



OVERVIEW

In this lesson, students will analyze Ebola sequences that were obtained from patients in Sierra Leone during the 2014 outbreak in West Africa. Students are challenged to place sequences into groups based on similarities to determine the transmission history of the virus. Students then compare their results to those of scientists at the Broad Institute of MIT and Harvard, who followed a similar procedure at the beginning of the outbreak.

KEY CONCEPTS AND LEARNING OBJECTIVES

- As viruses reproduce, they accumulate mutations in their genomes.
- Since mutations accumulate over time, analyzing virus sequences from infected individuals can help researchers track, understand, and treat diseases.

Students will be able to

- analyze and interpret sequence data.
- develop visuals to summarize and convey their findings.

CURRICULUM CONNECTIONS

Curriculum	Standards
NGSS (April 2013)	HS-LS-3, HS-LS-4, S.P.4, Cross-cutting Concept: Patterns
AP Biology (2012-13)	1.A.1, 3.A.1, 3.C.1, 3.C.3, Science Practices 1 and 5
IB Biology (2016)	5.2, B.4
AP Environmental Science (2013)	I.C, II.C, III.B.3

KEY TERMS

Ebola, mutation, outbreak, virus, sequencing

TIME REQUIREMENTS

This activity was designed to be completed in a 45-minute class period with additional time required for watching a video and reading background information. An optional extension activity will require additional time and may be assigned as homework.

SUGGESTED AUDIENCE

This lesson is appropriate for high school biology (all levels including AP and IB) and introductory college biology.

PRIOR KNOWLEDGE

- Students should understand the basic mechanism of viral replication: that a virus invades a host cell in order to hijack its cellular machinery and reproduce.
- Although this activity does not mention cladistics, students would benefit from previous experience grouping organisms according to shared-derived characteristics, as this procedure is similar to what they will do with the virus sequences.



TEACHING TIPS

- Before conducting this activity, have students watch the 8-minute video Think Like a Scientist: Natural Selection in an Outbreak (<https://www.youtube.com/watch?v=Tq2GhPZvdkU>) featuring computational geneticist Pardis Sabeti and epidemiologist Lina Moses and complete the background reading: Introduction to Ebola.
- In the video Think Like a Scientist: Natural Selection in an Outbreak, changes in the virus sequences are portrayed as changes in how the virus looks using drawings of the virus on squares of paper. This video provides an opportunity for discussing how creating effective artistic representations of scientific processes can be challenging. You may want to discuss with students whether this is an effective visual for illustrating how mutations occur and how some mutations spread by natural selection. Make sure that students understand that the mutations they are looking at in this activity are single-nucleotide changes in the virus genome. Some changes will have no effect on the virus, while others could affect the structure or function of particular proteins in the virus, but these are subtle changes, unlikely to affect the overall structure of the virus.
- Printing tip: Print the sequence sheet in color on card stock so, you can reuse them (recommend one set of sequences per 2-4 students). Provide the background reading electronically or print off a class set and reuse. This activity has two parts. In the first part, students group virus sequences. In the second, they compare their groupings to ones selected by scientists at the Broad Institute. Do not provide students Part 2 of this activity (pages 3-4) until they have completed Part 1 (pages 1-2).
- Make sure students understand that Ebola is an RNA virus. Scientists studying Ebola and other RNA viruses use reverse transcription to copy the RNA to DNA prior to sequencing, so the data that students analyze is DNA, but the actual genetic material inside the virus particles is RNA.
- If you have already covered evolution in your class, you may point out to your students the parallels between the sequence analysis they did in this activity to grouping organisms by shared-derived characteristics. In both cases, there is an underlying assumption of parsimony (all other things being equal, the best hypothesis is the one that requires the fewest evolutionary changes) and trait accumulation.

ANSWERS

Background Questions:

1. *Thinking about what you saw in the video and what Drs. Sabeti and Moses discussed, identify three factors that contributed to the number of individuals infected in the Ebola outbreak.*

Students' answers will vary, but they may mention that the virus has a selective advantage that makes it spread more rapidly, that the infection occurred in the healthcare staff, or that there was a high density of bats in the area, a poor healthcare infrastructure, high human population density, or good roads to allow for rapid movement between populations.

2. *Define the term "mutation."*

Answers will vary, but the reading states that "random changes to the sequence of letters occur." Students should indicate that these letters are nucleotides within a sequence of DNA or in the case of Ebola, RNA.

3. *In your own words, why is it important to examine the sequence of the Ebola virus genome during an outbreak?*

Students' answers will vary but should include some of the following points: knowing how the virus is mutating over time is important for understanding how it is spreading; it can diagnose individuals who are infected; and it can help determine whether the virus is becoming more infectious.



Analysis Questions, Part 1:

1. Describe the criteria you used to assign the sequences to different groups.

Students could have used one or more of the following attributes: identical sequences; sequences that only differ by one mutation; mutation location; number of mutations; or viruses that share sets of mutations.

2. Describe alternate criteria you could have used, and explain why you opted for the criteria you described in your answer to Question 1.

See answer to Question 1 for additional criteria. Explanations should clearly indicate why the student selected the criteria that they did.

3. If a sequence has a larger number of mutations when compared to the reference sequence, does that mean it is from earlier or later in the outbreak? Explain your answer.

It's from later in the outbreak. Mutations accumulate over time. The more mutations there are between a sample and the reference sample, the later in the outbreak it was collected.

4. Create a visual that highlights the relationship between your groups. Examples of effective visuals include flowcharts and trees. Be sure that your visual includes an arrow indicating passage of time during the outbreak.

An effective visual should include the reference sample and show how the different groups are related to each other over time.

Analysis Questions, Part 2:

1. Compare the groupings in Figure 1 to your groupings. What are the similarities and differences?

Answers will vary. Students should mention the number of groups, the criteria used to group sequences, or the positioning of different sequences.

2. Using the grouping in Figure 1, list the core mutations that occurred between one group and the other. Core mutations are mutations shared by every virus in the group. Describe the mutation by indicating the nucleotide number in the sequence.

- Differences between the reference sample and the Group 1 sequences: C in position 2 and T in position 5
- Differences between groups 1 and 2: T in position 1, C in position 7, A in position 12, and C in position 13
- Difference between groups 2 and 3: C in position 14

3. What can you infer from this diagram about when each group of patients contracted Ebola relative to one another?

Patients infected with Group 1 viruses became infected after the reference patient; patients with Group 2 viruses became infected after Group 1; patients with Group 3 viruses became infected after Group 2.

4. Explain how the sequences and groupings support the hypothesis that mutations accumulate over time.

Group 2 and 3 sequences have the same core mutations present in the Group 1 sequences. Group 3 sequences had the core mutations present in Group 2 sequences. This observation is consistent with the hypothesis that mutations accumulate over time. Mutations that arise in one virus will be passed on as the virus replicates and infects other people, and additional mutations then occur in these viruses as they replicate, and so on.

5. How can you explain the fact that some sequences have additional (noncore) mutations that did not spread into other groups?

Mutations that interfere with the essential functions of the virus are rapidly lost from the population and do not spread to other groups.



6. If a particular mutation was advantageous to the virus, in that it allows the virus to spread faster, what would you notice happening over time to the sequences of the virus you collected in a population?

Students should explain that over time almost all the virus sequences collected in samples from that population will have that mutation.

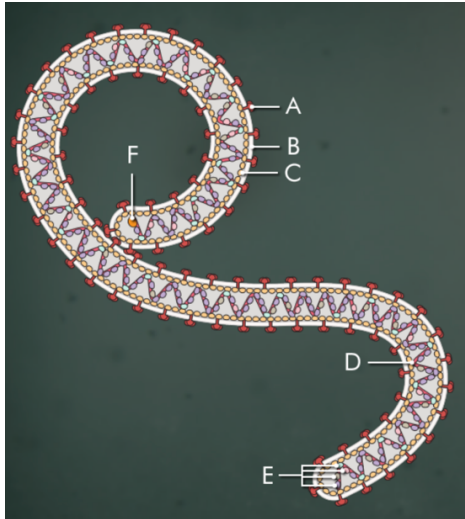
EXTENSION ACTIVITY

Depending on your students’ knowledge about viruses, it may also be useful for them to visit the Click and Learn interactive “Virus Explorer” (<https://www.hhmi.org/biointeractive/virus-explorer>). The Click and Learn explores different viruses, but you could have them focus on Ebola by completing the quick exploration activity below.

Categories of Exploration	Circle the correct choice(s) below	Notes
Host	Humans, Mammals, Birds, Reptiles, Plants, and Bacteria	Ebola can infect humans, other primates, and bats
Envelope	Enveloped or Naked	Some viruses exit their host cell by budding. In the process, part of the host cell membrane envelops the virus particle, forming an envelope.
Structure	Spherical, Helical, Icosohedral, or Conical	Typically described based on the overall shape of the protein layer that surrounds the virus genetic material.
Genome Type	ds DNA, ss + RNA, ss – RNA, Segmented, Linear, or Circular	Genomes vary by the type of nucleic acid, number of strands of nucleic acid, the sense or polarity of the strands, and the structure.
Transmission	Human-to-human, Zoonotic, Arthropod, Vector, Plant-to-plant, Bacterium-to-bacterium	The mechanism in which a virus passes from one host to another depends on several factors, including which organisms the virus is able to infect, which type of cells the virus infects, and how the virus is released from an organism. An organism that serves to transmit the virus from one host to another is termed a vector.
Vaccine	Vaccine Available or No Vaccine Available	A vaccine is a substance that, when taken into the body, should induce a protective immune response to a virus.



In the interactive, click on the cross section and write down the labels for Figure 2 below.



- A. Glycoprotein
- B. Lipid Envelope
- C. Matrix Protein
- D. RNA Genome
- E. Nucleocapsid Proteins
- F. Polymerase

Figure 2. Cross section of Ebola virus.

REFERENCES

1. Gire, Stephen K. *et al.* (2014). "Genomic surveillance elucidates Ebola virus origin and transmission during the 2014 outbreak." *Science* 345(6202):1369–1372.
2. Tam, Ruth. 2014. "This is how you get Ebola, as explained by science." PBS Newshour.
3. "Frequently Asked Questions on Ebola virus disease." World Health Organization. <http://www.who.int/csr/disease/ebola/faq-ebola/en/>

AUTHORS

This lesson was adapted from a teacher guide created by a collaborative group at the Broad Institute. The original activity is available at http://scienceintheclassroom.org/sites/default/files/disease_detectives_-_introduction_to_sequence_analysis.pdf

Edited by Melissa Csikari and Laura Bonetta, PhD, HHMI, and Stephanie Keep, consultant.

Reviewed by Nathan Yozwiak, PhD, Broad Institute.