

## **The Evolutionary Origin of the AIDS Virus.**

A BioinFORMatics Exercise by David Form,  
Minuteman Regional High School

In this exercise you will analyze nucleotide sequences from HIV and various related viruses in order to help determine the source of the AIDS virus and construct a scenario that will explain how the AIDS virus originated in Africa in the early 20<sup>th</sup> Century.

In the early 1980's it was demonstrated that AIDS results from an infection with HIV-1, the human immunodeficiency virus. Based on comparisons of HIV sequences from various sources, the origin of the AIDS virus can be pinpointed to Africa in the 1930's. These comparisons reveal that the earliest sample of HIV is from a blood sample drawn from an African in 1959.

HIV is a retrovirus, meaning that it contains single-stranded RNA. Its genome contains the gene for reverse transcriptase, an enzyme that converts RNA into complementary DNA. The gene that codes for reverse transcriptase in HIV is the POL gene. Pol genes are found in many retroviruses, including a number that are harbored by human populations.

Various possibilities exist for the origin of HIV. One possibility is that the virus was always present in the human population, originating by evolution from one of those other retrovirus, endemic to human populations. In this scenario, the AIDS virus then remained confined to isolated populations in Africa until conditions in the 20<sup>th</sup> century, such as population growth, increased intravenous drug use, widespread travel, etc. lead to its spread.

It has been found that nonhuman primate species in Africa harbor a group of HIV-related viruses, known as SIV (simian immunodeficiency virus). AIDS appears to have originated in areas in which these SIV-infected primates are found. Did HIV originate from one or more transfers of SIV from an African primate to humans?

The AIDS virus is not a single entity, but consists of a number of strains that can be grouped into two distinct types, HIV-1 and HIV-2. HIV-2 is less

virulent and is mostly restricted to populations in West and Central Africa. Although HIV-1 and HIV-2 are genetically distinct, they are related. DID HIV-2 evolve from HIV-1, or did it originate independently from another source?

As discussed above, various possibilities, or hypotheses, can be generated to explain the origin of HIV-1 and HIV-2. Your assignment is to generate a phylogenetic tree that shows the relationships among HIV-1, HIV-2, SIV from a number of African primates and various other retroviruses. You will then use this tree to construct a set of hypotheses that explain the origin of HIV-1 and HIV-2 as related or independent events. Analysis of your phylogenetic tree will support some hypotheses and exclude others. These hypotheses will then allow you to write a scenario explaining how AIDS arose in Africa.

### **Objectives:**

1. Use ClustalW to construct a multiple alignment of nucleotide sequences.
2. Use DrawGram to construct a rooted phylogenetic tree.
3. Interpret your phylogenetic tree, using the data to support or reject various hypotheses concerning the origin of the AIDS virus.
4. Use phylogenetic data to support the idea that viruses can cross the species barrier.
5. Describe a plausible scenario for the evolution and spread of the AIDS virus in the human population during the mid-20<sup>th</sup> Century.
6. Demonstrate how nucleotide sequence data can be used to support your scenario.

### ***Directions:***

Our phylogenetic tree will be prepared using a collection of bioinformatics tools available through the ***Biology Workbench*** @workbench.sdsc.edu/. Your first step is to set up an account, by clicking on ***set up an account*** and following the directions.

The nucleotide sequences are in FASTA format in a text file, Pol.doc. This file contains the nucleotide sequences for the pol gene from HIV-1, HIV-2, chimpanzee SIV, mangabey (a baboon-like primate) SIV, green monkey,

SIV, mandrill Siv, human lymphotropic virus, human hepatitis B, human foamy virus and the feline leukemia virus (FIV).

### ***HIV phylogeny***

#### **Step 1: Start a session**

1. Scroll down until you see a series of boxes.
2. Click on the box labeled ***Session Tools***. Now look through the menu of the long rectangular box and select ***Start New Session***. Name the session.
3. Select the box labeled ***Nucleotide Tools***.

#### **Step 2: You are now ready to upload the viral sequence Pol.doc s.**

4. From the menu, select ***Add a New Nucleotide Sequence***. A new box will appear on your screen.
5. Select ***Browse***, then find the file named and open it.
6. Select the ***Upload*** button. The sequences will be uploaded into your session window.
7. Go to the top or bottom of the screen and click on the ***Save*** button. You should now see a list of your sequences.
8. Click to the left of the sequence name to select it. Select the all of the sequences.
9. From the menu box, select ***ClustalW***. Select ***submit*** from the box on the screen. When the program has finished running, scroll down to view your alignment.

#### **Step 3: Creating a phylogenetic tree.**

10. Find the window labeled ***Import Alignments*** and click on it.
11. Under the rectangle labeled Alignment Tools, select ***DrawGram***.
12. Click on the small box next to the list of aligned sequences and select submit.
13. Scroll down to see your phylogenetic tree. The tree can be dragged to a word processing file.

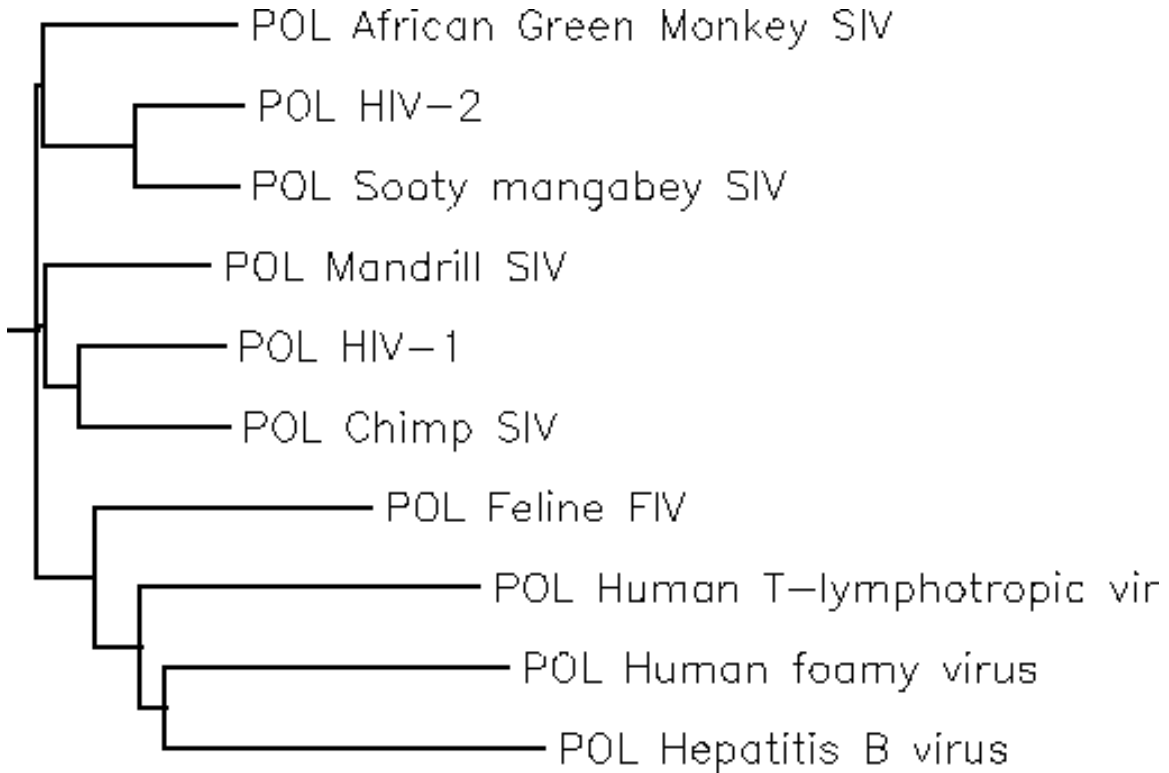
#### **Critical Thinking Exercises:**

1. What is the most likely source of HIV-1? HIV-2? Explain how your tree lead you to these conclusions.
2. Did HIV-2 evolve from HIV-1, or was the evolution of HIV-2 an independent evolutionary event? What does this tell us about the likelihood of new viral diseases emerging in Africa and other places.

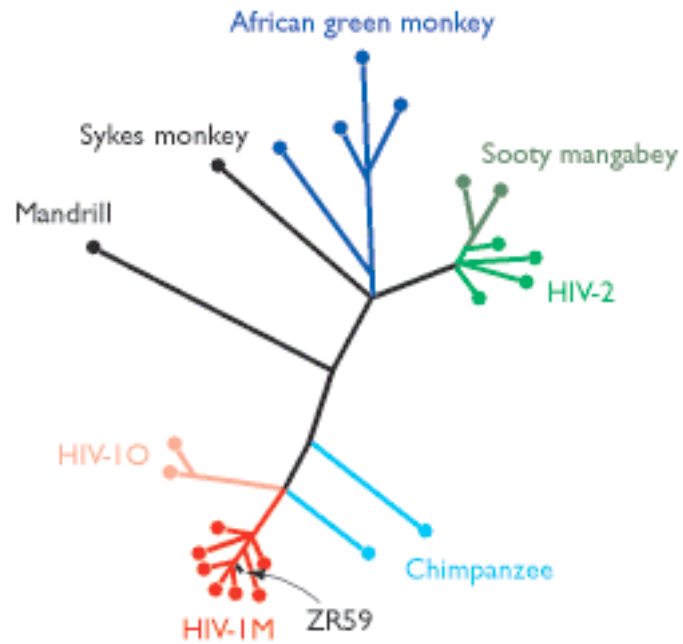
3. Is there any evidence that HIV-1 or HIV-2 evolved from retroviruses that were already endemic to human populations?
4. Write a scenario, in the form of a paragraph, describing how AIDS may have originated in Africa. Include a description of how a new strain of AIDS may emerge in the future.
5. In this exercise we analyzed the sequences of the Pol gene. Why was this gene chosen? Do you think that looking at another gene would give the same result? Look at the graphical representations of the genomes of HIV-1, HIV-2 and the other retroviruses. Which other gene sequences would you suggest analyzing? How did you arrive at this decision?
6. Wain-Hobson studied the sequences of various isolates of HIV-1, HIV-2 and SIV from a number of primates in Africa. The phylogenetic tree that they constructed is included with this exercise. How do your results compare with theirs? What are the major similarities? What are the major differences?

**Student Results**

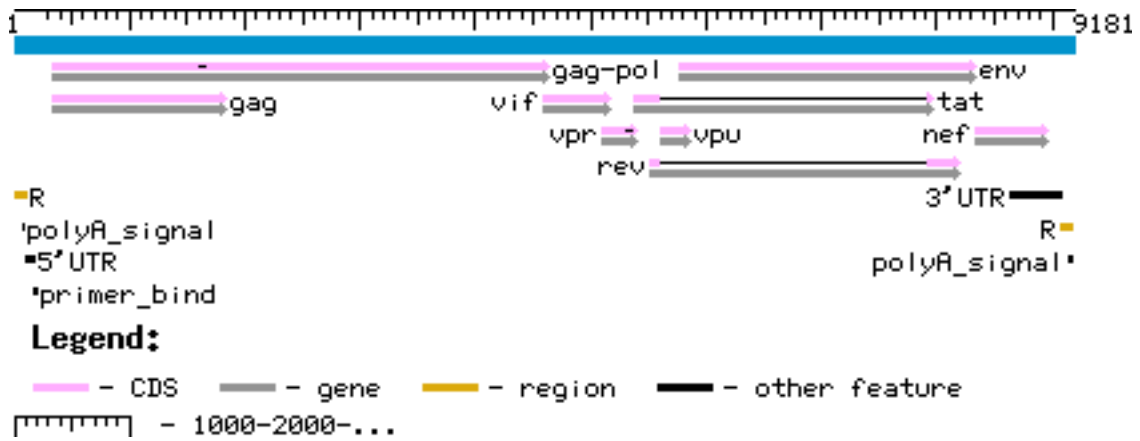
**Relationships among the POL genes from a variety of viruses**



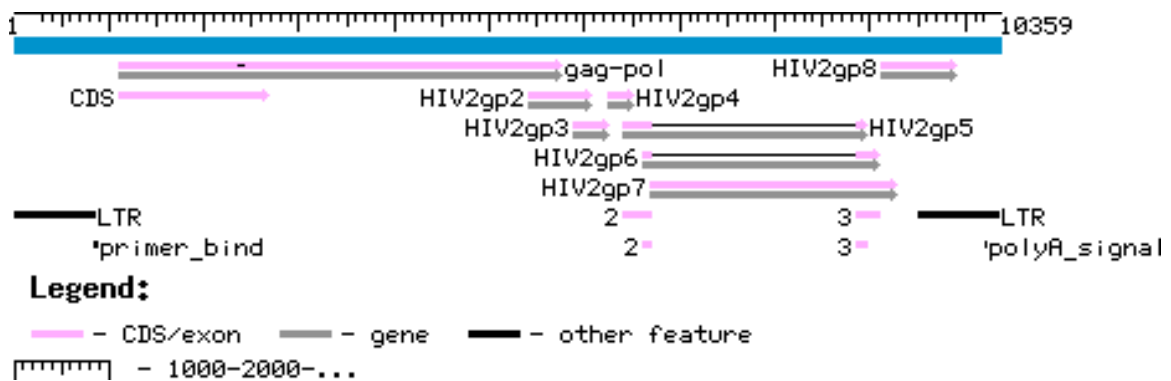
**Relationships Between SIV and HIV  
Determined by the research of Wain-Hobson, 1998.**



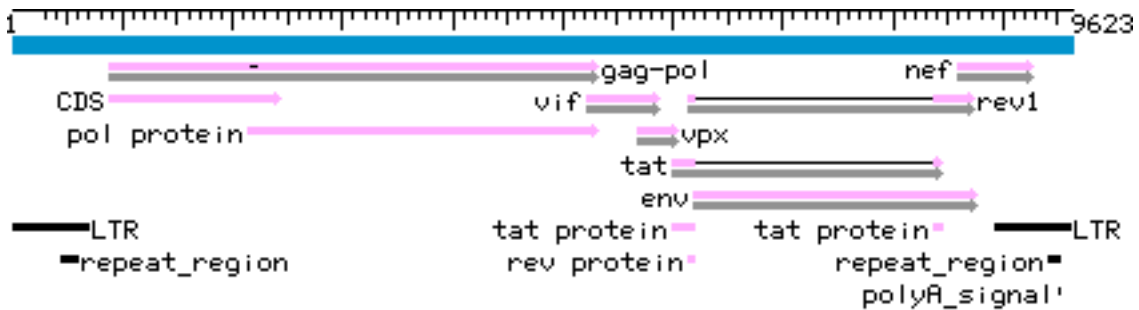
## HIV-1



## HIV-2



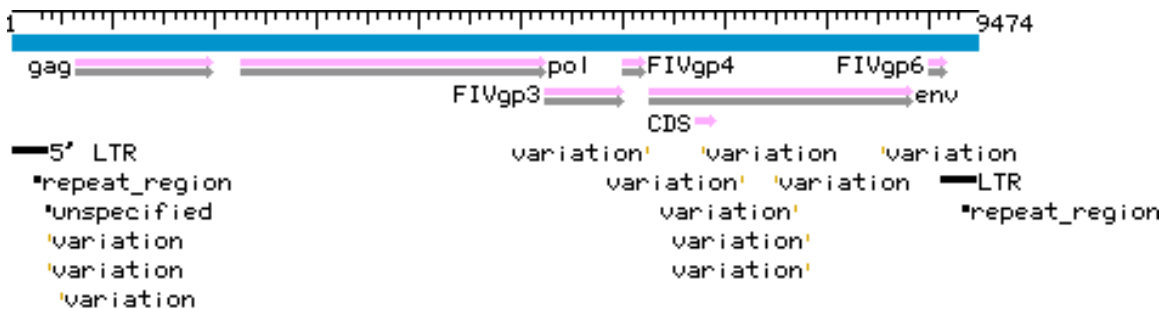
## SIV



### Legend:

→ - CDS/exon   
 → - gene   
 █ - other feature  
 [|||||] - 1000-2000-...

## FIV (Feline Immunodeficiency Virus)

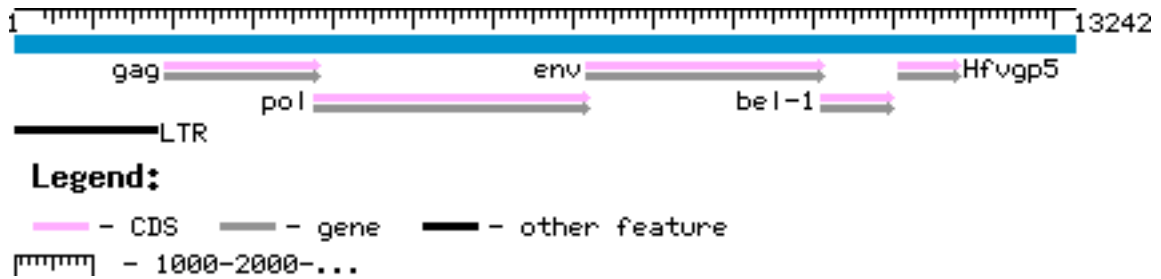


### Legend:

→ - CDS   
 → - gene   
 █ - region   
 █ - other feature  
 [|||||] - 1000-2000-...



## Human Foamy Virus



## Human T-Lymphotropic Virus

