![BioOne RESEARCH EVOLVED]

Drug-Resistant Tuberculosis: A Genetic Analysis using Online Bioinformatics Tools

Author(s): Jessica M. Taylor, Rebecca M. Davidson and Michael Strong

Drug-resistant Tuberculosis: A Genetic Analysis Using Online Bioinformatics Tools

● JESSICA M. TAYLOR,
REBECCA M. DAVIDSON,
MICHAEL STRONG

ABSTRACT

Tuberculosis (TB) continues to be a serious global health problem, resulting in >1.4 million deaths each year. Of increasing concern is the evolution of antibiotic-resistant strains of the bacterium that causes TB. Using this real-world scenario, we created a 90-minute activity for high school or undergraduate students to use online bioinformatics tools to detect single-nucleotide polymorphisms (SNPs) between a wild-type and a variant Mycobacterium tuberculosis *gene that could confer resistance to a commonly used TB antibiotic, rifampin. Students write a scientific explanation, providing evidence and reasoning, to support their claim of antibiotic resistance or susceptibility. The entire lesson can be found online at http://www.stronglab.org/taylor.*

Key Words: *Bioinformatics; single-nucleotide polymorphisms; mutations; antibiotic resistance; tuberculosis; scientific explanations.*

In today's scope of biology education, it is necessary to create a foundational understanding of DNA, genes, and how mutations create changes in the DNA sequence. The Next Generation Science Standard HS-LS1-1 expects students to be able to "construct an explanation based on evidence for how the structure of DNA determines the structure of proteins which carry out the essential functions of life" (NGSS Lead States, 2013). A common example of DNA mutations used in many high school and college classes is the example of the gene coding for β-hemoglobin (Herron et al., 2010). A single base-pair mutation in the β-hemoglobin gene can result in a change in the protein sequence, leading to sickle cell disease. If a teacher has already offered the sickle cell example in class, a complementary example is needed to provide an opportunity to challenge students to construct an explanation in demonstration of Standard HS-LS-1.

With the burgeoning fields of genomics and bioinformatics, there has been a call for early introduction into these fields, including increased prominence in high schools (Bloom, 2001;

> *Tuberculosis (TB) is a deadly infectious disease that primarily affects the lungs.*

Gallagher et al., 2011; Lewitter & Bourne, 2011; Machluf & Yarden, 2013). The activity we present here provides a mechanism of gathering evidence in a manner similar to what would be done by biologists in the modern laboratory. Students are introduced to important methods in genomics and bioinformatics through the active use of online bioinformatics research tools, applied to a scenario of global health importance: drug-resistant tuberculosis.

Tuberculosis (TB) is a deadly infectious disease that primarily affects 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 can then spread to other individuals. Symptoms of active TB include chest pain, prolonged cough, and blood in the sputum. Worldwide, ~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 multidrug 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. Pertinent to this lesson, the antibiotic rifampin binds to the beta-subunit of *Mtb*'s RNA polymerase, inhibiting the enzyme's normal function and resulting in bacterial death.

○ The Lesson: An Overview

During the classroom exercise, laptops or netbooks are positioned in the class to be shared between pairs of students. The learning objectives for this lesson are listed in Appendix 1, and specific student instructions and questions are listed in Appendix 2.

Part A: Identifying Single-nucleotide Polymorphisms

Students are asked to compare the *rpoB* gene sequence from both a wild-type and a hypothetical variant *Mtb* strain. Each pair of students is assigned a digital document containing a portion (nucleotides 1168–1429) of the 3519 nucleotide *rpoB* gene sequence from a wild-type *Mtb* strain, along with one of eight possible variant *Mtb* strains (Figure 1). The wild-type *rpoB* gene sequence of *Mtb* H37Rv was obtained from TubercuList (http://tuberculist.epfl.ch/), and the variant sequences were created using reported polymorphisms in the TB Drug Resistance Mutation Database (Sandgren et al., 2009).

Following Alaie et al. (2012), students first visually scan the two sequences for 1 minute to see if they can find a single-nucleotide polymorphism (SNP, pronounced "snip"). Because the portion of the *rpoB* gene provided is 320 nucleotides long, most students are not able to identify a SNP in the time allowed, and they are transitioned to the idea that computer programs and bioinformatics tools have been created to accomplish this task rapidly. Students are then introduced to the online bioinformatics sequence-analysis program ClustalW (http://www.genome.jp/tools/clustalw). After copying and pasting both their wild-type and variant *rpoB* sequences into ClustalW and running an alignment comparison, students examine the results to identify whether a SNP is present between their two sequences (Figure 2).

A SNP is a single-base-pair substitution that can be observed when comparing similar DNA sequences of the same gene – between organisms, strains, or homologous chromosomes (Figure 3A). 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, because multiple codons can code for the same amino acid (Figure 3B). Nonsynonymous polymorphisms, however, lead to a change in the protein sequence and can either cause

```
>wild-type_rpoB_gene
ATGGAGCGGGTGGTCCGGGAGCGGATGACCACCCAGGACGTGGAGGCGATCACACCGCAGACGTTGAT
CAACATCCGGCCGGTGGTCGCCGCGATCAAGGAGTTCTTCGGCACCAGCCAGCTGAGCCAATTCATGG
ACCAGAACAACCCGCTGTCGGGGTTGACCCACAAGCGCCGACTGTCGGCGCTGGGGCCCGGCGGTCTG
TCACGTGAGCGTGCCGGGCTGGAGGTCCGCGACGTGCACCCGTCGCACTACGGCCGGATGTGCCCGAT
CGAAACCCCTGAGGGGCCCAACATCGGTCTGATCGGCTCGCTGTCGGT

>variantD_rpoB_gene
ATGGAGCGGGTGGTCCGGGAGCGGATGACCACCCAGGACGTGGAGGCGATCACACCGCAGACGTTGAT
CAACATCCGGCCGGTGGTCGCCGCGATCAAGGAGTTCTTCGGCACCAGCCAGCTGAGCCAATTCATGG
ACCAGAACAACCCGCTGTCGGGGTTGACCCACAAGCGCCGACTGTCGGCGCTGGGGCCCGGCGGTCTG
TCACGTGAGCGTGCCGTGCTGGAGGTCCGCGACGTGCACCCGTCGCACTACGGCCGGATGTGCCCGAT
CGAAACCCCTGAGGGGCCCAACATCGGTCTGATCGGCTCGCTGTCGGT
```

Figure 1. Screen shot of one of the eight provided *Mtb* variant sequences to assign to students. Students are first asked to visually scan the two segments to find the SNP. They are then instructed to copy and paste the two sequences into ClustalW, a multiple-sequence-alignment program (http://www.genome.jp/tools/clustalw).

```
clustalw.aln

CLUSTAL 2.1 multiple sequence alignment


wild-type_rpoB_gene       ATGGAGCGGGTGGTCCGGGAGCGGATGACCACCCAGGACGTGGAGGCGATCACACCGCAG
variantD_rpoB_gene        ATGGAGCGGGTGGTCCGGGAGCGGATGACCACCCAGGACGTGGAGGCGATCACACCGCAG
                          ************************************************************

wild-type_rpoB_gene       ACGTTGATCAACATCCGGCCGGTGGTCGCCGCGATCAAGGAGTTCTTCGGCACCAGCCAG
variantD_rpoB_gene        ACGTTGATCAACATCCGGCCGGTGGTCGCCGCGATCAAGGAGTTCTTCGGCACCAGCCAG
                          ************************************************************

wild-type_rpoB_gene       CTGAGCCAATTCATGGACCAGAACAACCCGCTGTCGGGGTTGACCCACAAGCGCCGACTG
variantD_rpoB_gene        CTGAGCCAATTCATGGACCAGAACAACCCGCTGTCGGGGTTGACCCACAAGCGCCGACTG
                          ************************************************************

wild-type_rpoB_gene       TCGGCGCTGGGGCCCGGCGGTCTGTCACGTGAGCGTGCCGGCTGGAGGTCCGCGACGTG
variantD_rpoB_gene        TCGGCGCTGGGGCCCGGCGGTCTGTCACGTGAGCGTGCCGTGCTGGAGGTCCGCGACGTG
                          ****************************************  *******************

wild-type_rpoB_gene       CACCCGTCGCACTACGGCCGGATGTGCCCGATCGAAACCCCTGAGGGGCCCAACATCGGT
variantD_rpoB_gene        CACCCGTCGCACTACGGCCGGATGTGCCCGATCGAAACCCCTGAGGGGCCCAACATCGGT
                          ************************************************************

wild-type_rpoB_gene       CTGATCGGCTCGCTGTCGGT
variantD_rpoB_gene        CTGATCGGCTCGCTGTCGGT
                          ********************
```

Figure 2. The results displayed from the ClustalW sequence alignment. The SNP is located where there is no asterisk.

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 3B).

Part B: Translating DNA into Amino Acids

From prior instruction, students should know how to transcribe and translate a DNA sequence into RNA and then into amino acids using a codon chart. Part B asks students to use an online DNA-to-protein translation tool, GeneMarkS (Besemer et al., 2001), to translate their two DNA sequences directly into amino acid sequences.
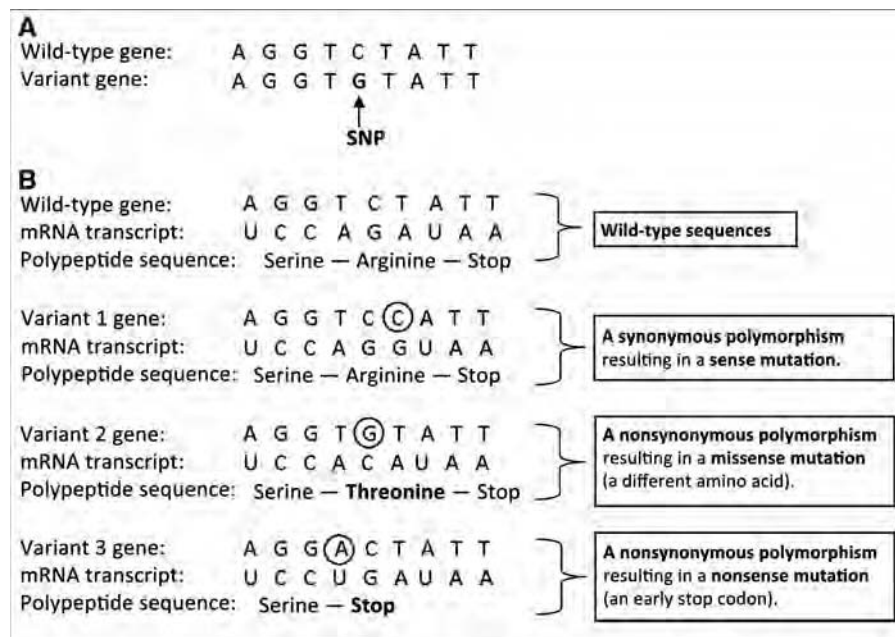
Part C: Identifying Synonymous & Nonsynonymous Mutations

A wild-type strain of *Mtb* does not have any gene mutations that confer antibiotic resistance, which makes 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 SNPs that 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.

Part C instructs students to use ClustalW for a second time to align their two amino acid sequences to determine whether their variant *Mtb* protein sequence contains a mutation (Figure 4). If there is a SNP in the DNA sequence but not an amino acid change, this indicates a synonymous mutation. If a SNP in the DNA sequence leads to an amino acid change, this indicates a nonsynonymous mutation. If a nonsynonymous mutation is observed, students are asked to write down the location of the mutation in the amino acid sequence as well as the chemical properties of both the wild-type and the variant amino acid (Table 1).

Part D: Hypothesizing Antibiotic Resistance

Antibiotics often target microbial proteins at particular binding sites, which then disrupt the normal function of the protein. Because the protein no longer works when the antibiotic is bound, the bacterium can't carry out its normal functions and will ultimately perish. For example, rifampin binds to and inhibits a component of the *Mtb* RNA polymerase protein, which is coded by the *rpoB* gene. If the bacterium's RNA polymerase
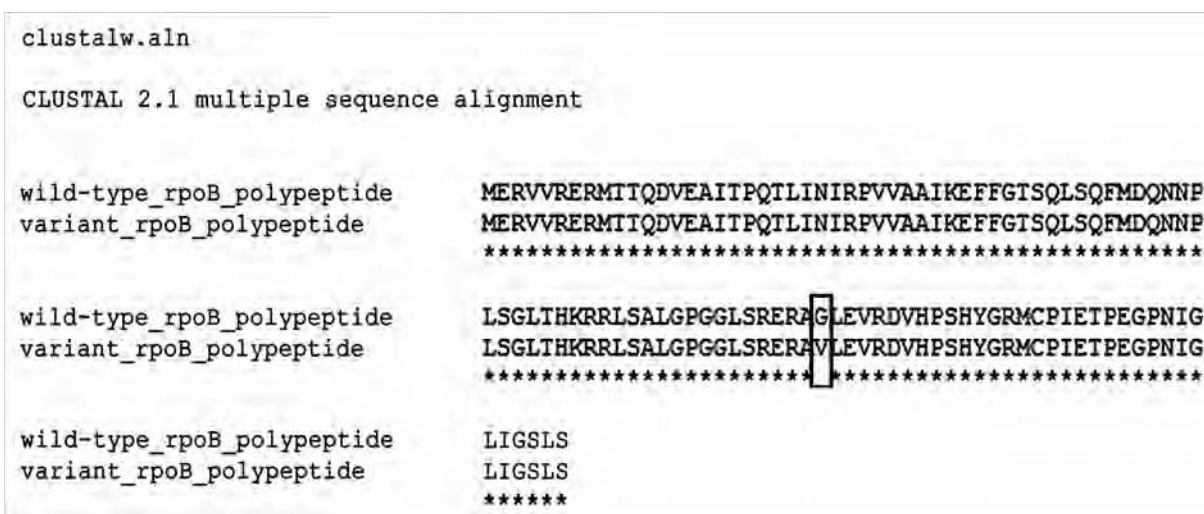

Figure 3. (a) A single-nucleotide polymorphism between a wild-type and a variant gene. (b) Possible consequences of SNPs in the resulting polypeptide.

Figure 4. Results from the ClustalW polypeptide sequence comparison. Asterisks indicate amino acids that are identical. A colon, period, or blank space indicates a changed amino acid. In this case, a valine (V) in the variant substitutes for a glycine (G) in the wild type.

Table 1. Part C: Amino acid comparison table for student completion. Answers for the Variant D TB protein are provided.

	Wild-type TB Protein	Variant D Protein
Numerical position of the amino acid difference	(Count the letters until you get to the difference.) Number 74	
Amino acid single letter symbol	G	V
Amino acid full name	Glycine	Valine
Chemical property of the amino acid	Nonpolar	Nonpolar

Notes: In this comparison between the wild-type polypeptide and variant D polypeptide for RNA polymerase, the mutation occurs at amino acid position 74, which is outside of the rifampin binding site (amino acids 36–67). Additionally, despite the change in the amino acid sequence, the chemical property of the variant amino acid remains the same as the wild type. Therefore, the structural binding site for rifampin will likely not change. Thus, a student is expected to claim that the variant D *Mtb* strain remains susceptible to rifampin.
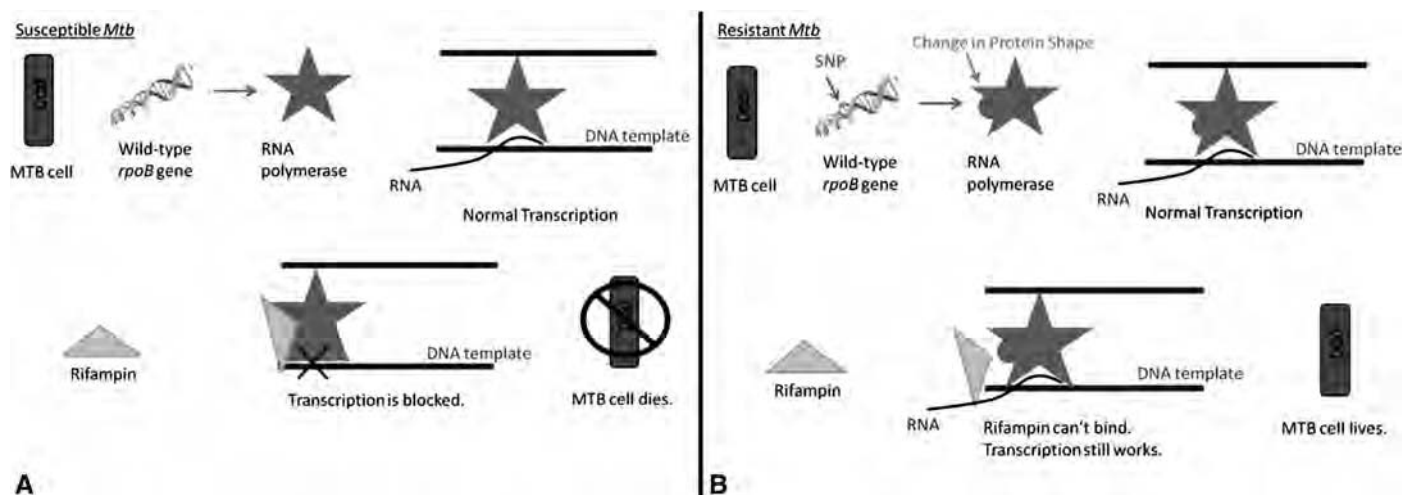

Figure 5. Diagram of the action of rifampin on (a) an *Mtb* RNA polymerase that rifampin can bind to and (b) an *Mtb* RNA polymerase that rifampin is not capable of binding to, because of a change in the structure of the rifampin binding site.

doesn't work, bacterial DNA can no longer be transcribed into RNA, and the production of RNA and proteins subsequently stops (Figure 5A).

Bacteria that are resistant to rifampin often have a SNP in their *rpoB* gene that alters the shape of the binding site where rifampin typically binds to the RNA polymerase protein. The resulting mutation prevents rifampin from binding, allowing the polymerase to continue to function and enabling the bacteria to survive (Figure 5B).

In general, if there is no amino acid change between the wild-type and variant *Mtb* strains, we expect the strain to remain susceptible to the drug. If, however, an amino acid change is observed, the chemical property and the location of the mutation in relation to the rifampin binding region can be used to assess impact on resistance. For our lesson plan, if the amino acid mutation has a different chemical property than the wild-type amino acid, and if it occurs within the rifampin binding site, we assume resistance. If the mutation occurs outside the rifampin binding site, or if the substituted amino acid has the same chemical property as the wild type's amino acid, we assume that the strain remains susceptible.

To engage the students in critical thinking, Part D asks students to evaluate the observations they made of the *rpoB* DNA and protein alignments in ClustalW. As background information, students are given the location of the binding site of rifampin in relation to the protein sequence (between amino acids 36 and 67 of the 106-amino-acid translated sequence). Using a claims–evidence–reasoning approach for writing a scientific explanation (McNeill & Krajcik, 2007), students make a claim based on their collected observations, stating whether their *Mtb* variant strain will be resistant or susceptible to rifampin. Students provide specific evidence from their comparisons to support their claim and are asked to write a reasoning paragraph that justifies how their evidence supports their claim. Final versions of the scientific explanation are collected and graded using the provided rubric (Table 2). A key to the likely claims for the eight variant *Mtb* strains is displayed in Table 3. Groups who finish early are allowed to choose a second variant *Mtb* strain to investigate. All background information, student directions, and DNA sequences for this lesson can be found at our corresponding website (http://www.stronglab. org/taylor).

⊙ Results

This 90-minute activity was piloted to four sections of ninth- and tenth-grade biology students at a traditional, comprehensive urban high school during the spring semester of 2013. Prior instruction to this lesson included a segment on drug-resistant tuberculosis from PBS's *Evolution* video series (PBS, 2001), a β-hemoglobin sickle cell disease lesson (BSCS, 2006), and an overview of the four levels of protein structure.

Table 2. Rubric for the assessment of students' scientific explanations.

	Excellent	Proficient	Developing	Needs Improvement
Claim	Sentence is an accurate statement of ALL outcome(s) of the experiment and is supported by the evidence provided.	Sentence is a statement of MOST of the outcomes of the experiment and is supported by evidence provided.	Sentence is a claim that is not supported by the evidence.	No claim is written.
	5	4	3	0
Evidence	Results of your SNP comparison between the wild-type and variant TB strains are reported. Results of your amino acid comparison between the two strains are reported.	A minor change to the evidence statement(s) is needed.	Moderate changes to the evidence statement(s) are needed.	No evidence is written OR evidence is not reported using numbers or direct observations.
	5	4	3	2 1 0
Reasoning	Reasoning correctly shows that the student understands, *to a level above proficiency*, that • A change in amino acid sequence AND properties within the binding site for rifampin will cause the bacterium to be resistant to rifampin OR • Because the amino acid sequence and properties stay the same, the bacterium will be susceptible to rifampin. • The structure of the protein CHANGES (or DOESN'T) depending on the amino acid sequence.	Reasoning shows that the student understands that • A change in amino acid sequence AND properties within the binding site for rifampin will cause the bacterium to be resistant to rifampin OR • Because the amino acid sequence and properties stay the same, the bacterium will be susceptible to rifampin.	Reasoning shows that the student has a partial understanding of • How a change in DNA can produce a change in amino acid sequence (or not). • A different amino acid sequence can change the structure of the protein. Some gaps exist in the student's understanding of the concepts of gene expression.	No reasoning is written OR the reasoning shows that the student has an incomplete or incorrect understanding of • Gene expression • Antibiotic resistance Many gaps exist in the student's understanding of gene expression.
	10	9 8	7 6	5

Students were intentionally paired into groups of two based on academic performance. Because of the variety of possible outcomes of a SNP in the eight variant *rpoB* genes, there is an inherent discrepancy in the level of difficulty in the analysis of one variant from the next. This feature of the activity allows for strategic differentiation for each pair of students. Lower-performing students were paired with middle-performing students and assigned *Mtb* variants A or B with synonymous mutations. Variants with nonsynonymous mutations that required more analysis (C–H) were assigned to the middle- to higher-performing pairs. Informal observations were made of the students while they were doing the activity, and we also gathered student feedback via a short survey.

During the activity, students were fully engaged in the prospect of determining whether their variant strain was resistant to rifampin or not. They eagerly worked to find the SNP by visual recognition, and they were pleasantly surprised with the ease of producing and interpreting the sequence comparisons of their gene and polypeptide sequences using ClustalW. Seventy-four percent of students said that this lesson increased their understanding of the molecular consequences of mutations and that they would recommend this lesson be taught to other high school students (Figure 6). Most students seemed to appreciate the experience of using bioinformatics websites to make observations regarding the *rpoB* gene and polypeptide sequences, as evident in their feedback. When asked what the best part of this lesson was, student comments included the following:

- "Getting to work on the molecular consequences of mutations hands-on."
- "Learning how different tools can be used to find SNPs."
- "Using the program[s] to [translate] and match up the sequences."
- "I liked this. It helped to apply what we've learned to something bigger."
- "I liked how the different websites can help you compare different genes."

Table 3. Answer key for the 8 variant *Mtb* strains.

| *Mtb* Variant	What is the location of the mutation in the polypeptide sequence? Is the mutation in rifampin's binding site (aa 36–67)?	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 mutation result in sense, missense, or nonsense in the variant polypeptide?	What is your claim? Will this variant be resistant or susceptible to rifampin?
A	Not applicable	No change	Not applicable; no amino acid change	Sense	Susceptible
B	Not applicable	No change	Not applicable; no amino acid change	Sense	Susceptible
C	30, outside	Alanine to Threonine	Yes, from nonpolar to polar	Missense	Susceptible
D	74, outside	Glycine to Valine	No, from nonpolar to nonpolar	Missense	Susceptible
E	56, inside	Histidine to Aspartic acid	Yes, from basic to acidic	Missense	Resistant
F	51, inside	Leucine to Valine	No, from nonpolar to nonpolar	Missense	Susceptible
G	43, inside	Glutamine to STOP	A stop codon comes too early	Nonsense	Resistant
H	46, inside	Aspartic acid to Tyrosine	Yes, from acidic to polar	Missense	Resistant
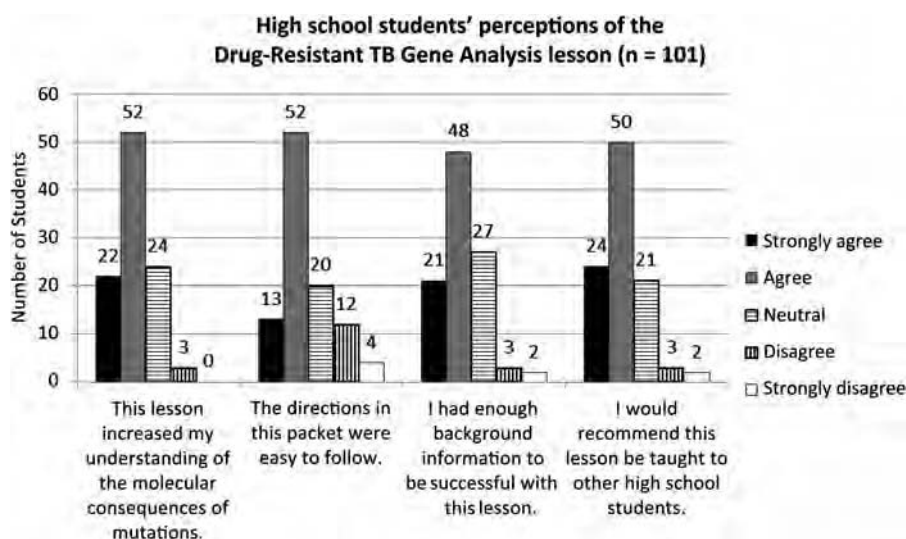
Figure 6. Results from a feedback survey given after the completion of the Drug-Resistant TB Gene Analysis lesson.

Discussion

Many of the bioinformatics lessons currently reported in the literature are geared toward AP Biology or undergraduate biology student populations (Honts, 2003; Gallagher et al., 2011; Alaie et al., 2012; May, 2013). From our experience with this lesson, ninth- and tenth-grade students are also capable of accurately using bioinformatics tools to gather evidence to support a claim.

This activity introduces students to the tools that many contemporary biologists use to identify SNPs and serves as a logical extension from translating DNA by hand to online alignment and SNP visualization using online bioinformatics tools. The inquiry and analysis of the molecular consequences of SNPs within the context of antibiotic resistance in *Mtb* challenges students to transfer their understanding of the structure and function of DNA and proteins into a new disease model of global importance – tuberculosis. Through the assignment of writing a scientific explanation, students gain more experience writing "arguments to support claims in an analysis of a topic, using valid reasoning and relevant and sufficient evidence" (Common Core State Standard W.9–10.1).

Although some students felt that the directions were at times long or difficult to follow (as assessed by student evaluations), all students were able to finish through Part D of the lesson, and all students were able to write rough drafts of their scientific explanations during the same class period.

In addition, this lesson revisits the disciplinary core idea of the process of natural selection taught earlier in the year by showcasing, at the molecular level, the effect that a single base-pair substitution can have on the survival of a bacterium. This activity also highlights the important role of online bioinformatics tools and resources that enable biologists to perform analyses efficiently and accurately. We believe that this lesson plan can be a useful addition to any high school or undergraduate biology class, particularly those interested in emphasizing the concepts and impacts of DNA mutations. In addition, instilling an appreciation of bioinformatics methods early on in the high school curriculum may help reach an extended segment of the student population, enticing students with an interest in technology or computers to consider potential career paths in biology or computational biology.

○ Acknowledgments

References

Alaie, A., Teller, V. & Qiu, W. (2012). A bioinformatics module for use in an introductory biology laboratory. *American Biology Teacher, 74,* 318–322.

Besemer, J., Lomsadze, A. & Borodovsky, M. (2001). GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. *Nucleic Acids Research, 29,* 2607–2618.

Bloom, M. (2001). Biology in silico: the bioinformatics revolution. *American Biology Teacher, 63,* 397–403.

BSCS. (2006). *BSCS Biology: A Human Approach, 3rd Ed.* Dubuque, IA: Kendall Hunt.

Gallagher, S.R., Coon, W., Donley, K., Scott, A. & Goldberg, D.S. (2011) A first attempt to bring computational biology into advanced high school biology classrooms. *PLoS Computational Biology, 7,* e1002244.

Herron, S.S., Parr, J., Davis, B. & Nelson, P. (2010). Theme-based instruction: making conceptual ties with the sickle cell story. *American Biology Teacher, 72,* 422–426.

Honts, J.E. (2003). Evolving strategies for the incorporation of bioinformatics within the undergraduate cell biology curriculum. *Cell Biology Education, 2,* 233–247.

Lewitter, F. & Bourne, P.E. (2011). Teaching bioinformatics at the secondary school level. *PLoS Computational Biology, 7,* e1002242.

Machluf, Y. & Yarden, A. (2013). Integrating bioinformatics into senior high school: design principles and implications. *Briefings in Bioinformatics, 14,* 648–660.

May, B.J. (2013). Engaging students in a bioinformatics activity to introduce gene structure and function. *Journal of Microbiology & Biology Education, 14,* 107–109.

McNeill, K.L. & Krajcik, J. (2007). Inquiry and scientific explanations: helping students use evidence and reasoning. In J. Luft, R.L. Bell, & J. Gess-Newsome (Eds.), *Science as Inquiry in the Secondary Setting*. Arlington, VA: NSTA Press.

National Governors Association Center for Best Practices and Council of Chief State School Officers. (2010). *Common Core State Standards in Writing*. Washington, DC: Common Core State Standards Initiative.

NGSS Lead States. (2013). *Next Generation Science Standards: For States, By States* (HS-LS1 From Molecules to Organisms: Structures and Processes). Washington, DC: National Academies Press.

PBS. (2001). Episode 4: the evolutionary arms race. *Evolution.* PBS video.

Sandgren, A., Strong, M., Mutukrishnan, P., Weiner, B.K., Church, G.M. & Murray, M.B. (2009). Tuberculosis drug resistance mutation database. *PLoS Medicine, 6,* 2.

JESSICA M. TAYLOR is a Biology Honors Teacher at East High School in Denver, CO 80206; e-mail: jessica_taylor@dpsk12.org. REBECCA M. DAVIDSON is an Instructor Level Bioinformatics Researcher at National Jewish Health, Denver, CO 80206. MICHAEL STRONG is an Assistant Professor at National Jewish Health in Denver and at the University of Colorado, Denver; e-mail: strongm@njhealth.org.

Appendix 1: Student Learning Objectives

By the end of this activity, students 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 their 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, and write a scientific explanation justifying a claim with evidence and reasoning.

Appendix 2: Student Instructions & Questions

Essential question: Does your assigned variant *rpoB* gene sequence result in resistance to the antibiotic rifampin?

Part A: Identifying SNPs

1. 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. Be sure to stop after one minute.

Student Questions

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

3. Comparing allele sequences by hand is a time-consuming process. Fortunately, computer programs have been created to make this task happen almost instantaneously.

    (a) 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.

    (d) Click "Execute Multiple Alignment."

    (e) On the page that comes up, scroll down to the section under the heading "clustalw.aln."

    (f) Look at the alignment of your two sequences. Stars (***) indicate bases that are identical. **An empty space indicates a SNP**.

Student Questions

What is the base change? Answer: The base in the wild-type TB allele is a(n) _____ while the base in my variant allele is a(n) _____. 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? Look back at the top of your ClustalW alignment results. Locate how many base pairs (bp) are in this allele segment: _____ bp. Describe mathematically how you could calculate how many amino acids are coded for by this segment of the allele. Perform that calculation (Round down to the nearest whole number). Answer: This segment of the rpoB allele codes for a polypeptide sequence 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.

   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 both gene sequences into the large box. Be sure to include the sequence labels again.

    (d) Under "Output options," mark the boxes "Protein sequence" and "Gene nucleotide sequence."

    (e) Click "Start GeneMarkS."

    (f) On the next page, click on the "gms.out.faa" link to retrieve protein sequences.

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

    (g) Copy both amino acid sequences.

Part C: Identifying Synonymous & Nonsynonymous Mutations

In order to determine if the SNP in your variant DNA 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 allele sequences) and determine if the SNP causes a synonymous or nonsynonymous mutation in the variant protein. Here's how to do that.

1. Return to (or reopen) ClustalW at http://www.genome.jp/tools/clustalw/.

    (a) Next to "Enter your sequences…" click on **Protein**.

    (b) Copy and paste the **wild-type TB and variant TB amino acid sequences** from GeneMarkS into ClustalW.

(c) Click "Execute Multiple Alignment."

(d) 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.**

Student Questions

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 Step 4. If they are different, continue with Step 2.

2. For variants C–F and H, fill in the chart (Table 1) to compare the different amino acid between the two sequences. For variant G, skip to Step 3. To determine what amino acid each letter represents and the chemical property of each amino acid, click on "Copymasters" for reference charts.

3. For Variant G, provide an explanation for why you think there are missing amino acids at the end of the variant protein.

4. 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

A SNP in the gene region that changes the structure of the binding site for rifampin in the RNA polymerase 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 bacterium 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 questions and the information above, make a claim as to whether the bacteria containing your given variant protein will be resistant or susceptible to rifampin.

> **Claim:** *I predict that* Mycobacterium tuberculosis *with the rpoB variant ____ allele will be (resistant or susceptible) to rifampin.*
>
> **Evidence:** *The evidence that supports my claim is (provide evidence that supports your claim)….*
>
> **Reasoning:** *The reason for my claim is that (provide scientific justification for your claim)….*