Name _____Period _____ Date _____

Drug-Resistant TB: A Gene Analysis

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' or 'missense' in the resulting amino acid sequence.
4) hypothesize whether a SNP will likely cause antibiotic resistance.

Background

Tuberculosis (TB) is a serious bacterial infection, primarily of the lungs, that is most commonly caused by *Mycobacterium tuberculosis* (MTB). Transmission of TB occurs when a person with an active case of TB coughs, sneezes or otherwise aerosolizes the bacteria and that bacteria is inhaled by another person.

Worldwide, 8.8 million new cases of TB were diagnosed and 1.5 million people died from TB in 2010, with most of the new cases and deaths occurring in developing countries. It is estimated that one-third of the world's population is infected with *M. tuberculosis*. However, 90% of those infected will not develop an active case of TB. For those who develop active TB, symptoms include chest pain, prolonged cough, blood in the sputum, fever, night sweats, and weight loss. Without effective treatment, 66% of people with active TB will die.

Because it's caused by a bacterium, TB can be treated using antibiotics, and there are currently eight classes of drugs available to treat TB. Two antibiotics that are commonly used to treat TB are **isoniazid** and **rifampin.** However, there is a growing prevalence of multi-drug resistant (MDR) TB infections with ~650,000 cases in 2010. MDR-TB infections are resistant to both isoniazid and rifampin. Extensively- and totally- drug resistant (XDR, TDR) TB have also been reported.

Because *M. tuberculosis* colonies are very slow-growing, 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.

Recall that bacteria have a single, circular chromosome. The genome of *M. tuberculosis* is approximately 4 million base pairs with 4,000 genes. A wild-type strain of *M. tuberculosis* does not have any genes that confer antibiotic resistance, making it susceptible to all 8 classes of TB antibiotics. Other strains of *M. tuberculosis*, however, could have mutations such as **single-nucleotide polymorphisms (SNPs,** pronounced "snips") which could create resistance to certain antibiotics.

A SNP is a single base pair substitution that can be observed when comparing sequences from the same genes between two organisms or between homologous chromosomes (See Fig. 1.) A SNP which codes for the same amino acid sequence as its homologue is called a **synonymous polymorphism**, or a 'silent mutation.' A SNP that codes for a different amino acid sequence is called a **nonsynonymous polymorphism**. A nonsynonymous polymorphism can either cause **missense**, which results in a different amino acid, or **nonsense**, which results in a stop codon that comes too early in the gene sequence. (See Fig. 2.)

Fig. 1 An example of a single-nucleotide polymorphism

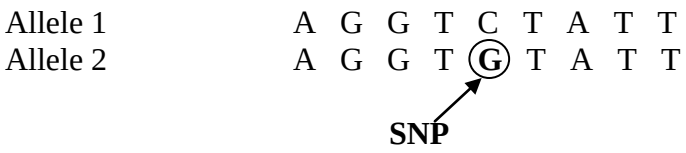

Fig. 2 Possible consequences of SNPs

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 or SILENT or SYNONYMOUS MUTATION**
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) ____.

Part A: Identifying SNPs

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

2) Take two minutes 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) Which row is your SNP in? Circle one: 1^{st} 2^{nd} 3^{rd} 4^{th} 5^{th}

6) What is the base change? Wild-type TB base is a(n) _____; Variant ___TB base is a(n) _____.

7) Do you think this base change will change the amino acid sequence that the variant allele codes for? _____ Explain your reasoning.

_____
_____
_____
_____

8) 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?

_____
_____
_____
_____
_____
_____
_____
_____

9) Look back at your Clustalw results. Locate how many base pairs (bp) this gene is. _____ bp. Calculate how many amino acids this gene codes for. _____ Show your work below:

Part B: Translating DNA into Amino Acids

10) There are websites that scientists use that will translate a DNA sequence into an amino acid sequence! **You will now *translate* each gene sequence into a polypeptide (amino acid sequence).**
    a. Go to *Softberry* at http://linux1.softberry.com/berry.phtml
    b. In the left column, hover over 'Operon and Gene Finding in Bacteria.'
    c. Click on 'FGENESB Gene Finding in Bacteria.'
    d. Go back to the original data file from your teacher.
    e. 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.
    f. In the field 'Choose closest organism,' use the drop-down menu to select 'BACTERIAL generic.' [Hint: It's 5th from the bottom.]
    g. Click 'Process.'
    h. You will get a result that looks like this. Important information is identified for you.

```
Prediction of potential genes in microbial  genomes
 Time:   Tue Jan  1 00:00:00 2005
 Seq name: wild-type_TB ← Your sequence's name
 Length of sequence - 282 bp  ← Length of the sequence in number of base pairs
 Number of predicted genes - 1
 Number of transcription units - 1, operons - 0
     N      Tu/Op   Conserved  S             Start         End    Score
                   pairs(N/Pv)
     1     1 Tu  1     .       +    CDS          1 -       280    145
Predicted protein(s):
>GENE     1         1  -       280    145     93 aa, chain +  ← Number of amino acids in the polypeptide
                                                                (protein) sequence
MPKSKVRKKNDFTVSAVSRTPMKVKVGPSSVWFVSLFIGLMLIGLIWLMVFQLAAIGSQA  ← Amino acid  sequence
PTALNWMAQLGPWNYAIAFAFMITGLLLTMRWH
```

The amino acid sequence is given with single-letter codes for the 20 different amino acids. See Copymaster 1: Genetic Code chart to determine which letter stands for which amino acid. For example, M = Met = Methionine; K = Lys = Lysine.

i. Open a new Word document.
j. Copy and paste the wild-type amino acid sequence into a new Word document. Give it a label like >WT_rpoB_protein.
k. Repeat for the variant protein.
l. Go back to *Softberry*, click your 'back arrow,' and delete the WT sequence (Ctrl-A selects all, then hit 'delete').
m. Paste in your variant gene sequence. Give it a label like >X_rpoB_protein. (Substitute "X" with your variant's letter.)

Part C: Identifying Synonymous and Nonsynonymous Polymorphisms

11) 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 polymorphism in the variant protein. Here's how to do that.
    a. Return to (or reopen) ClustalW at http://www.genome.jp/tools/clustalw/.
    b. On *ClustalW*, next to 'Enter your sequences…' click on **Protein**.
    c. Copy and paste the wild-type TB and variant TB amino acid sequences from Word into *Clustalw*. Be sure each sequence is labeled with a > and a name.
    d. Click 'Execute Multiple Alignment.'
    e. On the next page that comes up, scroll down to the section headed with 'clustalw.aln.'
    f. 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? _____
    - If they are identical, skip to Question 14. If they are different, continue with Question 13.

13) Fill in the chart below to compare the different amino acid between the two sequences.

	Wild-type TB protein	Variant ___ protein
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*		

*See Copymaster 2: Properties of Amino Acids chart. Determine if the amino acid is grouped as 1) **positively-charged, 2) negatively-charge, hydrophilic, (polar but with uncharged side chains)** or **hydrophobic**.

14) Does the amino acid change cause **missense** or **nonsense**? _____ (Refer to the Background section, page. 2.) Why did you classify the change that way?

_____
_____
_____
_____
_____
_____

15) Does your variant TB allele produce a synonymous or nonsynonymous polymorphism in the variant protein? _____

Part D: Hypothesizing Antibiotic Resistance

As mentioned in the Background section earlier, a SNP in the gene region that creates a binding site in the protein for rifampin could change the protein's (RNA polymerase) structure in such a way that rifampin can't bind. If that happens, then the bacterium with that mutation would be considered resistant to rifampin. **Your job now is to hypothesize whether or not the bacteria from which your variant TB allele came is resistant or not resistant 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 65 and 80.
- 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 for #12, and the information above, predict whether the bacteria containing your given variant protein will be resistant to rifampin.

Prediction: I predict that *Mycobacterium tuberculosis* with the rpoB variant ___ allele will be _____ to rifampin.
resistant OR not resistant

Rationale: I think this because (provide evidence that supports your prediction) _____

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

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

Fill in the table as results are displayed on the board in class.

MTB Variant	Sense, missense or nonsense in the amino acid sequence	Amino acid change between wild-type and variant (From ___ to ___ OR *no change*)	Synonymous or Nonsynonymous Polymorphism	Is the amino acid change in the binding site for rifampin? (Between 65 & 80 aa)	Inference: Will this variant be resistant to rifampin?
A					
B					
C					
D					
E					
F					
G					
H					