Tracing the Source of Infection for HIV using Evolutionary Changes in DNA

Written By: Cami Senior
Section 001
2016

ABSTRACT

HIV is a highly infectious disease the rapidly evolves to create different strains with different levels of drug resistance. Because of the rapid evolution however, it can be easy to trace back the origins of a strain. Meaning, that there is the ability to trace back which individual infected which. Which is the point of this experiment, to compare DNA sequences of several infected people to find out who infected the patient of interest, and then creating a phylogenetic tree. The hypothesis made at the beginning was that nurse #2 had infected the patient. And by comparing the DNA sequences and finding the percent match is was discovered that nurse #2 had the most similar strain of HIV as the patient's. Then by looking at the constructed phylogenetic tree, it was seen that nurse #2's strain and the patient's strain often close together in the phylogenetic tree, sometimes on their own branch. By observing this data the conclusion was made that nurse #2 was the one to infect the patient, proving the initial hypothesis.

INTRODUCTION

HIV stands for the human immunodeficiency virus, which is a virus that causes AIDS, or acquired immune deficiency syndrome. The thing that makes HIV so famous is the fact that it rapidly evolves drug resistance, making it very difficult to treat.

To begin, let's look at its history. AIDS was first recognized in 1981. Initially doctors thought it a weird case of pneumonia and cancer among the gay population. Soon after this was linked to HIV. At this time medical developments like vaccines and antibiotics had managed to get other infectious disease under control, so many people didn't fear AIDS as much because they believed that in a couple years there would be a vaccine for AIDS. Obviously that didn't happen. By 2012 HIV had infected 65 million people, with about 30 million dead due opportunistic infections that could get to them with the help of AIDS. But where did this virus originate? By looking at the virus's genome and tracing back similar strains scientist have come to a hypothesis. The genome and life cycle is very similar to that of simian immunodeficiency viruses, which are viruses that affect a variety of primates. This means that the viruses have jumped from primates to humans, and it appears that the virus has jumped back and forth between chimps and humans at least three times. (This can be seen by the three groups of HIV-1: group O, group P, and group M. With group M being responsible for 95% of HIV infections)

But how is it transmitted? HIV is transmitted when a bodily fluid, which can be semen, vaginal and rectal secretions, blood, or breast milk, that contains the virus, meaning coming from the infected individual, comes in contact with a mucous membrane or into the bloodstream of an uninfected individual. The most common activities that spread the disease are sex, homosexual, heterosexual, or oral, needle sharing, blood transfusions with infected blood, childbirth, breastfeeding, and unsafe medical procedures. There are many choices out there to help prevent infection, such as antiviral drugs for infected mothers soon to give birth, circumcision for men (which reduces the risk by half), use of condoms, or general avoidance of risky behavior.

But what exactly is HIV? Technically HIV is an intracellular parasite that is incapable of reproducing on its own. Which means to reproduce it must infect and hijack cells, and all enzymes and materials that it contains, to reproduce itself. And that would be ok, if in the process it didn't also kill the host cell. The life cycle of HIV is this; First it is in it extracellular form, which contains 2 copies of RNA genome, reverse transcriptase, integrase, and protease, and it lined by gp41 anchor protein and gp120 surface protein. From here the gp120 surface protein binds to the host cell's CD4 and its coreceptor. After this the HIV membrane and the host cell membrane fuse, releasing all of the HIV's contents into the cell. This is now when the virus uses the host cell to replicate. It begins with the reverse transcriptase enzyme synthesizing DNA from HIV's RNA genome; this is then integrated into the host cell's DNA by integrase. Then the host cell's own RNA polymerase reads this integrated HIV DNA and transcribes it to mRNA, and then again the host cell's own ribosomes translated the HIV mRNA into HIV precursor proteins. Now that the host cell has made all the parts, a new generation of HIV virus can assemble, and bud from the host cell's membrane. But the virus isn't complete until the protease in the HIV cleaves the precursor protein made by the host cell, which finally ends with a mature HIV virion. This virion can then go onto infect other host cells.

A side note, of how HIV turns into AIDS. HIV attacks cells by binding with host cells' surface proteins CD4 and coreceptor. There are various strains of HIV, the can bind to various coreceptors, but mostly the bind to proteins called CCR5. The cells that most commonly contain CCR5 and CD4, meaning cells that HIV can attack, are macrophages, effector helper T cells, and memory helper T cells, which are

used in the immune system to fight off viruses. There are three phases of HIV infection. The first phase is the acute phase in which the HIV virus enters the body and rapidly replicated, which causes the concentration of T cells that contain CD4 to very quickly plummet. The individual at this stage generally show symptoms of a viral infection. This phase begins to end when the replication of HIV slows, and the T cells are able to recover a bit. The individual then enters the chronic stage. HIV virions continue to replicate, but at a slower, more steady pace, at the same time the concentration of CD4 T-cells steadily falls. The individual tends to not show any symptoms at this stage. Finally the individual enters the third stage, the AIDS stage. This is reach when the concentration of CD4 T-cells in the blood stream reaches a critical low of 200 cells per cubic millimeter. The immune system has begun to collapse, and opportunistic virus may attack. If not treated with anti-HIV drug therapy the individual will likely live for less than three years.

Now how do we treat it? Or at least postpone its final effect of death? Most patients when diagnosed with HIV begin HIV therapy. HIV therapy consists of exposing the individual to a cocktail of drug. Initially there was just one drug that was used to fight HIV. This drug was AZT, or Azidothymidine, which resembles the nucleotide thymidine, except for the replacement of a hydroxyl group with an azide group. This drug was targeted at the reverse transcriptase, by permanently binding it up with AZT at the azide group site, and initially it worked. But over time patients stopped responding, and began to have a decreasing concentration of T cells. This happened because over time populations of virions evolved resistance to AZT. That is, they evolve mutations in their genes that made for an altered reverse transcriptase that is able to remove the AZT that it added, and correct the mistake. So over time the virions without the mutation were killed off, while the virions with it were free to replicate, until all that was left was the mutated virions. This mutation likely occurred because reverse transcriptase is prone to making mistakes in its transcription. This is a case of evolution by natural selection, and this is what makes it so difficult to treat individuals with HIV. Nowadays we have several drugs that target different parts of the virion such as, coreceptor inhibitors, fusion inhibitors, reverse transcriptase inhibitors, integrase inhibitors, and protease inhibitors.

This rapid evolution of the HIV leads to different strains that have different drug resistance. However, this also mean that by comparing the DNA similarities one can trace the origin of HIV between different individuals that might of infected each other. With this data we can create an evolutionary tree, both within an individual, and within a population of infected individuals. And that is exactly what is done in this experiment. To begin we are presented with a patient who has been diagnose with HIV, and must figure out whom the patient got HIV from. We were also given DNA from various other infected individuals in the populations. My hypothesis for this set of data is that the patient was infected by nurse #2

MATERIALS AND METHODS

To begin this experiment three sets of data were presented. Each set of data contained DNA that corresponded to the HIV virus in several individuals from a different loci in HIV. This data was then copied into the tool 'MUSCLE', or '**MU**ltiple **S**equence **C**omparison by **L**og- **E**xpectation', on the website for 'The European Bioinformatics Institute'. This tool then ran the DNA and aligned all the samples, and by comparing the differences and similarities, constructed a probable phylogenetic tree. Then each was compared to come to a conclusion on who infected the patient.

RESULTS

Table 1: Alignment of DNA env
The alignment can be found at: http://www.ebi.ac.uk/Tools/services/rest/muscle/result/muscle-I20160402-224908-0416-49707767-oy/aln-clustalw A "*" means that all the DNA samples have the same nucleotide at that spot, while a " " (blank space) means there are different nucleotides.

Table 2: Percent Match. The percent of overall sequence that matches the patient for env	
Individual	Percent match
Nurse #1	75.68%
Nurse #2	94.76%
Person #1	84.08%
Person #2	84.63%
Person #3	85.81%
Person #4	82.48%
Person #5	85.26%
Person #6	82.48%
Person #7	95.18%
Partner	82.38%

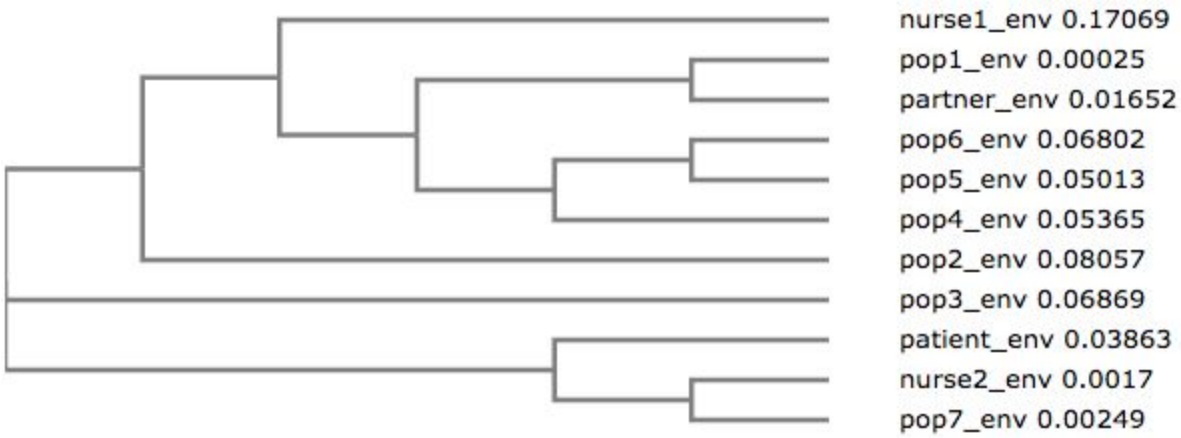

Image 1: Cladograms Phylogenetic tree for env: this is the tree that the tool made by comparing the DNA. Each branch leads to a different individual, and one that are close to each other, and have evolved at later times are closely related strains.

Table 3: Alignment of DNA for pol
The alignment can be found at: http://www.ebi.ac.uk/Tools/services/rest/muscle/result/muscle-I20160402-182825-0138-83181768-es/aln-clustalw A "*" means that all the DNA samples have the same nucleotide at that spot, while a " " (blank space) means there are different nucleotides.

Table 4: Percent Match. The percent of overall sequence that matches the patient for pol	
Individual	Percent match
Nurse #1	87.92%
Nurse #2	94.45%
Person #1	88.08%
Person #2	88.25%
Person #3	94.00%
Person #4	93.60%
Person #5	88.23%
Person #6	88.12%
Person #7	88.19%
Partner	87.58%

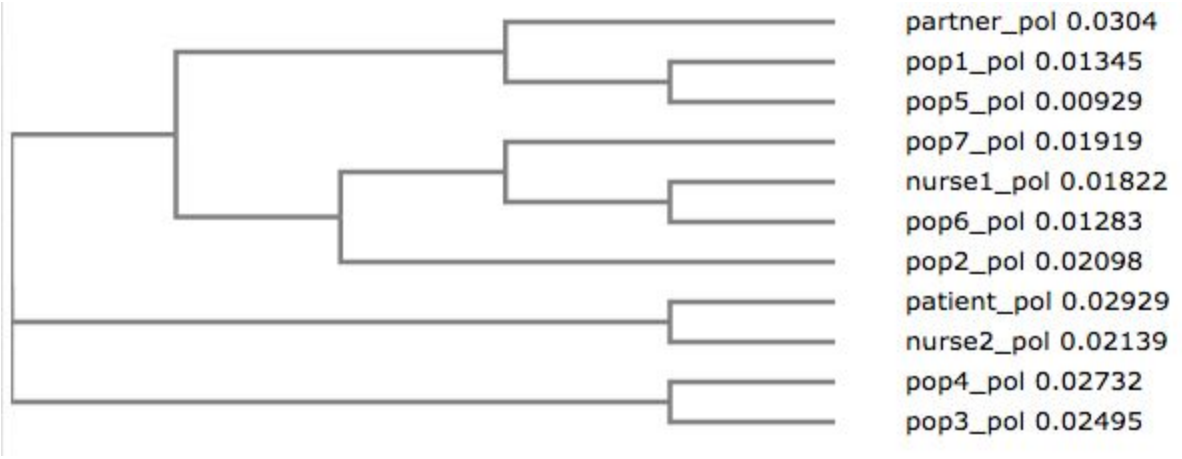

Image 2: Cladograms Phylogenetic tree for pol: this is the tree that the tool made by comparing the DNA. Each branch leads to a different individual, and one that are close to each other, and have evolved at later times are closely related strains.

Table 5: Alignment of DNA for gag
The alignment can be found at: http://www.ebi.ac.uk/Tools/services/rest/muscle/result/muscle-I20160402-225108-0020-45169897-oy/aln-clustalw A "*" means that all the DNA samples have the same nucleotide at that spot, while a " " (blank space) means there are different nucleotides.

Table 6: Percent Match. The percent of overall sequence that matches the patient for gag	
Individual	Percent match
Nurse #1	90.18%
Nurse #2	86.62%
Person #1	87.44%
Person #2	90.85%
Person #3	91.30%
Person #4	90.62%
Person #5	90.85%
Person #6	91.53%
Person #7	91.08%
Partner	91.76%

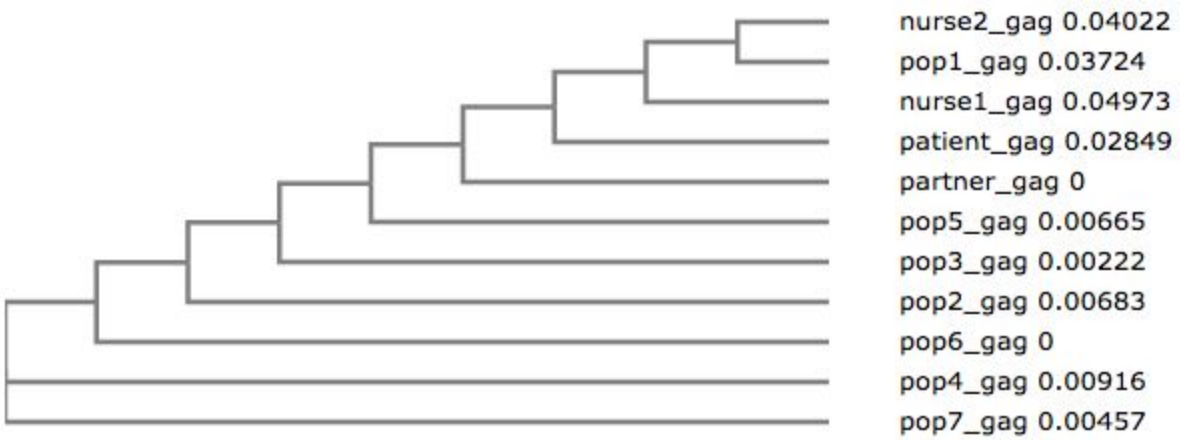

Image 3: Cladograms Phylogenetic tree for gag: this is the tree that the tool made by comparing the DNA. Each branch leads to a different individual, and one that are close to each other, and have evolved at later times are closely related strains.

DISCUSSION

The purpose of this experiment to was to figure out who infected the patient with HIV. My hypothesis from the beginning was that nurse #2 had infected the patient. Through observing the data the hypothesis was proved.

To begin I looked at the alignment of the three different loci. First I noticed that for the alignment of DNA at the gag loci, which is table 5 and 6i, they all had a pretty high match. This led me to believe that this loci was a common to most HIV, and thus meant that it could not be used to identify who infected the patient. Though it did show some change from one person to the next, which can be seen in image 3, but still this data could not help to come to a conclusion.

Next I looked at the alignment for pol. This is where I began to suspect nurse #2. Nurse #2 had the highest match with the patient, which means that it's a similar strain, and that there is a high chance that the patient's infection was from the nurse. Then I looked at the phylogenetic tree, image 2. This provided more information proving my initial hypothesis. It is fairly obvious that both the patient and nurse #2 are on the same branch, separated from everybody else, which, again, means that they have similar strains of HIV.

Finally I looked at alignment for env, which can be found in table 1. I couldn't make much of it until I had a look at the percent of matched sequence in the alignment. This had some conflicting information. The top percent match was to person #7. So which was it, person #7 or nurse #2? I then saw that the second highest match was nurse #2, so this definitely kept them in the running for who infected the patient. To come to a final conclusion I had to look at the phylogenetic tree, image 1. This gave me the final information that I needed. Person #7, nurse #2, and the patient were all on the same branch, which nurse #2 and person #7 being closest. Adding in the data from the pol loci, I realized that nurse #2 had probably infected the patient, and by the looks of it, probably also person #7, that or person #7 had infected the nurse. But ultimately it meant that the infection of HIV in the patient was a most closely related to the strain from nurse #2. To have similar strains that means one has to pass it to the other, and seeing as I am investigating who infected the patient, that means that the nurse was the one who passed it onto the patient.

CONCLUSION

I began this experiment in the hopes of finding who had infected our patient with HIV. From the beginning I suspect nurse #2. And through testing and comparing the DNA of HIV from a selection of infected individuals I came to prove my hypothesis. Now there is a possibility that the nurse #2 did not infect the patient, and that somebody else infected the patient, and then that strain mutated into one that is similar to nurse #2's strain. But that is highly unlikely.

There have been similar experiments such as this. An example would be one of the first in a court case of *Louisiana v. Richard J. Schmidt*, in which a nurse by the name of Janet Trahan had a multifarious relationship with a doctor who was her colleague. This doctor is also known as Dr. Richard Schmidt. Now this wasn't just a one-time thing, it was a 10-year affair, as well as her being his patient. But however, as many affairs do, Trahan ended it, in this case. But she still continued as the doctor's patient and colleague. Soon after the break up Schmidt went to Trahan's apartment to administer her regular vitamin injection. And things turned for the worse. Soon after the injection Trahan had symptoms of a general viral infection, and a couple months after she test positive for HIV. There was a theory that on the day of the injection Dr. Schmidt had drawn extra blood from an HIV patient, and then intentionally injected Trahan with it, ultimately infecting her. To test this theory samples of HIV DNA were taken from Trahan, the patient, and various infected people in the city. These were then compared to see if it was indeed HIV from the patient. Ultimately it was proved that the HIV strain was one that had branched off in the patient, and Dr. Richard Schmidt was prosecuted with second-degree attempted murder.

Ultimately I think this experiment went pretty well. We were given enough information to understand the objective, and how to move from there. I wouldn't change anything about this experiment.

REFERENCES

Freeman, Scott, and Jon C. Herron. "Chapter 1: A Case for Evolutionary Thinking: Understanding HIV." *Evolutionary Analysis*. 5th ed. Upper Saddle River, NJ: Pearson Prentice Hall, 2007. N. pag. Print.

Klug, William S., and Michael R. Cummings. *Essentials of Genetics*. 8th ed. Upper Saddle River, NJ: Pearson Education, 2005. Print.

Karcher, S. and SJA Kimble. (2016) BIOL 24200 Lab 9- Phylogenies I, Lab 10- Phylogenies II

MUltiple Sequence Comparison by Log- Expectations. Computer software. *The European Bioinformatics Institute*. N.P., n.d Web. <http://www.ebi.ac.uk/Tools/msa/muscle/>.