

ALIGNMENT OF HIV-1/HIV-2/SIV GENOMES

This alignment contains a sampling of HIV-1 genomes (one or two per subtype from the HIV-1 M group, one from the N group and one from the O group), at least one of each type of SIV that has been sequenced (Chimpanzee; Vervet, Grivet, Sabaeus, and Tantalus subspecies of African green monkeys; Sooty mangabeys and macaques infected with Sooty mangabey virus; L'Hoest monkey; Mandril) and sequences of HIV-2 subtypes A and B which are closely related to the Sooty mangabey viruses. Table 1 lists the sequence name, database accession number, country of isolation, first author and publication for each of the sequences in this alignment.

Together, these genomes represent the full breadth of diversity discovered to date in the primate lentivirus lineage. Non-primate lentiviruses such as equine infectious anemia virus (EIAV) and caprine arthritis/encephalitis virus (CAEV) are more distantly related to each other and to the primate lentiviruses, and are not included in this alignment.

The new alignment was based on the previous version, but entirely re-done using manual editing with the program BioEdit,

<http://www.mbio.ncsu.edu/RNaseP/info/programs/BIOEDIT/bio-edit.html>

created by Tom Hall. It is now entirely codon-aligned (meaning that the correct translation reading frame has been maintained as much as possible; in the case of overlapping frames, obviously one will not translate correctly). We have tried to 'reset' the alignment so that each new gene starts in the first frame; this means in some places empty columns have been inserted. The annotation is unchanged from last year's compendium.

The annotation is based on known protein coding regions in HIV-1 and on annotations found in SIV sequence database entries. The protein cleavage sites that create Gag p17, Gag p24 and other mature peptides from the Gag and Gag-Pol precursor polyproteins have been experimentally determined for HIV-1 and at least one strain of HIV-2, the study of analogous cleavages in SIV polyproteins have not been published. Two representative genomes have been translated in all three reading frames; HIV-1 subtype B strain HXB2 and SIV from a Sykes monkey. The translations are provided as a visual aid for finding landmarks in the genomes.

The HIV-2/SIV-SMM vpx gene is postulated to be a duplication of the vpr gene (Tristem et al. *Nature* **347**:341–342 (1990)) and thus there may be two alternative alignments of this region of the genome, as there are for the duplicated stem-loops of the TAR element.

HIV Database compendia from previous years, as well as electronic copies of these alignments in a variety of formats are available on our WWW site at

<http://hiv-web.lanl.gov/HTML/compendium.html>

and

http://hiv-web.lanl.gov/ALIGN_CURRENT/ALIGN-INDEX.html

respectively.

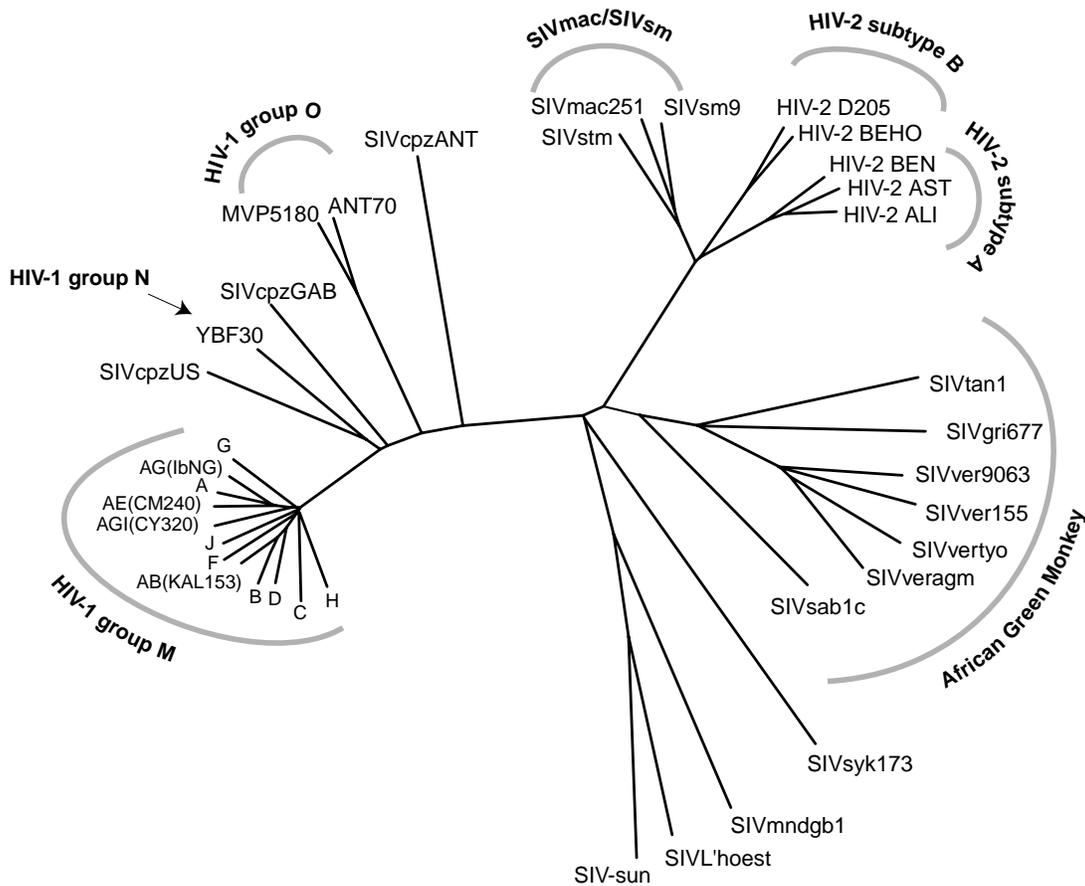


Figure 1. This phylogenetic tree was constructed from the complete lentivirus genome alignment. Columns in the alignment for which any sequence is represented by a gap were first removed. The tree was constructed using PHYLIP DNADIST with maximum likelihood distance estimation and a transition/transversion ratio of 1.6. The DNADIST output was then used to create a neighbor-joining tree using the WEIGHBOR program (The PHYLIP neighbor program gave similar results). WEIGHBOR is available in UNIX, Mac, and PC at

<http://t10.lanl.gov/billb/weighbor/index.html>

A cut and paste interface to weighbor is available at

<http://bioweb.pasteur.fr/seqanal/interfaces/weighbor.html>

Some of the genomes are known to represent recombinants or mosaics between two or more different lineages, so this tree should not be assumed to represent the true phylogenetic history of these viruses. For example, the SABIC genome from the sabaeus subspecies of African green monkeys is recombinant between an African green monkey lentiviral lineage and a Sooty mangabey viral lineage. When the AGM-like regions of its genome are used to build a phylogenetic tree, the SABIC lineage shares the same major branch with the other AGM lineages. Given these limitations, the tree is still useful for graphically representing the diversity and relationships between these viruses. It is readily apparent that the HIV-2 viruses are related to Sooty mangabey viruses.