Names _____Period _____Date _____

Drug-Resistant TB: A Genetic Analysis Using Online Bioinformatics Tools

Student Activity

Overview

This lesson asks you to compare gene sequences between one wild-type and one of a variety of mutant *Mycobacterium tuberculosis* (TB) strains. You will identify mutations as single-nucleotide polymorphisms (SNPs) and then make an inference on whether your variant strain will be resistant to a TB drug or not.

Objectives

By the end of this activity you will be able to:

1) identify and explain what a single-nucleotide polymorphism (SNP) is when comparing two gene sequences.
2) navigate online scientific tools to translate DNA into polypeptide sequences and to compare and contrast wild-type and variant polypeptide sequences.
3) determine whether your given SNP will result in 'sense,' 'missense,' or 'nonsense' in the resulting amino acid sequence.
4) hypothesize whether a SNP will likely cause antibiotic resistance.

Background

Tuberculosis (TB) is a deadly infectious disease, primarily affecting the lungs. It is caused by the pathogenic bacterium *Mycobacterium tuberculosis* (MTB). Transmission occurs when a person with active TB disease coughs and aerosolizes the bacteria, which then spread to other individuals. Symptoms of active TB include chest pain, prolonged cough, and blood in the sputum. Worldwide, approximately 1.4 million people die from TB each year, with most of the new cases and deaths occurring in developing countries.

Because TB is caused by a bacterium, it can often be treated using antibiotics, although the course of treatment typically extends 6 months or longer. Alarmingly, there is an increasing prevalence of TB strains that are resistant to the antibiotics typically prescribed, including multi-drug resistant (MDR) and extensively drug resistant (XDR) strains. MDR-TB infections are resistant to two of the first line TB antibiotics, isoniazid and rifampin, and XDR-TB infections are resistant to four types of antibiotics.

M. tuberculosis colonies are very slow-growing, so the recommended treatment of a new case of TB is 6 months of a combination of 2-4 different types of antibiotics. For MDR-TB, treatment with at least 4 effective antibiotics should last from 18 to 24 months. It is important that antibiotics with a strong likelihood of success be prescribed as quickly as possible and *taken as directed for the entire prescribed length of time.* Failure to do so may allow certain MTB with random mutations to survive, reproduce, and create a second infection that is resistant to the antibiotic(s) originally used to treat the first infection.

A wild-type strain of *M. tuberculosis* does not have any gene mutations that confer antibiotic resistance, making it susceptible to all standard classes of TB antibiotics. Other strains of MTB, including MDR and XDR TB strains, have mutations in the form of **single-nucleotide polymorphisms (SNPs)** which lead to resistance to antibiotics. Once a bacterium is resistant to a particular antibiotic, that antibiotic is no longer effective in killing the bacteria or curing the infection.

A SNP (pronounced "snip") is a single base pair substitution that can be observed when comparing similar DNA sequences of the same gene—either between organisms, strains, or homologous chromosomes (Figure 1). A SNP can lead to a **synonymous polymorphism** or a **nonsynonymous polymorphism.** Synonymous polymorphisms (also known as 'silent mutations') do not lead to an amino acid change in the translated protein sequence, since multiple codons can code for the same amino acid (Figure 2). Nonsynonymous polymorphisms, however, lead to a change in the protein sequence and can either cause a **missense** mutation, which results in a different amino acid in the protein sequence or a **nonsense** mutation that results in an early stop codon (Figure 1b).

Fig. 1 An example of a single-nucleotide polymorphism

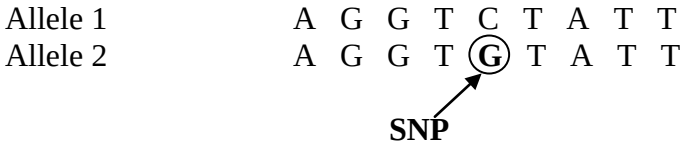


Fig. 2 Possible consequences of SNPs

Wild-type allele	A G G T C T A T T		
mRNA transcript	U C C A G A U A A		
Original amino acid sequence	Serine—Arginine—Stop	=	**ORIGINAL**
Synonymous polymorphism	A G G T C **C** A T T		
mRNA transcript	U C C A G G U A A		
Alternative amino acid sequence	Serine—Arginine—Stop	=	**SENSE**
Nonsynonymous polymorphism	A G G T **G** T A T T		
mRNA transcript	U C C A C A U A A		
Alternative amino acid sequence	Serine—Threonine—Stop	=	**MISSENSE**
Nonsynonymous polymorphism	A G G **A** C T A T T		
mRNA transcript	U C C U G A U A A		
Alternative amino acid sequence	Serine—Stop	=	**NONSENSE**

Recall that the order of amino acids in a polypeptide determines the polypeptide's three-dimensional shape and, consequently, the protein's structure and function.

Antibiotics often target proteins at particular **binding sites** that disrupt the function of the protein. Because the protein no longer works, the bacterium can't carry out its normal functions and it will perish. For example, **rifampin**, an antibiotic against TB, binds to and inhibits a subunit of MTB's RNA polymerase. If the cell's RNA polymerase doesn't work, what are the consequences? Genes can no longer be transcribed into proteins, and no proteins means no functional cellular machinery!

Bacteria that are resistant to rifampin can have a mutation in their rpoB gene which alters the site of where rifampin binds to RNA polymerase. Therefore, rifampin isn't able to bind to the polymerase, the polymerase continues to work, and the bacteria live!

In this activity, you will be given the DNA sequence of the rpoB (beta subunit of RNA polymerase) gene from a wild-type *Mycobacterium tuberculosis* and the rpoB gene sequence of a variant strain of MTB. Your job is to identify any SNPs in the variant gene sequence, determine the amino acid sequences of both the wild-type and variant alleles and whether the SNP is a synonymous or nonsynonymous polymorphism. You will then need to critically evaluate what effect the SNP may have on conferring antibiotic resistance of the variant strain of *M. tuberculosis*.

Materials

- Computer with internet access
- Wild-type *MTB* rpoB gene sequence (wild-type allele)
- Variant *MTB* rpoB gene sequence (variant allele)
- Copymaster 1: Genetic Code chart
- Copymaster 2: Properties of Amino Acids chart

PROCEDURE AND QUESTIONS

I have been assigned MTB variant (letter) ____.

➔ **Go to www.stronglab.org/taylor to begin the lesson at Part A.**

Part A: Identifying SNPs

1) Per your teacher's directions, open the digital document that has been assigned to you.

2) Take one minute to visually scan the two sequences to see if you can find a single-nucleotide polymorphism (SNP) in the variant allele.

3) Did you find it? _____ Describe the experience of comparing these two allele sequences. Is it easy? Difficult? Explain. What could make this process easier?

_____
_____
_____
_____
_____
_____

4) Comparing gene sequences by hand is a time-consuming process. Fortunately, computer programs have been created to make this task happen almost instantaneously.
   a. Open your computer's internet browser and go to ClustalW at http://www.genome.jp/tools/clustalw/.
   b. On ClustalW, next to 'Enter your sequences…' click on DNA.
   c. Copy and paste both of your gene sequences into the large empty box. Be sure to include the sequence labels (e.g. >wild-type_TB) for each sequence. The > symbol should start a new line.
   d. Click 'Execute Multiple Alignment.'
   e. On the page that comes up, scroll down to the section headed with 'clustalw.aln.'
   f. Look at the alignment of your two sequences. Stars (***) indicate bases that are identical. **An empty space indicates a SNP**.

5) What is the base change? The base in the wild-type TB allele is a(n) _____ while the base in my variant allele is a(n) _____.

6) Based on what you have done in class so far, what would you have to do in order to determine if the amino acid sequence changes due to the SNP?

_____
_____

7) Look back at the top of your ClustalW results. Locate how many base pairs (bp) this gene is: _____ bp. Explain how you could calculate how many *amino acids* this gene codes for?

_____

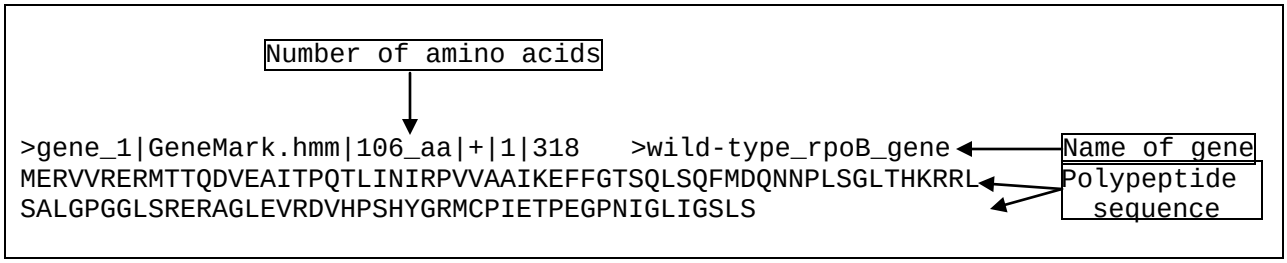
_____

Perform that calculation. The rpoB gene codes for a protein that is _____ amino acids long.

Part B: Translating DNA into Amino Acids

You will not need to do transcription and translation of your DNA sequences by hand!! There are websites that biologists use that will translate a DNA sequence into an amino acid sequence.

8) **You will now *translate* each gene sequence into a polypeptide (amino acid sequence).**
   a. Go to *GeneMarkS* at http://exon.gatech.edu/genemarks.cgi.
   b. Go back to the original data file of your wild-type and variant gene sequences.
   c. Copy and paste the wild-type TB gene sequence into the large box. Be sure to include the sequence label (e.g. >wild-type_TB) for the sequence.
   d. Under 'Output options,' mark the boxes on 'Protein sequence' and 'Gene nucleotide sequence.'
   e. Click 'Start GeneMarkS.'
   f. On the next page, click on the "gms.out.faa" link.
   g. You will get a result that looks like this. Important information is identified for you.



The amino acid sequence is given with single-letter codes for the 20 different amino acids. For example, M = Met = Methionine; K = Lys = Lysine.

9) Next, you need to copy and paste your amino acids sequence into a new Word document.
   a. Open a new Word document.
   b. Back in GeneMarkS, copy the wild-type amino acid sequence. (Ctrl-C)
   c. Paste the sequence into the new Word document. (Ctrl-V)
   d. Give it a label like >WT_rpoB_polypeptide.

10) Repeat Steps 8 a-g to get the amino acid sequence for the VARIANT allele.
    a. Go back to the homepage of GeneMarkS.
    b. Delete the WT sequence (Ctrl-A selects all, then hit 'delete').
    c. Paste in your variant gene sequence. Verify all webpage settings as in Step 8. Click 'Start GeneMarkS.'
    d. From the output page, copy the variant amino acid sequence and then paste it into the Word document you created for the wild-type polypeptide sequence.
    e. Give it a label like >X_rpoB_polypeptide. (Substitute "X" with your variant's letter.)

Part C: Identifying Synonymous and Nonsynonymous Mutations

In order to determine if the SNP in your variant sequence will affect the structure and function of the protein, you will need to align the two amino acid sequences (like you did with the gene sequences) and determine if the SNP causes a synonymous or nonsynonymous mutation in the variant protein. Here's how to do that.

11) Return to (or reopen) ClustalW at http://www.genome.jp/tools/clustalw/.
    a. On ClustalW, next to 'Enter your sequences…' click on **Protein**.
    b. Copy and paste the **wild-type TB and variant TB polypeptide sequences** from Word into ClustalW. Be sure each sequence is labeled with a > and a name.
    c. Click 'Execute Multiple Alignment.'
    d. On the next page that comes up, scroll down to the section headed with 'clustalw.aln.'
    e. Look at the alignment of your two sequences. Stars (***) indicate amino acids that are identical. **A semicolon (:), a period (.), or a blank space indicates a changed amino acid. A series of hyphens (-----) indicates missing amino acids.**

12) Are your amino acid sequences identical or are they different? _____
Therefore, does the SNP in your variant rpoB gene sequence cause a synonymous mutation or a nonsynonymous mutation? _____

- If they are identical, skip to Question 15. If they are different, continue with Question 13.

13) For variants C-F and H, fill in the chart below to compare the different amino acid between the two sequences. For variant G, skip to Question 14.

a. Use Copymaster 1: Amino Acid Chart and Copymaster 2: Properties of Amino Acids.

	Wild-type TB polypeptide	Variant ___ polypeptide
Numerical position of the mutation	(Count the letters until you get to the mutation.)* Number _____	
Amino acid single letter symbol		
Amino acid 3-letter symbol		
Amino acid full name		
Chemical property of the amino acid**		

*An easy way to do this in Word is to highlight the amino acids from the beginning of the sequence, stopping when you've just included the different amino acid. Then, click on 'Review' at the top of Word, then click on 'Word Count.' The Word Count will give you the numerical position of the mutation.

Click on the Properties of Amino Acids chart. Determine what chemical property each amino acid has: **acidic, basic, polar, or nonpolar.

14) For Variant G, provide and explanation for why there are missing amino acids at the end of the variant protein.

_____
_____
_____
_____
_____
_____

15) Does the amino acid change cause **sense, missense** or **nonsense**? _____ (Refer to the Background section) Why did you classify the change that way?

_____
_____
_____
_____
_____
_____

Part D: Hypothesizing Antibiotic Resistance

As mentioned in the background section earlier, a SNP in the gene region that changes the structure of the binding site in the polymerase for rifampin could cause rifampin to no longer work. If that happens, then the bacterium with that mutation is considered to be resistant to rifampin. If there is no change to the rifampin binding site in RNA polymerase, then the bacterium is considered to be susceptible to rifampin. **Your job now is to hypothesize whether or not the bacteria from which your variant TB allele came is resistant or susceptible to rifampin.**

But first you need some more information to help you with this. Keep reading!

- The binding site for rifampin is located between amino acids 36 and 67.
- Amino acids with the same properties will likely cause the protein to fold in its original way.
- Amino acids with different or opposite properties will likely cause the protein's shape to change.

Using the information in the Background, your completed chart in #13, and the information above, make a claim as to whether the bacteria containing your given variant protein will be resistant or susceptible to rifampin. (This portion of the work will be a DRAFT of your final scientific explanation.)

Claim: I predict that *Mycobacterium tuberculosis* with the rpoB variant ___ allele will be _____ to rifampin.
resistant OR susceptible

Evidence: The evidence that supports my claim is (provide evidence that supports your claim)

_____
_____
_____
_____
_____
_____
_____
_____
_____
_____
_____
_____

Reasoning: The reason for my claim is that

_____
_____
_____
_____
_____
_____
_____
_____
_____
_____
_____
_____

Going Even Further (If you finish early or if you want to try it again.)

Choose a different TB variant allele (C-H) to analyze by repeating the entire process. Pick up a "Second Variant TB Allele" worksheet from your teacher. Use this packet as your guide. Write down your new answers on your worksheet.

Part E: Comparing All Variant rpoB MTB Genes in the Class

If time is available, fill in the table below with data and observations from your classmates who investigated other variant MTB strains. Your teacher may display this chart on the board.

Mtb variant	Is the mutation in rifampin's binding site (aa 36 to 67)?	What is the amino acid change between the wild-type and variant *Mtb* polypeptides? (From ____ to ___ OR *No change*)	Does the chemical property of the variant amino acid change? If so, what is the change?	Does the polypeptide mutation result in sense, missense, or nonsense in the amino acid sequence?	Will this variant likely be susceptible or resistant to rifampin?
A					
B					
C					
D					
E					
F					
G					
H					