



Structure of Human Immunodeficiency Virus (HIV)

## Hands-On Five Origins of HIV

To reconstruct the history of HIV, researchers sequenced the gene that codes for the reverse transcriptase, for the *env*, *gag*, and *pol* genes in several simian immunodeficiency viruses (SIVs) and compared them (or their translated proteins) to the sequences found in a variety of HIV strains. The SIVs are parasites that infect the immune systems of chimpanzees and monkeys.

We are going to build phylogenetic trees with 12 sequences from HIV1, HIV2 and SIV. The sequences are from NCBI. The table lists, from left to right:

- **Num**: the number representing the order in which the sequences were obtained from NCBI
- **Isolate**: the isolate names of the genomes as given by NCBI
- **Accession Number**: GenBank accession numbers of the whole genomes
- **Length in base pairs**: the length of the genome
- **Posted Entry Date**: the date found in the corresponding database entry
- **Organism**: the organism to which the genome belongs.

Num	Isolate	Accession Number	Length in base pairs	Posted Entry Date	Organism
1	HIV1_ELI	K03454	9176	27-FEB-2002	Human
2	HIV1_BRU	K02013	9229	02-AUG-1993	Human
3	HIV1_MAL	X04415	9229	18-APR-2005	Human
4	HIV1_NDK	M27323	9143	02-AUG-1993	Human
5	HIV2_D205	X61240	10269	14-NOV-2006	Human
6	HIV2_ROD	M15390	9671	23-MAY-1996	Human
7	HIV2_ST	M31113	9672	29-MAY-1996	Human
8	HIV2_UCI	L07625	10271	02-AUG-1993	Human
9	SIV_Mm251	M19499	10277	29-NOV-2000	Macaque
10	SIV_CPZ	X52154	9811	18-APR-2005	Chimpanzee
11	SIV_AGM	M58410	9623	31-MAR-2006	Af. green monkey
12	SIV_SMM	X14307	10241	18-APR-2005	Sooty mangabey mk.

Three files, *env\_protein\_sequences.txt*, *gag\_protein\_sequences.txt*, and *pol\_protein\_sequences.txt* containing the amino acid sequences for the *env*, *gag* and *pol* proteins from the 12 isolates were created. In other words:

- *env\_protein\_sequences.txt* contains 12 *env* protein sequences, one from each of the 12 genomes found in the table.
- *gag\_protein\_sequences.txt* contains 12 *gag* protein sequences, one from each of the 12 genomes found in the table.
- *pol\_protein\_sequences.txt* contains 12 *pol* protein sequences, one from each of the 12 genomes found in the table.

**A)**

- Retrieve “env\_protein\_sequences.txt” and copy all 12 sequences.
- Go to <http://www.phylogeny.fr/index.cgi>
- Click on “One Click” under “Phylogeny analysis”
- Paste all 12 sequences in the window.
- Click on the “Submit” button.
- The new page contains the actual phylogenetic tree.
- Save the tree in pdf by clicking on “PDF” following the caption of the tree: “Figure 1: *Phylogenetic tree*”.

Remarks:

- 1) A different view of the tree can be obtained by scrolling to “Tree style”. For example, choose the “Radial (by Drawtree)” radio button and save the resulting tree in Pdf.
- 2) The steps that the sequences undergo to produce the phylogenetic tree are depicted in the flowchart that can be obtained by clicking on “1. Overview” at the top of the page.

**B)**

- 1) Repeat the procedure described in part A with gag\_hiv\_siv\_sequences.txt and save the resulting consensus trees under gag\_consense\_outfile.txt.
- 2) Repeat the procedure described in part A with pol\_hiv\_siv\_sequences.txt and save the resulting consensus trees under pol\_consense\_outfile.txt.

**C)**

Study carefully the three consensus trees and answer the following questions:

- 1) What do the trees show with regards to the HIV and SIV relationships?
- 2) Why do SIV’s cluster with both HIV-1 and HIV-2?
- 3) Which HIV type, HIV-1 or HIV-2, is more closely related to the SIV from the sooty mangabey? Which type is more closely related to the SIV from the chimpanzee? What does this tell you about the origin of HIV-1 and HIV-2?

This project was adapted from “The Origin and Evolution of HIV” from Siv Andersson’s laboratory at Uppsala University, Sweden.

## **More about the HIV Genome and Virus**

<http://www.hiv.lanl.gov/>

The HIV databases contain data on HIV genetic sequences, immunological epitopes, drug resistance-associated mutations, and vaccine trials. The website also gives access to a large number of tools that can be used to analyze these data. This project is funded by the Division of AIDS of the National Institute of Allergy and Infectious Diseases (NIAID), a part of the National Institutes of Health (NIH).