

Phylogenetic analysis of gp41 envelope of HIV-1 groups M, N, and O strains provides an alternate region for subtype determination

Danuta Pieniazek, Chunfu Yang, and Renu B. Lal

HIV/AIDS and Retrovirology Branch, Division of AIDS, STD, and TB Laboratory Research, National Center for Infectious Diseases, Centers for Disease Control and Prevention, 1600 Clifton Road, Mail Stop G19, Atlanta, GA 30333; email: RBL3@CDC.GOV.

Human immunodeficiency virus type 1 (HIV-1) is characterized by an unusually high degree of genetic variability in vivo. Analysis of HIV-1 *env* sequences of virus strains from different geographic locales has revealed that HIV-1 can be divided into three main groups, M (major), O (outlier), and N (new). HIV-1 group M has been further subdivided into genetically equidistant clusters of HIV-1 *env* genes, comprising subtypes A–J [1, 2]. Some HIV-1 variants reported as “uncertain” or “unclassifiable” (category U) may represent either a new subtype or a recombinant sequence and their designation may be clarified as more reference sequences become available [2]. Likewise detailed molecular analysis of many group O strains have not revealed any specific clustering into subtypes [3]. More recently, a new variant of HIV-1, termed group N, has been discovered in Cameroon [4]. These genotypic analyses have resulted in better understanding of the molecular epidemiology of HIV-1 and improved tracking of the epidemic, in addition to defining the geographic distribution and transmission patterns of various HIV-1 group M subtypes [2].

While full-length genomic analyses are needed for new subtype designation and to identify recombinant viral genomes [2], they are not suitable for large molecular epidemiologic studies due to labor and time expense. Thus the shorter fragments of the genome still remain the primary tool used to monitor HIV-1 genetic diversity worldwide. Since the *env* sequences provide information on all known and potentially new subtypes circulating in a given geographic area, the *env* region remains the principal target for HIV-1 subtyping to assist in continued identification of old and new variants. The most commonly used procedure for HIV-1 subtyping has relied on analysis of the C2–V3 segment of *env*. While most PCR-derived sequences contain a suboptimal length for phylogenetic analysis, a comparative analysis with larger *env* sequences has revealed that limited V3 region sequences can generally serve as a reliable basis for subtype determination [2, 5]. However, because of the broad heterogeneity within the C2–V3 domain of HIV-1 group M viruses and the constant nucleotide changes in this region over time [5], a plethora of different sets of primers are needed to maximize the efficiency of PCR amplification and sequencing worldwide. Moreover, creating the DNA sequence alignments from clinical specimens may be extremely difficult because of tremendous nucleotide variations due to deletions and insertions in this region. In addition, different sets of the C2–V3 primers are needed to generate data for group O and N viruses.

To circumvent these practical problems, we sought to determine if a highly stable transmembrane *env*-gp41 could serve as an alternate region for a global phylogenetic analysis of HIV-1 subtypes. Our initial approach was to determine whether HIV-1 gp41 and C2–V3 sequences would cluster into the same phylogenetic subtypes. A parallel phylogenetic analysis of the C2–V3 and gp41 sequences was done using group M, group N and group O reference sequences from full length genomes, as well as commonly used marker sequences of subtypes A–H and chimpanzee viral sequences of CPZ-ANT and CPZ-GAB strains [6, 7]. The results of the phylogenetic analysis demonstrated an agreement between phylogenetic clustering of gp41 and C2–V3 HIV-1 sequences into group M subtypes A–H viruses, group N, and group O by using both the maximum-likelihood and neighbor-joining methods (Fig. 1). In the neighbor-joining method, bootstrap values at branch nodes connected with subtypes range from 95% to 100% for both C2–V3 and gp41 trees. More importantly, while there were differences in the overall topologies of the phylogenetic C2–V3 and gp41 trees, the subtype assignment did not change in any case. Likewise, a comparison of phylogenetic analysis of *env* C2–V3 versus V3–V5 fragments showed differences in tree topology, although the subtype designations remain unchanged [10]. Thus, based on the remarkably similar phylogenetic clustering of C2–V3 and the gp41 region, we conclude

that the global diversity of the gp41 region is sufficient to provide a reliable marker for phylogenetic clustering of HIV-1 subtypes.

Despite the nucleotide divergence within the gp41 region, which allowed phylogenetic analysis for subtype determination, some fragments of the gp41 region showed enough conservation to permit the design of consensus PCR primers. Thus, we designed one set of *env* gp41 primers, namely gp41M/O, for a nested PCR amplification and sequence analysis of group M, N, and O viruses [2]. The primers are:

Name	Direction	Primer Sequence	Genomic Position*
gp40F1	forward	5' TCTTAGGAGCAGCAGGAAGCACTATGGG	7789–7816
gp41R1	reverse	5' AACGACAAAGGTGAGTATCCCTGCCTAA	8347–8374
Nested PCR			
gp46F2	forward	5' ACAATTATTGTCTGGTATAGTGCAACAGCA	7850–7879
gp47R2	reverse	5' TTAAACCTATCAAGCCTCCTACTATCATTAA	8281–8310

*Position by alignment to HXB2, GenBank accession number K03455.

See HXB2 numbering system at <http://hiv-web.lanl.gov/NUM-HXB2/HXB2.MAIN.html>

The primer sequences are highly conserved for group M, group N, and group O sequences (<http://hiv-web.lanl.gov/ALIGN-99/subtype-alignments.html>). The PCR conditions included denaturation at 94° C for 2 min, followed by 35 cycles of denaturation at 94° C for 30 sec, annealing at 50° C for 30 sec, and extension at 72° C for 60 sec, with a final extension at 72° C for 5 min. The PCR amplification conditions using DNA-PCR and RT-PCR assays for peripheral blood mononuclear cells (PBMC) and plasma, respectively, have been described elsewhere [11]. A fragment of about 460 bp spanning approximately 40% of the gp41 region, which includes the immunodominant region, was successfully amplified from as few as 1 to 5 copies of viral DNA from nearly full-length HIV-1 clones representing group M subtypes A–H. The assay is highly sensitive in detecting plasma viral RNA from HIV-1 strains of diverse geographic origins representing different subtypes of HIV-1 group M as well as HIV-1 group O. Of the 253 group M plasma specimens (subtypes A = 68, B = 71, C = 19, D = 27, E = 23, F = 33, and G = 12), 250 (99%) were amplified using the gp41M/O primer set. More importantly, all 32 (100%) group O plasma samples were also amplified with these primers [11]. *In vitro* spiking experiments further revealed that the assay could detect as few as 10 copies of viral RNA/mL of plasma and gave positive signals in selected HIV-1 seropositive plasma, with viral copy numbers below the detection limits of all commercially available viral load assays. Moreover, the gp41M/O primer set amplified viral DNA of the group N YBF30 strain indicating the potential of these primers to amplify new divergent viruses. Additionally, the gp41M/O primers were highly specific for HIV-1; no PCR amplification of HIV-2, SIVs, and SIV-cpz strains was observed. Therefore, our findings indicated that the universal gp41M/O primer set that we designed was indeed highly specific, sensitive, and efficient in a PCR amplification of the gp41 region, regardless of the genotypes and geographic origin of the viruses. Finally, we have performed phylogenetic analysis of gp41 sequences from more than 500 group M and 32 group O specimens that were successfully amplified from PBMCs and plasma as a part of various ongoing studies worldwide. The representative group M subtype A–G and group O sequences from selected geographic regions (Thailand, Egypt, Uganda, Ghana, India, Zimbabwe, Argentina, and Cameroon) are shown in Fig. 2. As expected, sequences from Thailand specimens clustered on a distinct branch within the subtype B, as had previously been shown for C2–V3 sequences. One of four unclassifiable variants had variable positions in the tree topology depending on the phylogenetic method used (Fig. 2). Analysis of other parts of viral genomes from these four variants, including the entire *env* region, is now underway in order to classify them into either potential new subtypes or recombinants along the *env* gene. Taken together, these studies indicate that gp41 provides a simple and practical region for subtype determination of group M HIV-1.

In conclusion, we have found that genetic analysis of PCR-amplified HIV-1 gp41 sequences provides a powerful approach for identifying groups M, N, and O viruses. Furthermore, group M

Phylogenetic Analysis of gp41

viruses can be phylogenetically classified into at least seven subtypes, the same that have been previously established by C2—V3 sequences. Detection of HIV-1 groups M, N, and group O infections with a very high efficiency is feasible by using only one set of primers for all known HIV-1 strains worldwide. Thus, using gp41 region for phylogenetic analysis of HIV-1 strains worldwide provides an alternate and effective tool for determination of distinct HIV-1 variants. In addition, the gp41M/O primer set may have a broad spectrum of future applications, ranging from quantitative measurement of viral load for clinical management of infected patients to utilization as a diagnostic tool for early detection of HIV-1 infection in blood donors. Moreover, the gp41 region has multiple functions ranging from membrane fusion, endocytosis signals, and calmodulin-binding that potentially affect critical cellular signal transduction pathways [12]. These functions along with structural elements that interact with the assembling capsid precursors suggest that this region might play an important role in both virus replication and pathogenesis *in vivo* [12]. Thus, the gp41 protein sequences generated worldwide would provide useful information related to immune response, as well as the information related to T20/DP178 epitope which has recently been used in therapeutic trial to block HIV entry [13]. Finally, the gp41 sequences generated would provide valuable information regarding the diversity and its diagnostic implication within the distinct domains of this HIV-1 multi-functional protein.

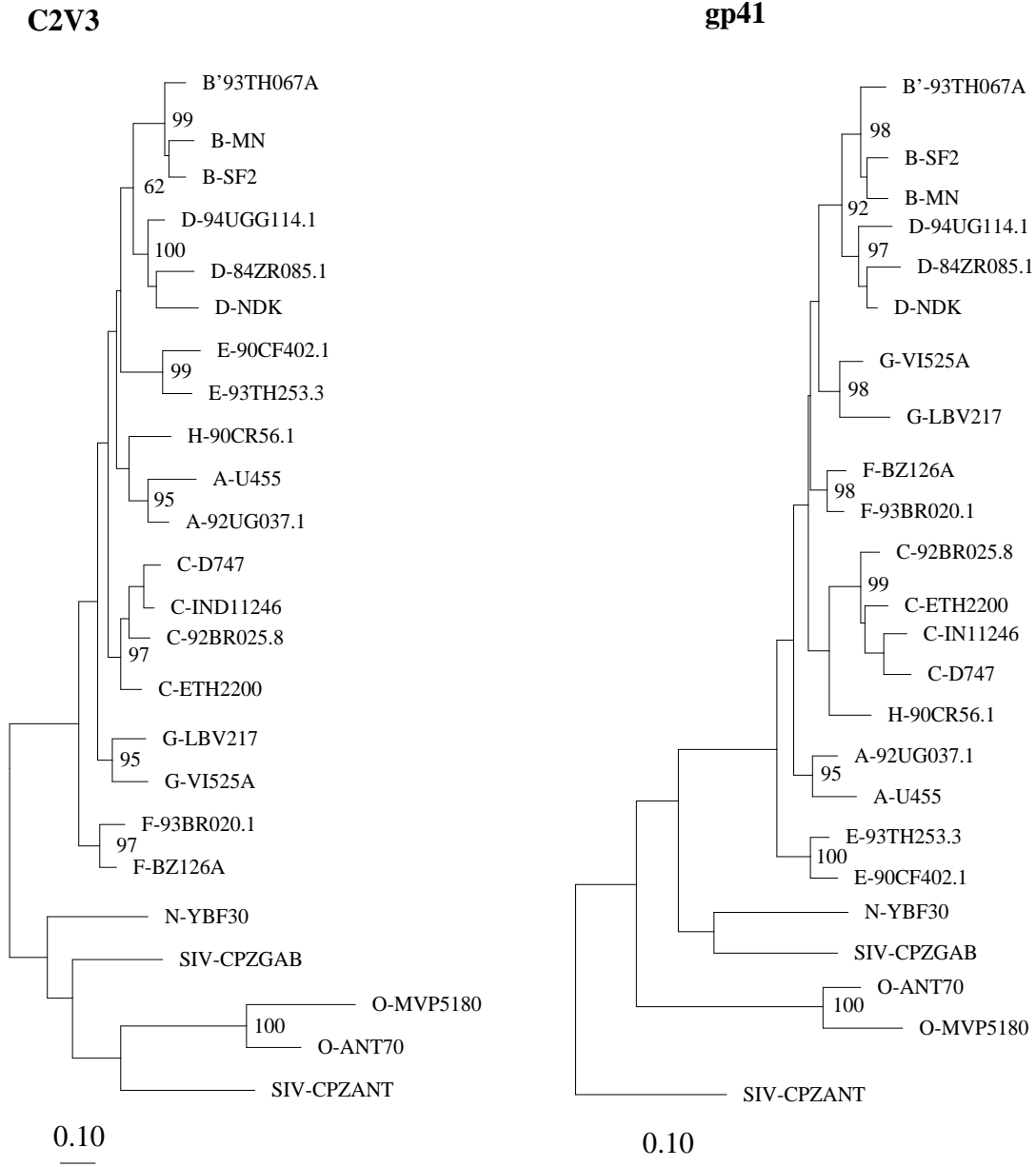


Figure 1. Assignment of HIV-1 reference strains from groups M, N, and O into subtypes by parallel phylogenetic analysis of the *env* C2–V3 and gp41 sequences. After gap stripping, 281 and 354 nucleotide positions were included in the *env* C2–V3 and gp41 alignments, respectively, with a transition/transversion ratio of 1.61 and 1.46. The trees were constructed on the basis of DNA sequences by the maximum-likelihood method using fastDNAm1 program (8). HIV-1 subtypes and groups are denoted with prefixes A, B, C, D, E, F, G, and H, and N and O, respectively. SIV-CPZGAB and SIV-CPZANT sequences were included in the analysis. The numbers at the nodes of subtypes A, B, C, D, E, F, H, G, and group O correspond to the bootstrap values in neighbor-joining trees (9) of C2–V3 and gp41. Trees were re-rooted by using HIV-2ROD as an outgroup. The scale bar indicates an evolutionary distance of 0.10 nucleotide per position in the sequence. Vertical distances are for clarity only.

gp41

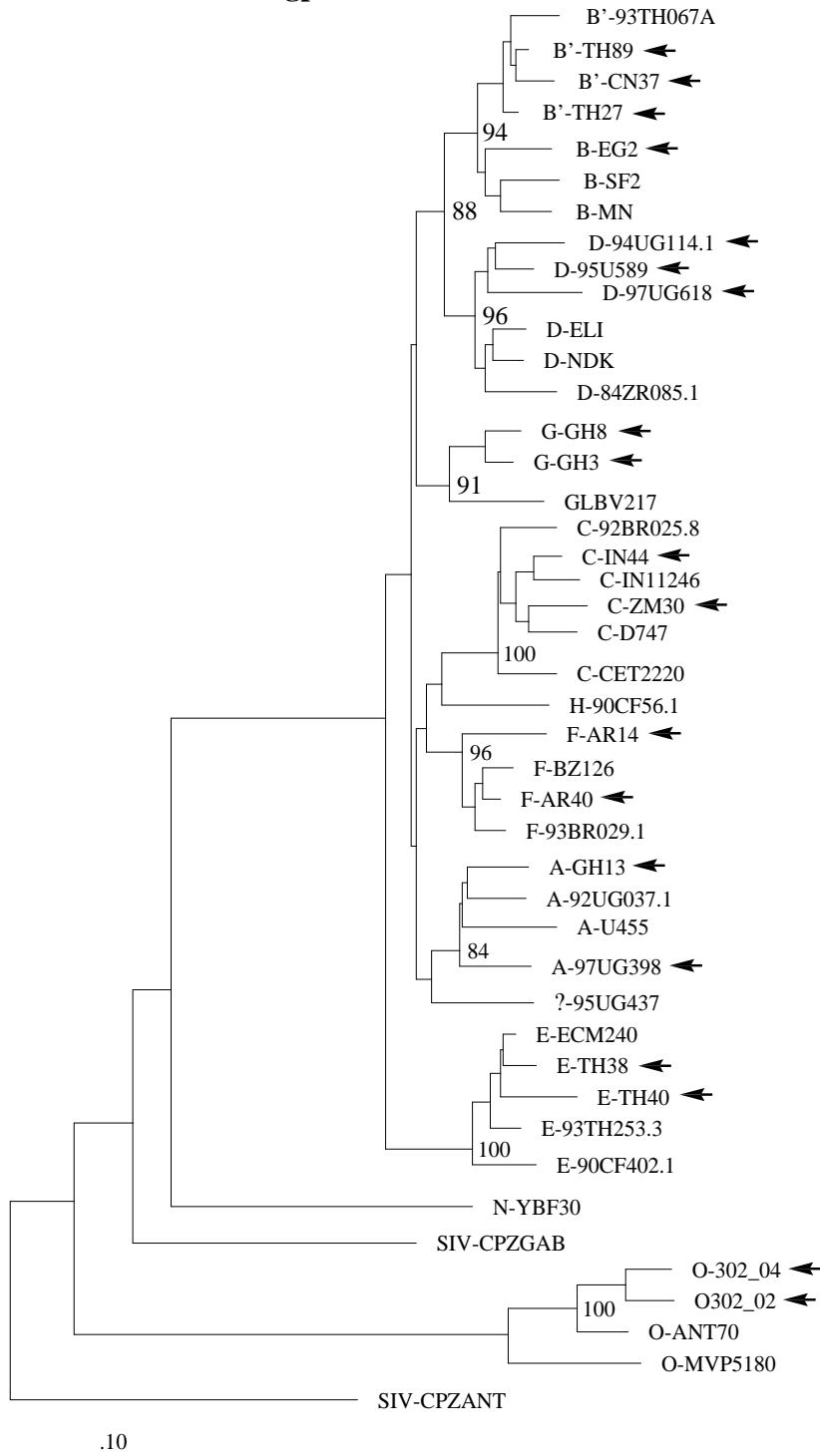


Figure 2. Phylogenetic classification of the new HIV sequences (←) from representative group M subtypes A–G and group O worldwide based on *env* gp41 region. ? represents the unclassified position of sample 95UG437. Genbank accession #AF113577–AF113595. The tree was generated using the neighbor-joining method (9). Numbers at the nodes indicate the percentage of bootstrap values (only values above 80% are shown). The tree was re-rooted by using HIV-2ROD as an outgroup.

REFERENCES

- [1] Hu, D.J., Dondero, T.J., Mastro, T.D., and Gayle, H.D. Global and molecular epidemiology of HIV. Pp. 27–40 In: *AIDS and other manifestations of HIV infection*. Ed. Wormser, GP. Lipponcott-Raven Publishers, 1998.
- [2] Carr JK, Foley BT, Leitner T, Salminen M, Korber B, McCutchan F. Reference sequences representing the principal genetic diversity of HIV-1 in the pandemic. Pp. III-1–III-8 In: *Human Retrovirus and AIDS 1998*. Eds. Korber, B., Foley, B., McCutchan, F., Mellors, J., Hahn, B. H., Sodroski, J. and Kuiken, C. Los Alamos National Laboratory, Los Alamos, New Mexico.
- [3] Bibollet-Ruche, F., Loussert-Ajaka, I., Simon, F., Mboup, S., Ngole, E.M., Saman, E., Delaporte, E., and Peeters, M. Genetic characterization of accessory genes from human immunodeficiency virus type 1 group O strains. *AIDS Res. Hum. Retroviruses* **14**:951–961, 1998.
- [4] Simon, F., Mauclore, P., Roques, P., Loussert-Ajaka, I., Muller-Trutwin, M., Saragosti, S., Georges-Courbot, M.C., Barre-Sinoussi, F., and Brun-Vezinet, F. Identification of a new human immunodeficiency virus type 1 distinct from group M and group O. *Nat. Med.* **4**:1032–1037, 1998.
- [5] Dighe, P.K., Korber, B.T., and Foley, B.T. Global variation in the HIV-1 V3 region: divergent patterns off of V3 loop evolution in M group subtypes. Pg. III74–III204. In: *Human Retrovirus and AIDS 1997*. Eds. Korber, B., Hahn, B. H., Foley, B., Mellors, J., Leitner T., Myers, G., McCutchan, F., and Kuiken, C. Los Alamos National Laboratory, Los Alamos, New Mexico.
- [6] Gao, F., Robertson, D.L., Carruthers, C.D., Morrison, S.G., Jian, B., Chen, Y., Barre-Sinoussi, F., Girard, M., Srinivasan, A., Abimiku, A.G., Shaw, G.M., Sharp, PM and Hahn B.H. 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**:5680–8, 1998.
- [7] Leitner, T., Korber, B., Robertson, D., Gao, F., and Hahn B. Updated proposal of reference sequences of HIV-1 genetic subtypes. Pg. III19–III24. In: *Human Retrovirus and AIDS 1997*. Eds. Korber, B., Hahn, B. H., Foley, B., Mellors, J., Leitner T., Myers, G., McCutchan, F., and Kuiken, C. Los Alamos National Laboratory, Los Alamos, New Mexico.
- [8] Larsen, N., Olsen, G.J., Maidak, B.L., McCaughey, M.J., Overbeek, R., Macke, T.J., Marsh, L.T., and Woese, C.R. The Ribosomal Database Project. *Nucleic Acids Res.* **22**:3485–7, 1993.
- [9] Felsenstein, J. PHYLIP-phylogeny interference package (version 3.2). *Cladistics* **5**:164–6, 1989.
- [10] Janssens, W., Heyndrickx, L., Van de Peer, Y., et al. Molecular phylogeny of part of the *env* gene of HIV-1 strains isolated in Cote d'Ivoire. *AIDS* **8**:21–26, 1994.
- [11] Yang C., Pieniazek, D., Owen, S.M., Fridlund, C., Nkengasong, J., Mastro, T.D., Rayfield, M.A., Downing, R., Biryawaho, B., Tanuri, A., Zekeng, L., van der Groen, G., Gao, F., and Lal, R.B. Plasma detection of phylogenetically diverse human immunodeficiency virus type 1 using generic primers amplifying both group M and group O. (Submitted).
- [12] Hunter E. gp41, a multifunctional protein involved in HIV entry and pathogenesis. Pg. III55–III73. In: *Human Retrovirus and AIDS 1997*. Eds. Korber, B., Hahn, B. H., Foley, B., Mellors, J., Leitner T., Myers, G., McCutchan, F., and Kuiken, C. Los Alamos National Laboratory, Los Alamos, New Mexico.
- [13] Kilby JM, Hopkins S, Venetta TM, DiMassimo B, Cloud GA, Lee JY, Alldredge A, Hunter E, Lambert D, Bolognesi D, Matthews T, Johnson MR, Nowak MA, Shaw GM, Saag MS. Potent suppression of HIV-1 replication in humans by T-20, a peptide inhibitor of gp41-mediated virus entry. *Nat. Med.* **4**:1302–7, 1998.