# Murder by HIV? Grades 9–12 Edition

## бу

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 III-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

 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.



SNCBI Resources 🛛 How To 🔍		Sign in to NCBI
Nucleotide Nucleotide Advanced		Search
ACCCACACACATT	Nucleotide	
TGTAGCTTACCTCCTC	The Nucleotide database is a collection of sequences from s PDB. Genome, gene and transcript sequence data provide t	several sources, including GenBank, RefSeq, TPA and he foundation for biomedical research and discovery.
Using Nucleotide	Nucleotide Tools	Other Resources
Quick Start Guide	Submit to GenBank	GenBank Home
FAQ	LinkOut	RefSeg Home
Help	E-Utilities	Gene Home
GenBank FTP	BLAST	SRA Home
RefSeq FTP	Batch Entrez	INSDC

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.

Nucleolide	Nucleotide  V	Search	
_			
GenBank -	Send to: •	Change region shown	
HIV-1 clo	one V2.MIC.RT from USA reverse transcriptase (pol) gene, partial cds		
GenBank: AY	156807.1	Customize view	
EASTA Gran			
TASIA Ola	hics PopSet		
Go to: 🕑	hics PopSet	Analyze this sequence Run BLAST	
	hirs PopSet AY156807 805 bp DNA linear VRL 22-OCT-2002	Analyze this sequence Run BLAST Pick Primers	
Go to: 🔍 LOCUS DEFINITION	h <u>ics PopSet</u> AY156807 805 bp DNA linear VRL 22-CCT-2002 HIV-1 clone V2.MIC.RT from USA reverse transcriptase (pol) gene, partial cds.	Analyze this sequence Run BLAST Pick Primers Highlight Sequence Features	
Go to: V LOCUS DEFINITION ACCESSION VERSION	hics PopSet AY156807 805 bp DNA linear VRL 22-OCT-2002 HIV-1 clone V2.MIC.RT from USA reverse transcriptase (pol) gene, partial cds. AY156807 AY156807.1	Analyze this sequence Run BLAST Pick Primers Highlight Sequence Features Find in this Sequence	

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).

S NCBI Resourc	rces 🗹 How To 🗹	
Nucleotide	Nucleotide ~	
	Advanced	
FASTA 🗸		Send t
HIV-1 clone	e V2.MIC.RT from USA reverse transcriptase (pol) gen	e, partial cds
GenBank: AY15680	807.1	
GenBank Graphics	ics PopSet	
>AY156807.1 HIV-	V-1 clone V2.MIC.RT from USA reverse transcriptase (pol) gene,	
GTAGGACCTACACCTG	TGTCAACATAATTGGAAGAAATCTGTTGACTCAGATTGGTTGCACTTTAAATTTTC	
CATAAGTCCTATTG	GAAACTGTACCAGTAAAATTAAAGCCAGGAATGGATGGCCCAAAAGTTAAACAATG	
GCCACTGACAGAAGAA	AAAAAATAAAAGCATTAGTAGAAATTTGTACAGAAATAGAAAAGGAAGG	
TCAAAAATTGGGCCTG	rgaaaatccatacaatactccagtatttgccataaagaaaaaaaa	
GGAGAAAATTAGTAGA	GATTTCAGAGAACTTAATAAGAGAACTCAAGACTTCTGGGAAGTTCAATTAGGAAT	
ACCACATCCTGCAGG		
GTTCCCTTAGATAAAG	AGAGTTCAGGAAGTATACTGCATTTACCATACCTAGTATAAACAATGAGACACCAG	
CACADADATATCAGT	TRUKATGTGUTTUUKUKGGGATGGAAAGGATUKUUKGUKATATTUUKAAGTAGUAT	
GACAAAAAICIIAGAG		
TATCTACCATCTCACT	CI INDAMI INDOGENICI INDAMI INAAMENDAGONACI ANDACAACAI CI UI I DA	
TATGTAGGATCTGACT	ACACCAGACGAAAAAACACCAGAAAGAACCTCCATTCCGTTGGATGGGTTATGAACT	

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**).

BLAST <sup>®</sup>	Home Recent Results Saved Strategies Help
Basic Local Alignment Search Tool BLAST finds regions of similarity between biological sequences. The program compares nucleotide or protein sequences to sequence databases and calculates the statistical significance.	Search Betacoronavirus Database         We have created a new BLAST database focused on the SARS-CoV-2 (Severe acute respiratory syndrome coronavirus 2) Sequences. For further detail please visit         NCBI GenBank.         Mon, 03 Feb 2020 10:00:00 EST
Web BLAST	stx otide ▶ protein Astn lated nucleotide
blastn         blastn         blastx         tblastn           Enter Query Sequence           Enter accession number(s), gi(s), or           ATGATCTG           TATGTAGGATCTGACTTAGAAATAGGGCAGGCA           ISIGEGGATTITTCACACCAGACGAAAACAC           TATGAACT           CCATCCTGATAAATGGACAGTACAGCCTATAC           Or, upload file           Job Title	tblastx   FASTA sequence(s) ③ ATAGAATAAAAACAGAGGAACTAAGACAACA CCAGAAAGAACCTCCCATTCCGTTGGATEGGT GTGC Browse_ ② ve title for your BLAST search ④

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.

Choose Searc	ab Set
-	
Database	Standard databases (nr etc.):      PrRNA/ITS databases      Genomic + transcript databases      Betacoronavirus
	Nucleotide collection (nr/nt)
Organism	
Optional	Enter organism name or idcompletions will be suggested
	Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown 😡
Exclude Optional	□ Models (XM/XP) □ Uncultured/environmental sample sequences
Limit to	Sequences from type material
Entrez Query	You Tube Create custom database
Optional	Enter an Entrez query to limit search 🤢

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

Optimize for	<ul> <li>Highly similar sequences (megablast)</li> <li>More dissimilar sequences (discontiguous megablast)</li> </ul>
-	Somewhat similar sequences (blastn) Choose a BLAST algorithm (g)
	Search database Nucleotide collection (pr/pt) using Plasta (Optimize

- 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.

Desci	riptions	Graphic Summary	Alignments	Taxonomy							
Sequ	iences pr	oducing significant a	lignments		Download 🗡	Mana	ge Col	umns	⊻ Sh	iow 10	0 🗸 🕜
s	elect all 8	sequences selected			<b>1</b>	Gen	Bank	<u>Grap</u> l	nics [	Distance t	ree of results
			De	scription		Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
	HIV-1 clone \	/2.MIC.RT from USA reverse tra	anscriptase (pol) gene,	partial cds		1452	1452	100%	0.0	100.00%	AY156807.1
	HIV-1 clone \	/1.MIC.RT from USA reverse tra	anscriptase (pol) gene,	partial cds		1448	1448	100%	0.0	99.88%	AY156806.1
	HIV-1 clone F	P6.MIC.RT from USA reverse tra	anscriptase (pol) gene,	partial cds		1430	1430	100%	0.0	99.38%	AY156803.1
	HIV-1 clone F	24.MIC.RT from USA reverse tr	anscriptase (pol) gene,	partial cds		1421	1421	100%	0.0	99.13%	AY156801.1
	HIV-1 clone F	25.MIC.RT from USA reverse tra	anscriptase (pol) gene,	partial cds		1389	1389	100%	0.0	98.26%	AY156802.1
	HIV-1 clone F	23.MIC.RT from USA reverse tr	anscriptase (pol) gene,	partial cds		1389	1389	100%	0.0	98.26%	AY156800.1
	HIV-1 clone F	P1.MIC.RT from USA reverse tra	anscriptase (pol) gene,	partial cds		1380	1380	100%	0.0	98.01%	AY156797.1
	HIV-1 clone F	2.MIC.RT from USA reverse tr	anscriptase (pol) gene,	partial cds		1371	1371	100%	0.0	97.76%	AY156799.1
	HIV-1 isolate	5018-83 clone pbf4 from USA,	complete genome			1371	1371	100%	0.0	97.76%	AY835777.1
	HIV-1 isolate	PRRT_38 from USA pol proteir	<u>gene, partial cds</u>			1370	1370	100%	0.0	97.39%	KT167884.1
						4000	4000	4000/	0.0	07 500/	ME000000 4

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

Descrij	ptions	Graphic Summary	Alignments	Taxonomy						
Seque	ences pr	oducing significant a	lignments		Download 🖌 🛛 Manage	e Col	umns	✓ Sh	10w 10	00 🗸 🕜
🗌 sel	lect all 8	sequences selected			FASTA (complete sequence)	<u>nk</u>	Grap	hics I	Distance t	ree of results
			De	escription	FASTA (aligned sequences) GenBank (complete	otal ⇔ore	Query Cover	E value	Per. Ident	Accession
И н	IV-1 clone \	2.MIC.RT from USA reverse tra	anscriptase (pol) gene,	partial cds	sequence)	152	100%	0.0	100.00%	AY156807.1
М н	IV-1 clone \	1.MIC.RT from USA reverse tra	anscriptase (pol) gene,	partial cds	Hit Table (text)	148	100%	0.0	99.88%	AY156806.1
н	HV-1 clone F	P6.MIC.RT from USA reverse tra	anscriptase (pol) gene,	partial cds	Hit Table (CSV)	<b>‡30</b>	100%	0.0	99.38%	<u>AY156803.1</u>
М н	IV-1 clone F	24.MIC.RT from USA reverse tra	anscriptase (pol) gene,	partial cds	Text	121	100%	0.0	99.13%	<u>AY156801.1</u>
<u>н</u>	HV-1 clone F	25.MIC.RT from USA reverse tra	anscriptase (pol) gene,	partial cds	XML	389	100%	0.0	98.26%	<u>AY156802.1</u>

"Murder by HIV? Grades 9-12 Edition" by Regassa, Cheeptham, and Shuster

- 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:

MAFFT v6.864 is a multiple alignment program for amino acid or nucleotide sequences. The original website for this application is http://mafft.cbrc.jp/alignment/software/	
Pasta your sequences in the EACTA format into the window :	
Past four sequences in the Post four sequences in the Post formats into the Mindow . Past 5600 _Victim_2 gtaggacots cacotybicas catastiggs agaastigt testicagat tegitigoet traastitic costaagtot tatigaasti ptacogias attasspic aggaasta agigaissast toosaastig ggacatgas gaagaasas taasgot tatigaasti pasagaastigtacogias tagaasgga aggaasaatt toosaastig ggootgaas tootacogiasticoogiast tipocataas gaasaast gitagaastigat ggaasaatt toosaastig ggootgaas tootacogiasticoogiast tipocataas gaasaast gitagaastigat ggootgaasa tagagatta coppastit actopatta costacida tatascost ggacacog agatagast costacogiastico attagasti coppastat actopatta costacida tataspost tagagatt tagastast costacogiastica tatascost ggatgacog tagatta costacida tataggoga tataggoga costagatas cotogias tatagatost ggatgacog tagottaga tegattaga actopatta costacogi tataggoga tatagacos catocidat astoggoga tatagatos ggatgacog tagottaga tegattaga actopatta costogi tagotgasta saccoggoga taagotast attoggi tagogotta agtogo titoscogagogaasaa cocogaasa actocidito costogi ggatgi gatagasto costoci taastogi taagota tagota tastogi tagota tagota	
>AY156006_Victim_1 gtaggacota caccegtcaa cataattega agaaatetgt teestcagat tegttepaatttaaatttto coataagtoo tgttepaaset gtaccegtaa astataagoo aggaatggategocoaaaag ttaascaatg gocactgaca gaagaasaa taasagoatt agtagaastttgtacegaaa tagaaaagga aggaaaastt tosassattg ggoctgaaka tocatscaatactocegtat ttgccataas	Results for MAFF
File Pasted Choose File No file chosen	
Click on the "compute" button beneath the sequence window.	Input file(s)
Click on the "compute" button beneath the sequence window. Wait	Input file(s)
Click on the "compute" button beneath the sequence window. Wait You will see what looks like a mostly blank screen called "Results for MAFFT	Input file(s) Input data
Click on the "compute" button beneath the sequence window. Wait You will see what looks like a mostly blank screen called "Results for MAFFT Click on the "View Tree" button on the left.	Dutput file(s)
Click on the "compute" button beneath the sequence window. Wait You will see what looks like a mostly blank screen called "Results for MAFFT Click on the "View Tree" button on the left. 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?

S

Credit: Photo in title block © kozini - Fotolia.com, 1D#47098219. Case copyright held by the **National Center for Case Study Teaching in Science**, University at Buffalo, State University of New York. Originally published January 29, 2013. Please see our **usage guidelines**, which outline our policy concerning permissible reproduction of this work.