Investigating a Mutation in HIV-1

Adapted from:

HIV Problem Space, written by Anton E. Weisstein and Sam Donovan, BEDROCK* (www.bioquest.org/bedrock)

Based on data from:

Markham RB, Wang WC, Weisstein AE, Wang Z, Munoz A, Templeton A, Margolick J, Vlahov D, Quinn T, Farzadegan H, Yu X-F. 1998. "Patterns of HIV-1 evolution in individuals with differing rates of CD4 T cell decline." *Proc. Natl. Acad. Sci. USA* 95: 12568-12573.

Written by:

April Bednarski** with advice from Alan Templeton

Funding:

*BEDROCK is part of the BioQUEST Curriculum Consortium (www.bioquest.org)

**This project was funded in part by a Professorship Award to Washington University in support of Sarah C.R. Elgin from Howard Hughes Medical Institute (HHMI)

INFORMATION FOR INSTRUCTORS:

Purpose

This inquiry-based lab is designed to help students advance their critical thinking and problem-solving skills. In this lab, students use sequence data collected from human subjects infected with HIV to analyze a hypothesis that links sequence data to patient health. Students must become familiar with the **evolution of HIV** and with making **sequence alignments** and **haplotype trees**. Students then use the data they have collected to analyze a hypothesis about an HIV polymorphism and immune system health.

Educational Context and Background Information

This lab was written for undergraduate introductory biology students. It accompanies a lecture course in population genetics that discusses HIV patient data collected from a specific study, called the ALIVE study (Markham, et al. 1998). The ALIVE study took place in Baltimore, Maryland on 15 patients who had become infected with the virus between 1989 and 1992. This experiment gives students the opportunity to work with some of the raw data from this study.

This lab uses the BEDROCK HIV Problem Space. BEDROCK is part of the Bioquest Curriculum Consortium. The BEDROCK project created several different "Problem Spaces" which are websites containing background information about an area of research and a collection of data sets. Instructors and students are then challenged to ask their own question (or make an hypothesis) about the data and analyze the data sets to come to a conclusion.

We teach this lab to approximately 400 students split into lab sections of 20 students each. Students use computers with a web connection to access background information and data from the BEDROCK HIV Problem Space (http://bioquest.org/bedrock/problem_spaces/hiv/). This lab does not require any special software on the computers beyond a web browser. Students could work individually or in pairs. We found the lab time to take approximately 2 hours, with students spending an additional 30 min -1 hour of time outside of class to finish the lab report write-up.

The lab uses DNA sequence alignments of a portion of the virus to explore changes in this sequence over time in the same individual. Students perform a **sequence alignment** using ClustalW, and analyze the alignment to identify base changes. These changes are then correlated to changes in the patient's health by reading the CD4 cell count. The number of CD4 cells per µl of blood serum is used a measure of immune system health, and a level of < 200 cells per µL is labelled as "severely compromised." Students may choose different individuals to study and compare their results with results from others in the class. Of the 15 patients in the data set, all but 2 support the hypothesis. Finally, students create a **haplotype tree** by hand as an exercise in the lab report write-up.

Supplemental Reading

For more information and a short movie on the HIV lifecycle:
http://www.hopkins-aids.edu/hiv_lifecycle/hivcycle_txt.html

For background information on the ALIVE study:

Markham RB, Wang WC, Weisstein AE, Wang Z, Munoz A, Templeton A, Margolick J, Vlahov D, Quinn T, Farzadegan H, Yu X-F. 1998. "Patterns of HIV-1 evolution in individuals with differing rates of CD4 T cell decline." *Proc. Natl. Acad. Sci. USA* 95: 12568-12573.

Templeton, AR, Reichert, RA, Weisstein AE, Yu XF, Markham RB. 2004. "Selection in context: patterns of natural selection in the glycoprotein 120 region of human immunodeficiency virus 1 within infected individuals." *Genetics* (167): 1547-61.

Answers to the Lab Report

Reading Questions:

1. Patients A and B are both HIV positive. Patient A has a CD4 count of 650 cells/μL and patient B has a CD4 count of 160 cells/μL. Do both patients have AIDS? Explain why CD4 counts are used as a diagnosis of AIDS.

   No, only patient B has AIDS because CD4 count is below 200 cells/μL. CD4 levels are used to diagnose AIDS because they are the cells that are attacked by HIV.

2. What is meant by the term "lentivirus"?

   A slow virus. This means that symptoms don't appear until long after infection.

3. What is proviral DNA?

   Viral DNA that has been incorporated into the host genome.

4. Explain the relationship of HIV-1 and HIV-2 to SIV's (simian immunodeficiency virus).

   HIV-1 originated from an SIV from chimpanzee while HIV-2 originated from an SIV from mangabeys.

5. Besides separate origins, describe three differences between HIV-1 and HIV-2.

HIV-1 is more virulent than HIV-2. The genomes are 15% - 30% different. HIV-1 has infected people worldwide while HIV-2 is mostly located in western Africa.

6. What are the two hypotheses for how HIV causes AIDS?

   One hypothesis is that the high rate of evolution of the virus leads to multiple mutant forms that escape and overwhelm the immune system. Another hypothesis is that a random mutation leads to an SI strain and this is the cause of the rapid immune system decline.

7. What is an SI strain of HIV and why is it more dangerous than a NSI strain?

   SI strains infect T-cells and cause them to fuse to healthy T-cells. This speeds the decline of the immune system as compared to NSI strains.

8. What does S6V2-14 refer to?

   Subject 6, Visit 2, HIV sequence 14

9. What three factors were used to select the 15 patients we will be studying from the ALIVE study?

   The 15 patients all seroconverted during the study. There were data available for at least 3 visits after the patients tested positive. The disease progression varied within the 15 patients.

10. What portion of the viral genome was sequenced in this study?

    The variable (V3) region of the envelope gene. The V3 region is part of g120 which is the extracellular portion of the envelope protein and is involved in cell binding.

Analysis Questions:

Here are the answers students should get for each of the subjects:

Subject 1- 8/13 = SI, CD4 = 15
Subject 2-9/9 NSI, CD4 = 830
Subject 3-6/6 SI, CD4 = 45
Subject 4-13/13 SI, CD4 = 135
Subject 5-5/5 NSI, CD4 = 700
Subject 6-9/9 NSI, CD4 = 560
Subject 7-9/9 NSI, CD4 = 310
Subject 8-10/10 SI, CD4 = 250, doesn't support hypothesis
Subject 9-8/8 NSI, CD4 = 270
Subject 10-2/10 SI, CD4 = 15, doesn't support hypothesis
Subject 11-9/9 SI, CD4 = 175
Subject 12-8/8 NSI, CD4 = 850
Subject 13-6/6 NSI, CD4 = 975
Subject 14-12/12 NSI, CD4 = 350
Subject 15-8/10 SI, CD4 = 15

11. Write the subject number you chose, the number of visits for that subject, and the total number of clones.
    Answers vary.
    Example: Subject 5, visits = 5, clones = 43

12. What trends in CD4 cell count do you notice for your subject?
    Answers vary depending on subject used in question 11. Use the "Summary and References" table from the website, called "Markham dataset summary.pdf"
    Example, for subject 1, CD4 count decreased over the visits.

13. Explain why there is more than one sequence for each visit.
    As the virus replicates, mutations can occur. Each new mutation is considered a new sequence or "clone". Even at the first visit, subjects already have multiple HIV clones.

14. Summarize your data here. Include number of sequences with SI mutation/ total sequences and CD4 T cell count.
    Will vary depending on Subject number.
    Data should look like:
    Subject 1 = SI 8/13 sequences and CD4 cell count at last visit = 15.

15. Does the data in your alignment support the theory that SI – associated mutations in more than half of the viral sequences leads to a decrease in CD4 T cell count?.

    Will vary depending on Subject number:

    Example, for subject 1, 8/13 is greater than 50% of the sequences contained the SI mutation and the subject had a low CD4 cell count relative to previous visits, so the data supports the stated hypothesis.

16. Directions: Draw a haplotype tree for the following sequences. These sequences are from hypothetical subject 16 from visits 1 and 2. This subject was infected with a single clone of HIV which had already evolved into 4 clones by the time of the second visit. Keep in mind that the haplotype tree should show clones from the second visit evolving from clones from the first visit. (Hint: all clones evolved from S16V1-1)

```
                    10          20          30          40
S16V1-1         GAGATAGTAA TTAGATCTGC CAATTTCTCG GACAATACTA AAA  43
S16V2-3         GAGATAGTAA TTAGATCTGC GAATTTCACG GACAATACTA AAA  43
S16V2-2         GAGGTAGTAA TTAGATCTGC CAATCTCACG GACAATGCTA AAA  43
S16V2-4         GAGGTAGTAA TTAGATCTGC CAATTTCACG GACAATACTA AAA  43
S16V2-1         GAGGTAGTAA TTAGATCTGC CAATCTCACG GACAATGCTA AGA  43
```

Procedure:

1. Since all the sequences listed evolved from the S16V1-1 sequence, use that sequence as your root. Circle changes from the S16V1-1 sequence in all the other sequences. (Don't circle anything in the gray-highlighted sequence.)
2. Start drawing the haplotype tree with the S16V1-1 sequence as the root. The next sequence(s) should be the one(s) that is the least changed from S16V1-1. On each connecting line, write each nucleotide change(s) and number(s) that connects it to the next sequence.
3. Continue drawing the tree until all the sequences are in the tree.

S16V1-1 ---(T28A)---**S16V2-3**----(A4G)---**S16V2-4**---(T25C and A37G)---**S16V2-2**----(A42G)----**S16V2-1**

Investigating a Mutation in HIV-1

In this lab, we will be analyzing the data from the ALIVE study. In the ALIVE study, the evolution of HIV-1 in individuals was studied by examining HIV-1 sequence data over a period of 3 – 4 years. The study shows multiple HIV-1 sequences for each patient, and change in these sequences over time. The mutations occur randomly, but there are regions of HIV-1 that mutate at a higher rate than other regions. This study focuses on a highly variable region of HIV-1, and we will try to determine if there is any correlation between a specific mutation in this region and rapid immune decline. The health of the immune system will be measured by the number of CD4 cell count. The website we will be using contains the original data from the ALIVE study as well as background information about HIV-1 and CD4 cells. The overall objective for this lab is to use the HIV sequence data and CD4 cell count data from a patient in the ALIVE study and determine if the data supports the hypothesis stated below. Class data will be pooled to determine the strength of the hypothesis.

Hypothesis: Presence of a SI mutation in more than half of the HIV sequences present in an HIV+ patient is associated with a severly compromised CD4 cell count (<200 cells/μL).

Link for learning about the HIV lifecycle:
http://www.hopkins-aids.edu/hiv_lifecycle/hivcycle_txt.html

Procedure:

1. Go to the BEDROCK HIV Problem Space at the website below and read the "Introduction" then the "Background" section. Answer the "Reading Questions" portion of your lab report (questions 1 – 10) while reading the "Background" section.

   http://www.bioquest.org/bedrock/problem_spaces/hiv/index.php

2. Go to the "Sequence Data" site and bookmark this site. Choose the "Summary and Reference" link. A pdf file should load on your desktop. If it does not automatically open, double-click on this file and look over the table that comes up. You should see information about each patient in the study and their CD4 cell counts from each visit. Since CD4 cells are a type of immune cell that is attacked by HIV, the CD4 cell count is an indication of the advancement of AIDS. Choose a patient to study. Answer questions 11 and 12.

3. Return to the "Sequence Data" site. Choose the "amino acid sequences" link. This link will take you to a page where you can download all the sequences that were used in the ALIVE study. Choose the link that corresponds with the patient you chose ("Subject _, all visits").

4. Scroll through the page of sequence information. Be sure you understand how the sequence data is labelled before continuing. You should see several sequences for each visit. Answer question 13.

5. Scroll to the bottom of the page. Use the mouse to highlight all the sequences for the LAST visit. Then copy these sequences (under "Edit," choose "Copy").

6. Go to the ClustalW website at:
   http://www.ebi.ac.uk/clustalw/index.html

7. Paste the sequences into the data-entry box. (Click in the box, then choose "Paste" under "Edit.")

8. You don't need to change any of the settings in this program. Click "Run" at the bottom of the page and wait for your results. This web-based program will analyze the sequences that you entered and then align them so that as many similar amino acids are positioned above eachother as possible.

9. Scroll down the output page until you see all the sequences lined up. At the bottom of each position, there is a "*" if the the amino acid is the same in all the sequences at that position. Print your alignment.

10. The SI phenotype is associated with a lysine (K) or arginine (R) at position 306. Use the guide below to help you find position 306 on your alignment.

11. Now look at the CD4 T cell count on the table you downloaded in Step 3. Write the CD4 count on the alignment printout. Answer questions 14 – 16.

Guide for Finding "Position 306"

Consensus sequence for the V3 region of the coat protein of HIV-1:

```
                                      -306            -320
EVVIRSVNFT DNAKTIIVQL NTSVEINCTR PNNNTRKRIR IQRGPGRAFV TIGKIGNMRQ 60

AHCNISRAKW NNTLKQIASK LREQFGNNKT IIFKQSS                          97
```

Use the sequence boxed in gray as a guide for finding position "306" in your alignment. The position may not be numbered "306" in your alignment because random mutations can lead to insertions or deletions in the amino acid sequences and change the numbering. Also, the sequence may be slightly changed from what is shown. Instead, find the sequence that most closely matches the gray-boxed sequence in your alignment and analyze the next amino acid as position "306".

SI phenotype is associated with either a lysine (K) or arginine (R) at position 306 OR a glycine (G) at position 306 and a lysine (K) or arginine (R) at position 320. The "-" above is positioned directly above the residue numbered 306 or 320.

Name:_____Section:_____Date:_____

Lab Report: Investigating a Mutation in HIV-1

Reading Questions:

8. Patients A and B are both HIV positive. Patient A has a CD4 count of 650 cells/μL and patient B has a CD4 count of 160 cells/μL. Do both patients have AIDS? Explain why CD4 counts are used as a diagnosis of AIDS.

9. What is meant by the term "lentivirus"?

10. What is proviral DNA?

11. Explain the relationship of HIV-1 and HIV-2 to SIV's (simian immunodeficiency virus).

12. Besides separate origins, describe three differences between HIV-1 and HIV-2.

13. What are the two hypotheses for how HIV causes AIDS?

14. What is an SI strain of HIV and why is it more dangerous than a NSI strain?

17. What does S6V2-14 refer to?

18. What three factors were used to select the 15 patients we will be studying from the ALIVE study?

19. What portion of the viral genome was sequenced in this study?

Analysis Questions:

20. Write the subject number you chose, the number of visits for that subject, and the total number of clones.

21. What trends in CD4 cell count do you notice for your subject?

22. Explain why there is more than one sequence for each visit.

23. Summarize your data here. Include number of sequences with SI mutation/ total sequences and CD4 T cell count.

24. Does the data in your alignment support the theory that SI – associated mutations in more than half of the viral sequences leads to a decrease in CD4 T cell count? Please explain.

25. Directions: Draw a haplotype tree for the following sequences. These sequences are from hypothetical subject 16 from visits 1 and 2. This subject was infected with a single clone of HIV which had already evolved into 4 clones by the time of the second visit. Keep in mind that the haplotype tree should show clones from the second visit evolving from clones from the first visit. (Hint: all clones evolved from S16V1-1)

```
                    10          20          30          40
S16V1-1         GAGATAGTAA TTAGATCTGC CAATTTCTCG GACAATACTA AAA  43
S16V2-3         GAGATAGTAA TTAGATCTGC GAATTTCACG GACAATACTA AAA  43
S16V2-2         GAGGTAGTAA TTAGATCTGC CAATCTCACG GACAATGCTA AAA  43
S16V2-4         GAGGTAGTAA TTAGATCTGC CAATTTCACG GACAATACTA AAA  43
S16V2-1         GAGGTAGTAA TTAGATCTGC CAATCTCACG GACAATGCTA AGA  43
```

Procedure:

1. Since all the sequences listed evolved from the S16V1-1 sequence, use that sequence as your root. Circle changes from the S16V1-1 sequence in all the other sequences. (Don't circle anything in the gray-highlighted sequence.)
2. Start drawing the haplotype tree with the S16V1-1 sequence as the root. The next sequence(s) should be the one(s) that is the least changed from S16V1-1. On each connecting line, write each nucleotide change(s) and number(s) that connects it to the next sequence.
3. Continue drawing the tree until all the sequences are in the tree.