# Murder by HIV? Grades 9–12 Edition

*by*
Laura B. Regassa, Department of Biology, Georgia Southern University
Naowarat (Ann) Cheeptham, Department of Biological Sciences, Thompson Rivers University
Michèle I. Shuster, Department of Biology, New Mexico State University

## Introduction

HIV-1 mutates very rapidly. Because of its high mutation rate, the virus will continue to change (evolve) after a person is infected. Thus, within an infected individual, there may be multiple variants of the virus, all of which diverged from the same strain since the time of infection. Similarly, if many people were all infected by a common source (the same infected individual), over time we would expect to see different sequence variants arise in each infected individual but for all those variants to be genetically related to one another. We can use the genetic sequences to generate a phylogenetic tree and test hypotheses about the genetic (and evolutionary) relationships between different viral strains.

### *Student Background Knowledge*

Students should have the following knowledge prior to completing this activity:

1. Know how to use a web browser.
2. Have a basic understanding of the function of DNA, RNA, and proteins.
3. Be familiar with the ways in which scientists traditionally classify organisms.

### *Vocabulary*

*Bioinformatics:* the unified discipline formed from the combination of biology, computer science, and information technology.

*GenBank:* an open access sequence database that has the collection of all publicly available nucleotide sequences and their protein translations.

*Phylogeny:* a branching diagram or "tree" showing the evolutionary relationships among various organisms based upon their overall similarities and differences.

### *Materials Checklist*

Access to a laptop or desktop computer.

## Part I – An Ill-Fated Argument

In July of 1994, a nurse broke off her relationship with her married boyfriend (a doctor). On August 4, 1994, her ex-boyfriend showed up at her residence and administered a shot that he claimed was a vitamin B-12 injection. He had given her vitamin B-12 injections in the past, but this one was very painful. Prior to that time, the nurse had had several HIV tests (each time she gave blood, and one after having the saliva of an infected patient splash on her skin) and she had always tested negative. Her most recent blood donation was in April of 1994, and her blood tested negative for HIV at that time. In January of 1995, however, she tested positive for HIV. At that time, she accused her ex-boyfriend of deliberately infecting her during the argument back in August. He was brought to trial on charges of attempted second-degree murder.

### *Questions*

1. You can imagine that the defense team posed alternative means by which the woman could have become infected. What are some other possibilities? List them below.

2. What kinds of tests or information could be used to rule out these alternative hypotheses for her infection with HIV?

## Part II – Comparing Sequences

Other possible sources of the infection included the woman's prior sexual contacts and occupational exposure, given that she was a nurse.

All seven of the men that she had been in sexual contact with (including her former boyfriend) were tested and found to be HIV-negative.

Her employment records were examined, and there were no reports of any accidental or occupational exposures other than the saliva that was splashed on her skin sometime in the mid-1980s. Her file did not have any documentation of any needle sticks at work.

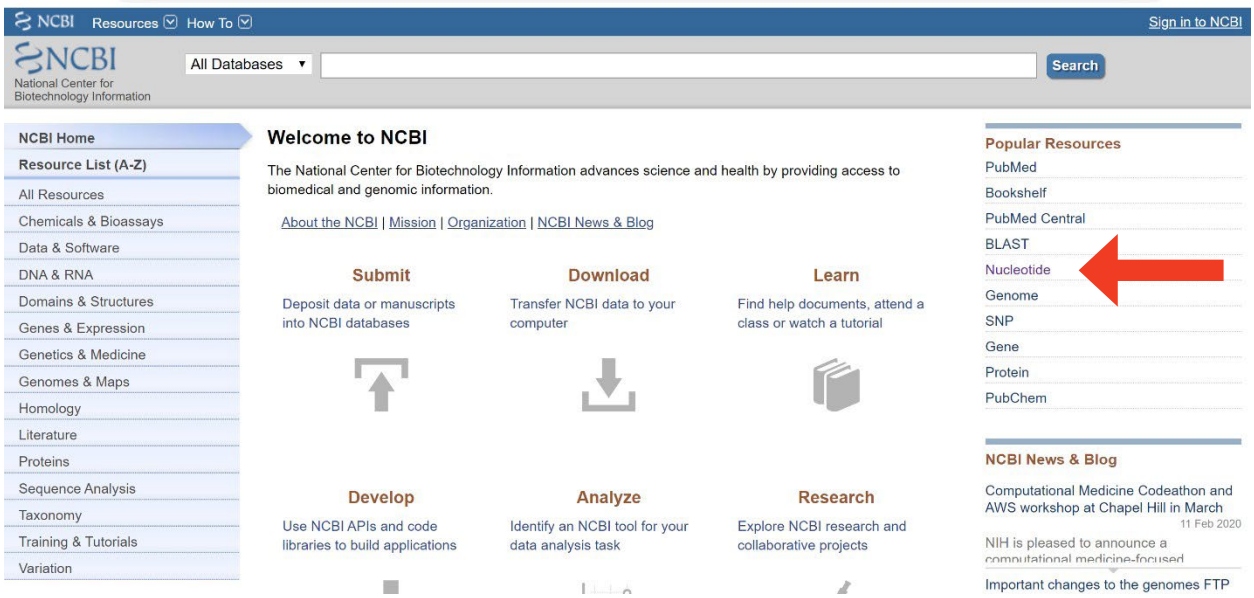
As the investigation proceeded, it was found that an HIV-positive patient under the care of the ex-boyfriend/doctor had blood drawn at the physician's offices on August 4, 1994. The paperwork for this procedure was deliberately hidden (the log book was found in a box of "1982 records" in a storage room with other records from the 1980s) and was not filled out in a manner that was consistent with standard office practices.

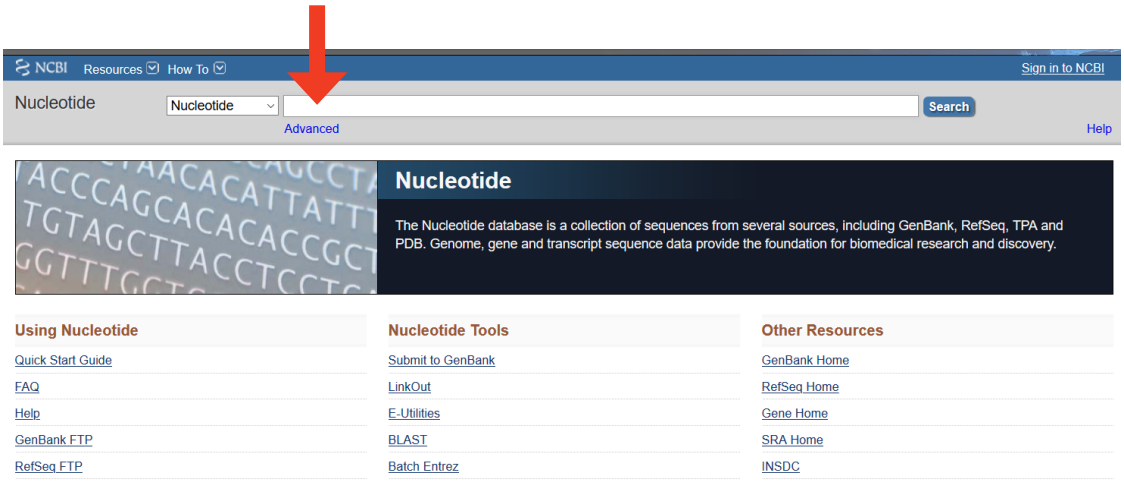
Based on the circumstantial case against the physician, the reverse transcriptase (RT) sequences from the victim (the nurse/ex-girlfriend) and from the physician's HIV-positive patient (the putative source of the nurse's infection via the injection administered during the ill-fated argument) were analyzed.

As HIV-1 mutates rapidly, we don't expect to find identical sequences in the victim and patient. Instead, we expect to find related sequences that share a common ancestor. We can investigate this by using patient and victim HIV RT sequences to generate a phylogenetic tree and look at the clustering of the sequences.

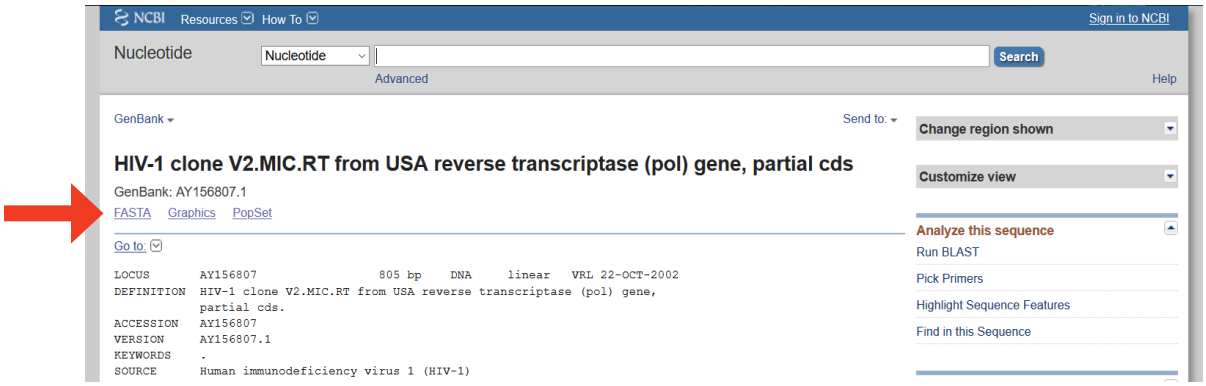
### *Procedure*

1. Go to the NCBI homepage (http://www.ncbi.nlm.nih.gov/). On the right toolbar, select **Nucleotide** (see red arrow in screen capture below) and then search for **AY156807**(see top of next page for where to enter it). AY156807 is the accession number for a reverse transcriptase gene sequence from an HIV isolate. The accession number is a way to locate or reference the sequence, like a book's call number in a library card catalog.

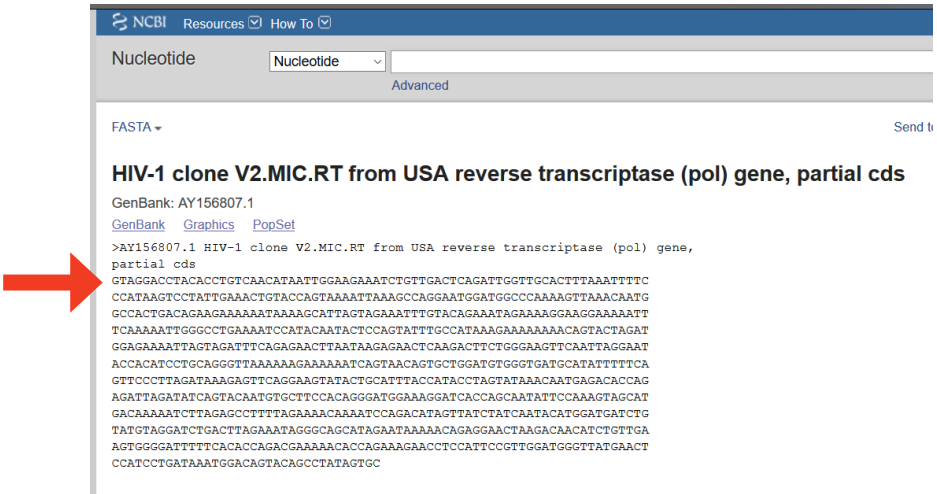




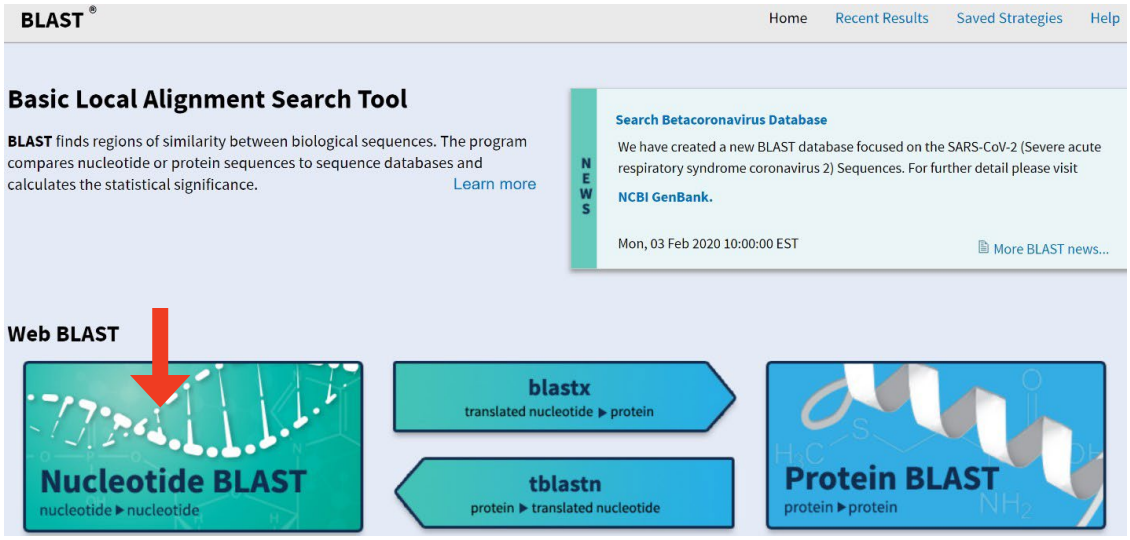

2. When you get to the page that opens with all of the record information, look near the top of the page and click on **FASTA**. FASTA is a format for DNA sequences that is compatible with programs used for bioinformatics analysis.

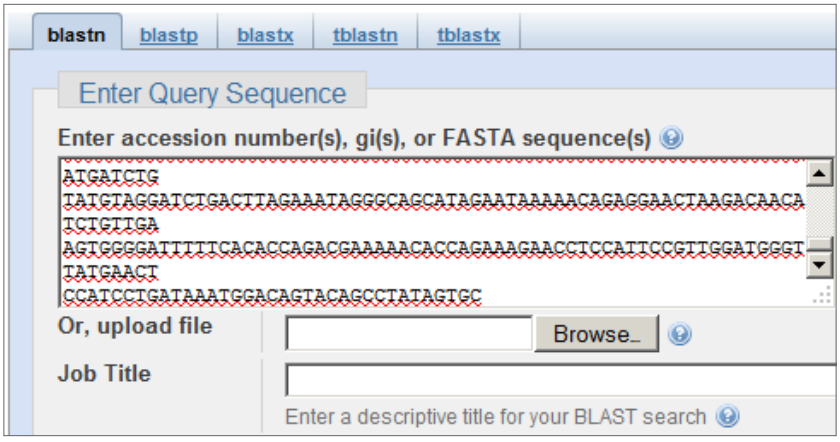



3. You will get the complete nucleotide sequence of that particular sequence. Highlight it and copy it (ONLY the sequence, not the blah blah at the beginning). Remember, the DNA sequence will only have four different letters representing the four nucleotides (A, T, C, G).

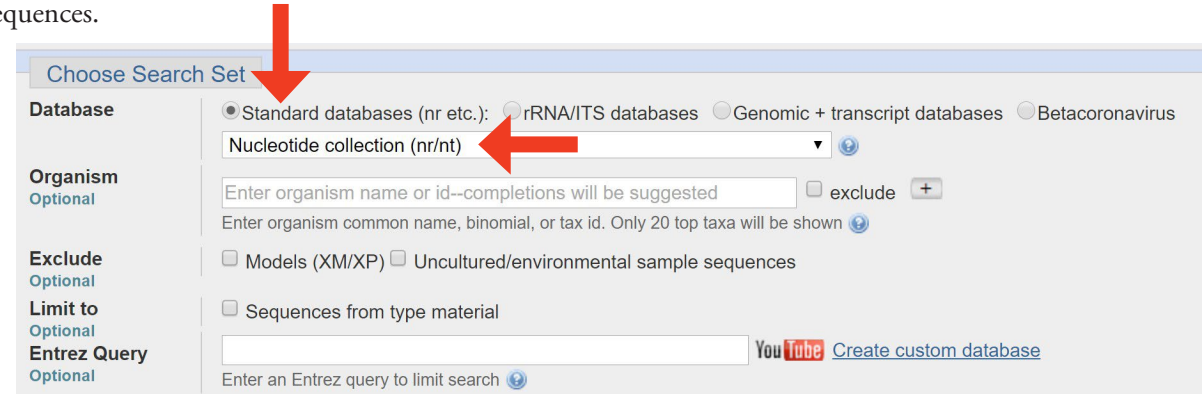


4. Now go back to the NCBI homepage and click on **BLAST** on the right. BLAST is a program that allows you to search for similar DNA sequences in a large database of sequences. When you get to the BLAST homepage, click on the "Nucleotide BLAST" link (see the green box below). When you get to the nucleotide blast page, paste your sequence into the top box (**Enter Query Sequence**).

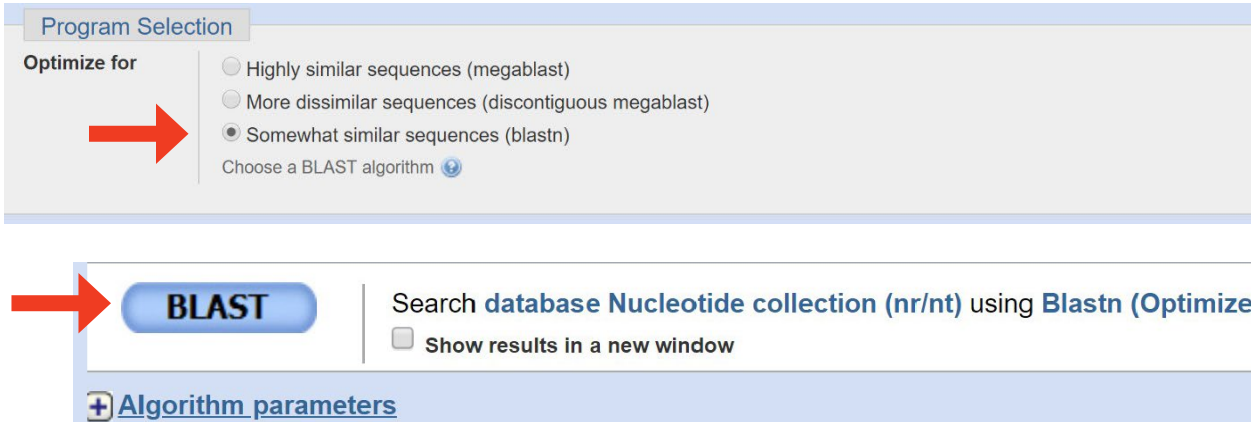


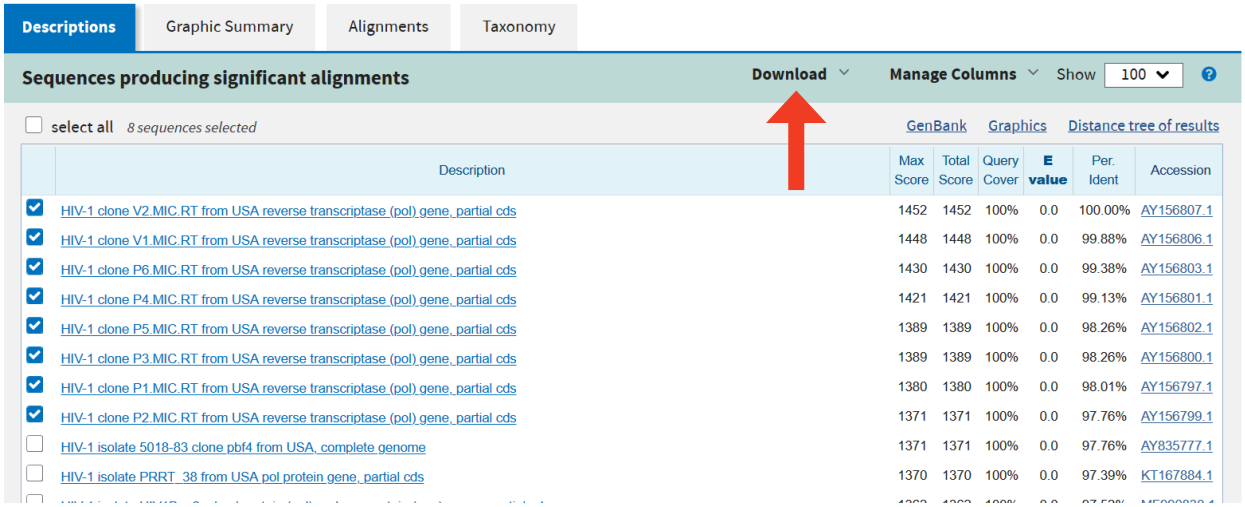

5. Under "**Choose Search Set,**" select "**Standard database (nr etc.)**" for the database and "**nucleotide collection (nr/nt)**" from the database dropdown menu. Some databases have just a subset of all available sequences (e.g., Human genome), but we are looking in a much larger database collection that includes viral nucleotide sequences.



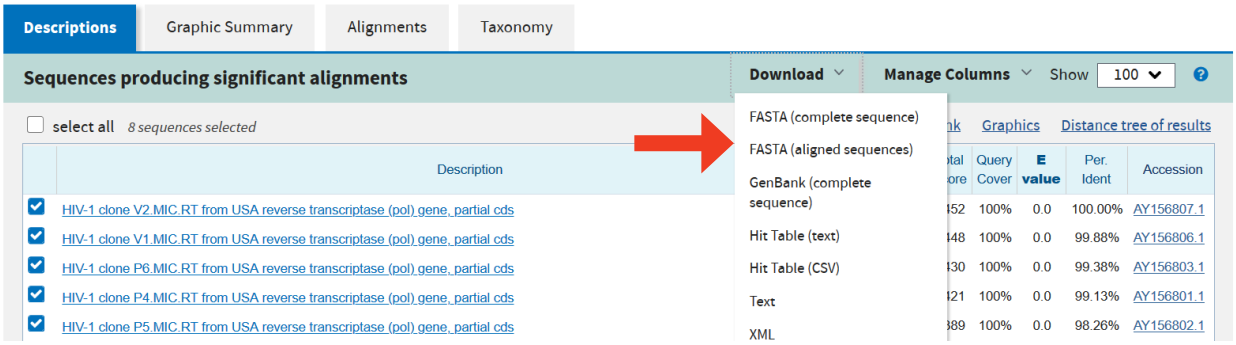
6. Under Program Selection, **Optimize for Somewhat similar sequences (blastn)**, then click on the BLAST button:




7. After a few moments, you will get a list of "hits" that have nucleotide similarities to your Query sequence (from the victim). Scroll past the graphical representation and the abbreviated list by accession number until you get to the listing of individual sequences. The most similar sequences will be listed first. Not surprisingly, the top hits are patient and victim sequences from this case. Click in the first 8 boxes (to check them) of the victim sequences (e.g., HIV-1 clone V2.MIC.RT) and patient sequences (e.g., HIV-1 clone P6-MIC-RT). You will have a total of 8 boxes checked.
8. Once you have selected the sequences you want to compare, click on the "download" button at the top.




9. A menu will appear—select "FASTA (complete sequence)."



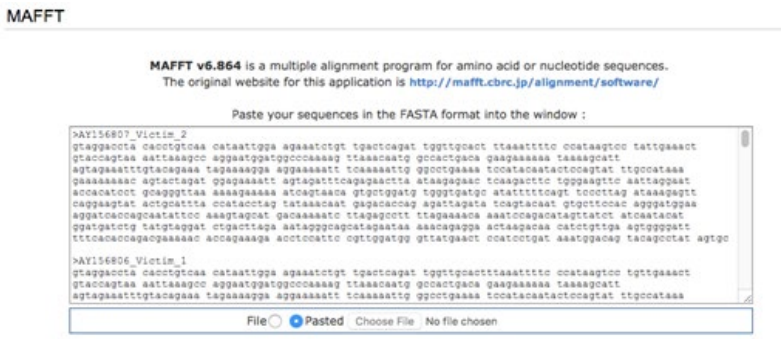
10. Save the resulting sequences into a text document (use Microsoft Notepad or Word, but save file as a txt file, *not* doc). Now you need to re-name the sequences to a shorter name that will be visible on the final phylogenetic tree. When re-naming the sequences, it is important to preserve the FASTA format. To ensure that you are preserving the FASTA format, be sure to keep the ">" at the start of the name of each sequence, and to use "_" instead of spaces (for example, >LA_1). Also remember that V represents HIV RT sequences from the victim and P represents HIV RT sequences from the patient. So, V2.MIC.RT can be changed to victim_clone2. The sequence title (e.g., V2.MIC.RT) is within the first line of each record that you copied (see the example below).

    Example:
    *change*
    >gi|24210021|gb|AY156807.1| HIV-1 clone V2.MIC.RT from USA reverse transcriptase (pol) gene, partial cds
    *to*
    >victim_clone2

11. Add six reverse transcriptase gene sequences from HIV isolates not related to the case to the text document; all of these HIV samples were isolated from patients in the U.S., but had no known connection to the case under study. You will need to retrieve each of these sequences using the nucleotide search engine (see Step 1) and then change the sequence to FASTA (text) format (see Step 10). Copy the sequence title line and title to your document that already has eight sequences from the victim and patient. Change the title for each of the sequences as indicated (e.g., change **HIV-1 isolate 5018-83 clone** to **USA_1**).

| USA Isolate | New title |
|---|---|
| AY835777 | >USA_1 |
| AY835778 | >USA_2 |
| AY835769 | >USA_3 |
| AY156793 | >LA_1 |
| AY156789 | >LA_2 |
| AY156788 | >LA_3 |

12. Go to **http://www.trex.uqam.ca**.
13. On the left hand menu, click on MAFFT.
14. Copy your sequences (from step 11), and past them into the window:




15. Click on the "compute" button beneath the sequence window.
16. Wait ...
17. You will see what looks like a mostly blank screen called "Results for MAFFT." Click on the "View Tree" button on the left.
18. The tree will appear in a new window on the same page.

## *Questions*

1. Describe the tree (in general terms). Draw a quick sketch of the tree.

2. Does there appear to be a relationship between the patient and victim sequences? Do they appear to diverge from a common ancestor?

3. What conclusion can you draw from this tree?

4. Given the circumstantial evidence and the phylogenetic evidence, what do you think the verdict was in this case?