

HIV: Sequencing and Common Ancestry

Introduction

In 1981, a mysterious infection began killing young healthy people in the United States. It took some time, but eventually **epidemiologists**, people who study the causes and spread of disease, solved the mystery of what was causing the illnesses—a virus they named the human immunodeficiency virus, or HIV. In the early 1980's, epidemiologists determined that HIV had spread to the United States from Central Africa, but there was much they still didn't know about it. Thirty years later, researchers are using genetic evidence to better understand when, where, and how HIV evolved into such a deadly disease that affects people around the world.

Viruses are very diverse. Let's say two people have a cold caused by the same virus, rhinovirus. Even though the both have rhinovirus, the rhinovirus making person 1 sick is almost certainly different from the rhinovirus making person 2 sick. Slightly different versions of the same virus are called strains.

In this lesson, you will explore how epidemiologists compare and analyze the genetic information within different strains of HIV. Researchers in the study you're going to read about today sampled HIV strains from hundreds of infected people to reconstruct how the virus mutated and changed as it spread. You will also learn how epidemiologists have used the genetic information to trace all of known HIV strains back to a specific time and location in Africa. Finally, you will see how scientists model evolutionary relationships among viruses using phylogenetic trees.

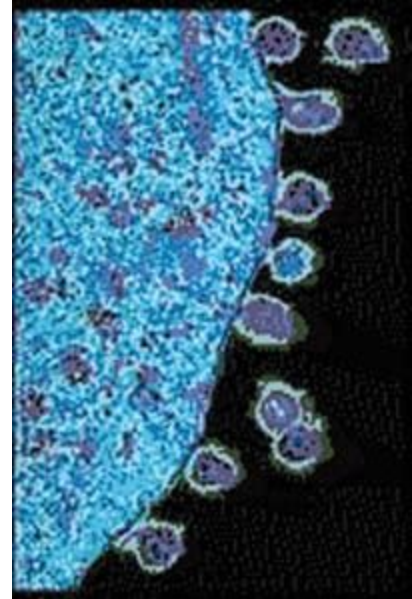


Figure 1. HIV. Several newly replicated HIV particles coming out of a host cell.

Source: [Wikimedia Commons](#).

What To Do

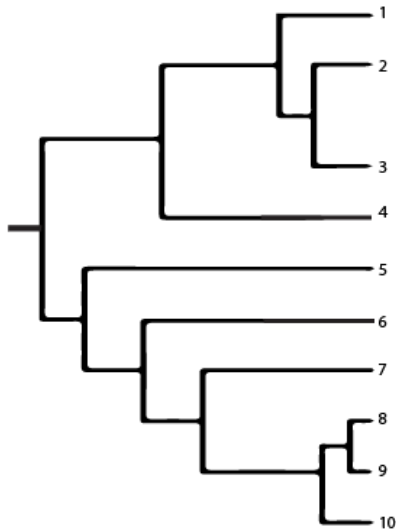
Answer the analysis questions below, reading the Bite when instructed.

Analysis Questions

1. A virus enters a host cell and makes the cell copy the virus' genetic material. This is an important step in how viruses reproduce.
 - a. Explain how viruses get mutations.

- b. Which virus is likely to accumulate (gather) more mutations: one that reproduces quickly, or one that reproduces slowly? Explain your reasoning.

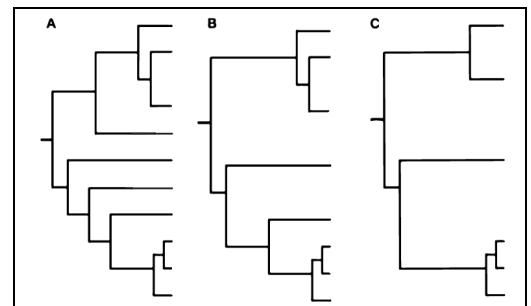
2. On the phylogenetic tree below, each number represents a group of closely related organisms.



- a. Put a square around the common ancestor of groups 1, 2, and 3.
- b. Put a circle around the common ancestor of groups 5, 7, and 8.
- c. Put a star around the common ancestor of all groups 1 through 10.

3. Look at the three phylogenetic trees below. Assume that each one is modeling the evolution of populations of a single virus over 50 years.

- a. Based on the trees, which population (A, B, or C) probably has the greatest amount of genetic diversity? Explain your reasoning.



- b. How would genetic diversity help a species to continue to survive and reproduce even if the environment changes?

- Even though some scientists do not consider viruses to be alive, there is no doubt that they do evolve. Based on the definition of evolution, explain what it means for a virus to evolve.

 & read **Science Bite**:
The Ancestry of an Outbreak: HIV from Past to Present

The HIV phylogenetic tree, **Figure 2** from the Bite, is shown again below.

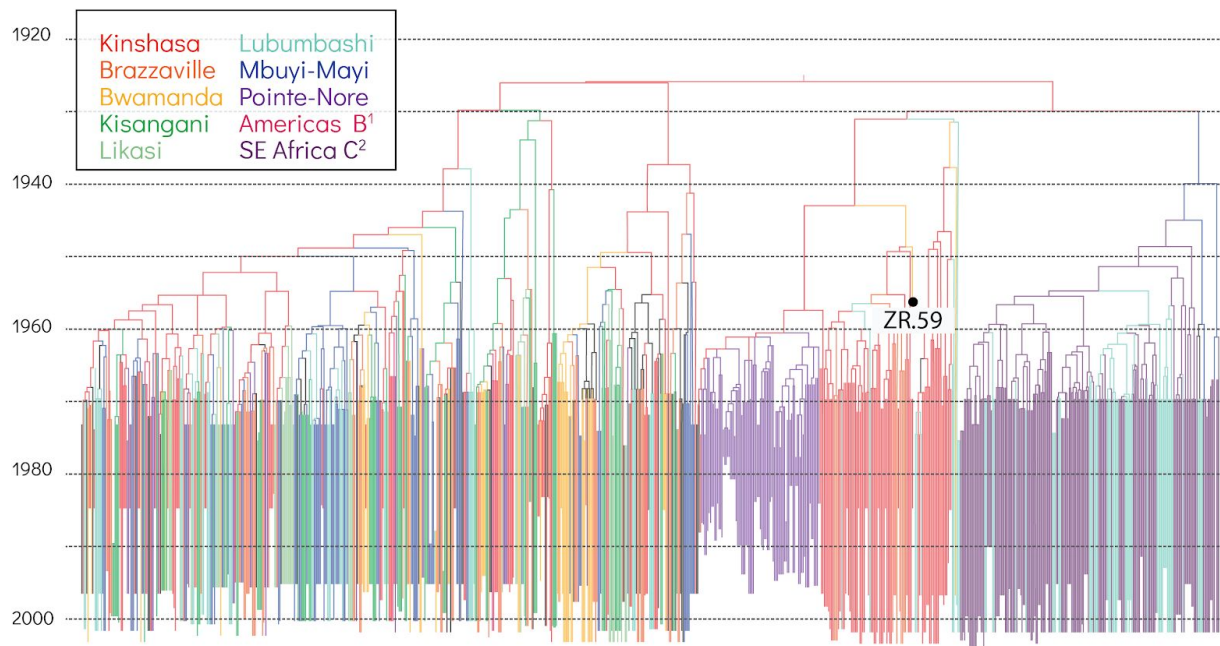
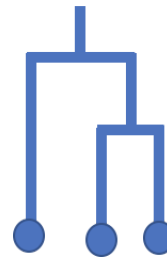


Figure 2. Phylogenetic tree of HIV diversity in Central Africa. This figure looks complicated, but you can get a lot of information from it. Each vertical line represents an individual patient sample, and the bottom point of the line ends at the time the sample was collected, with years on the Y-axis. Each line is also colored based on where the sample was collected. See the black circle labeled ZR59? That represents the earliest actual sample of HIV. Can you tell when and where the common HIV ancestor was found? You can! Look for the “trunk” of the tree and use the left axis to find the date, and the color to identify the region. *Source*: Modified from Faria *et. al.* 2014.

5. Use **Figure 2** to answer the following questions.
 - a. Circle the common ancestor of HIV in Central Africa.
 - b. According to **Figure 2** (and the Bite), where and when did the common ancestor of HIV first infect humans?
 - c. According to the Bite, the researchers compared only a small portion of the DNA sequence of each HIV strain, not the entire sequence. Why do you think they did that?

6. The researchers you read about compared almost a thousand genetic sequences to build their phylogenetic tree of HIV in Central Africa. In general, the more similar their genetic sequences, the more recently two viruses shared a common ancestor.
 - a. Given a seven-nucleotide section of a gene, compare the genetic sequences of these three specimens to each other and place them on the phylogenetic tree by writing 'A', 'B', or 'C' at the end of each branch. **Important!** In real studies, scientists would never rely on such a short sequence of DNA to draw conclusions. However, what you're doing here is the same thing that the researchers (or really, their computers did), only on a smaller scale.

Virus A	C	C	T	C	A	G	A
Virus B	C	C	T	G	A	C	T
Virus C	C	C	T	C	A	C	A



- b. Label the common ancestor of Viruses A, B, and C on the tree.
- c. Based on the information you have, what do you think the sequence of the viral common ancestor was? Fill in as many nucleotide bases as you feel confident doing. Leave others blank.

--	--	--	--	--	--	--

7. Researchers were able to use epidemiological data and genetic analysis to trace where and when HIV originated. Describe at least two ways researchers can build upon this knowledge to limit the global spread of HIV.

8. **Connect to the Big Idea** How could researchers use methods similar to what's in this study to find out more about humans' ancestry?

_v2: Swapped circle and square in Question 2; February 2020