<td></td>
<td>97CM766</td>
<td>43/M</td>
<td>None</td>
<td>O</td>
<td>AF229208</td>
</tr>
<tr>
<td></td>
<td>97CM798</td>
<td>35/M</td>
<td>AIDS</td>
<td>O</td>
<td>AF229207</td>
</tr>
<tr>
<td></td>
<td>97CM761</td>
<td>53/F</td>
<td>AIDS/TB</td>
<td>O</td>
<td>AF229216</td>
</tr>
<tr>
<td></td>
<td>98CM139</td>
<td>35/M</td>
<td>None</td>
<td>O</td>
<td>AF229224</td>
</tr>
<tr>
<td></td>
<td>98CM226</td>
<td>NA[^d^]</td>
<td>NA</td>
<td>UN[^b^]</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>97CM773</td>
<td>39/F</td>
<td>AIDS</td>
<td>O</td>
<td>AF229219</td>
</tr>
<tr>
<td></td>
<td>98CM010</td>
<td>26/F</td>
<td>AIDS</td>
<td>O</td>
<td>AF229221</td>
</tr>
<tr>
<td></td>
<td>98CM094</td>
<td>31/F</td>
<td>None</td>
<td>O</td>
<td>AF229223</td>
</tr>
<tr>
<td></td>
<td>98CM215</td>
<td>NA</td>
<td>NA</td>
<td>O</td>
<td>AF229225</td>
</tr>
</tbody>
</table>

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

[^a^]: The specimens 97ES202 and 97ES203 are from Spain, whereas 97ES204 and 97ES205 are from persons born in Guinea Bissau.

[^b^]: UN, Unamplifiable; amplification in the protease region was not successful.

[^c^]: These cases have previously been described in detail.

[^d^]: NA: not available.

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
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 previously 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.14–16 The gp41 sequences for Spanish and U.S. specimens have previously been published.15,21 Interestingly, the four remaining isolates from Cameroon that were
FIG. 3. Alignment of deduced amino acid sequences from the gp41 region (amino acids 569–688). The identical amino acids are indicated by dots; insertions or deletions are indicated by dashes. Positions of clusters I and II are shown by angle brackets (<>); the two conserved cysteine residues comprising the cysteine loop are indicated by asterisks (*); the potential N-linked glycosylation sites are indicated by carets ^^^ and the conserved tryptophan residues are in boldface.
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. Interestingly, the first known case of AIDS documented in Europe was a Norwegian sailor infected with a group O virus. He developed clinical manifestations of AIDS in 1966, followed by his wife and daughter, all of whom died in the 1970s. 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 revealed minor (<20%) substitutions (Fig. 3). Analysis of a cytotoxic T lymphocyte (CTL) epitope within cluster I (ALETIQNQLL: amino acids 591–602) revealed substitutions at L595 → F and I596 → M. Although the cysteine loop (CKGKLVC: amino acids 607–613) had greater numbers of substitution, most were conservative (K610 → R and V612 → 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, the biologic significance of this deletion remains to be determined. Likewise, for most specimens, the number of potential glycosylation sites (NX(S/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. 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


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