University at Buffalo *The State University of New York* | REACHING OTHERS

SEPA SCIENCE EDUCATION PARTNERSHIP AWARD
Supported by the National Institutes of Health

Introduction to STEM Using *Bordetella pertussis CS* Genes BPTD_0032 and BPTD_0034-0037

Morgan Farman, Tyler Norris, Nicholas Simpson, Cody Ullock and Lisa Mallaber
Livonia High School and the Western New York Genetics in Research Partnership

Abstract

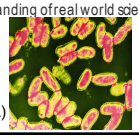
Students at Livonia High School manually annotated a portion of the *Bordetella pertussis CS* genome. Following a set of modules set out by the Genomics Education National Initiative Annotation Collaboration Toolkit (GENI-ACT), genes BPTD_0032 and BPTD_0034-0037 were analyzed. GENI-ACT allowed students to gather information and draw conclusions based on the data obtained from completing specific modules. This hands on approach, using actual genetic data, allowed students to discover and contribute to the scientific process. The Genbank proposed gene product name did not differ significantly from the proposed gene annotation for each of the genes in the group and as such, the genes appear to be correctly annotated by in their database. The one exception to this, is the call for a signal peptide by Phobius for the protein coded by gene BPTD_0035, may be in error.

Introduction

Bordatella pertussis is the causative agent of a respiratory illness, commonly known as whooping cough. Although vaccination programs world-wide have decreased the incidence of death from this disease, it still effects millions of people. This organism is an aerobic, non-motile, gram negative encapsulated coccobacillus, which releases a toxin that stops the cilia of epithelial cells found in the respiratory tract from beating. This prevents cilia from clearing debris form the lungs which sends the host into coughing fits. The bacteria is spread by airborne droplets from the host via coughing, laughing and sneezing and has an incubation period that averages 9-10 days.

Using the application of computer science and information technology to complex biological information, such as DNA, RNA or protein sequences, this technology can be used to better understand how DNA, amino acids, and proteins are related, and how they work within an organism. Through the GENI-ACT program and a partership created between the University of Buffalo and Western New York Genetics in Research Partnership a workforce has been created to check the accuracy of the computer output programs of BLAST, T-Coffee, TMHMM, WebLogo, TIGRfam, Pfam, Phobius, PSORT, SignalIP as well as others, to the proteins called for by genes from virulent organisms.

It is thought that *B. pertussis* has the ability to inhibit the function of the host's immune system. The toxin, known as pertussis toxin inhibits G protein coupling that regulates an adenylate cyclase-mediated conversion of ATP to cyclic AMP. This can cause disturbances in cellular signaling mechanisms and prevent phagocytes from correctly responding to an infection. In order to increase the workforce that can analyze such results, researchers at the NSF partnered with the University at Buffalo to work with local high schools, introducing younger students to the field of bioinformatics. This program allows students to gain a deeper understanding of real world science and STEM.

A) 

B) 

A) Magnified image of *B. Pertussis*. B) Magnified image of *B.Pertussus* infecting cilia lining the human respiratory tract

Methods and Materials

Modules of the GENI-ACT (http://www.geni-act.org/) were used to complete *Bordatella pertussis CS* genome annotation. The modules are described below:

Modules	Activities	Questions Investigated
Module 1- Basic Information Module	DNA Coordinates and Sequence, Protein Sequence	What is the sequence of my gene and protein? Where is it located in the genome?
Module 2- Sequence-Based Similarity Data	Blast, CDD, T-Coffee, WebLogo	Is my sequence similar to other sequences in Genbank?
Module 3- Cellular Localization Data	Gram Stain, TMHMM, SignalP, PSORT, Phobius	Is my protein in the cytoplasm, secreted or embedded in the membrane?
Module 4- Alternative Open Reading Frame	IMG Sequence Viewer For Alternate ORF Search	Has the amino acid sequence of my protein been called correctly by the computer?
Module 5- Structure-Based Evidence	TIGRfam, Pfam, PDB	Are there functional domains in my protein?

Results

BPTD_0032: This gene codes for Glycyl-tRNA synthetase. It catalyzes the synthesis of glycyl-tRNA, which is required to insert glycine into proteins.

BPTD_0034: The gene annotated identified with this strain is found to be most similar to a phosphatase enzyme whose function is to remove a phosphate groups from proteins. This is interesting because this specific enzyme could be attributed to the phosphatases of an entire group of phosphates from a protein. This gene varied greatly from the other tested genes. Its WebLogo varied significantly showing it being much less conserved than the WebLogo in Figure I which shows a trend between the other genes in which they are all highly conserved varying from this one. This may show an interesting comparison between the conservation of energy in the gene and the phosphatases going on. Because phosphatases is a reverse kinases which means instead of creating ATP energy it may be losing it.

BPTD_0035: This protein is found in the cytoplasmic trans membrane. Its function is to transport lipids through the membrane. It is not a signal peptide and therefore does not leave the cell. It is very conserved and has not changed much over the years.

BPTD_0037: This gene codes for glyoxalase. This enzyme detoxifies methylglyoxal, an organic metabolic substance, which is very toxic and can harm other parts of the cell.. One very different characteristic about glyoxalase is that it doesn't use a specific metal ion to bond the atoms together in it's compound. While some enzymes need certain metals such as iron, magnesium, and copper or it is unable to carry out it's function.

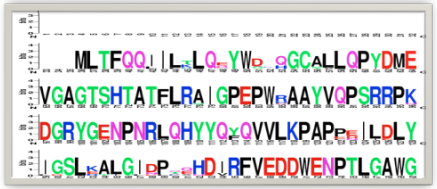


Figure I – Above is the WebLogo sequence generated from T-Coffee alignment. The letters representing amino acids are rather large even though they stem from different genera. The large gap at the beginning is caused by one protein having an extra amino terminal.

This WebLogo was an interesting output because the size of the letters representing was so large. Thus leading to the conclusion that this sequence of DNA is highly conservative. This means that this gene is high up on the phylogenic tree and codes for basic cellular function, stability or reproduction. The information provided by WebLogo supports the initial idea that the BPTD_0032 locus of *Bordetella pertussis* CS is a Glycyl-tRNA synthetase.

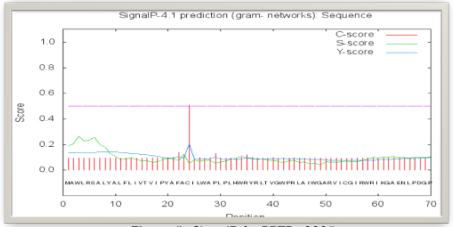


Figure II: SignalP for BPTD_0035

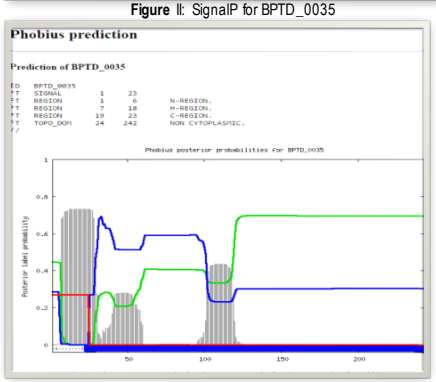


Figure III: Phobius for BPTD_0035
The colored lines show the probability of this protein's location. The higher the line is the higher probability. The grey area is where the trans membrane is located.

(**Figure III** continued…..) According to the graph my protein is found in the membrane due to the grey areas and it shows that it's a trans membrane helix while also showing that it is a signal peptide. This is unique because earlier in my SignalP (Figure II) it was shown that the protein was not a signal peptide. This may indicate something unique about my protein.

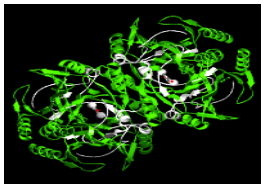


Figure IV: This figure shows the protein folding for the enzyme Glyoxalase, coded for by the gene BPTD_0037

Conclusion

Figure V Map of the gene neighborhood annotated

The GENI-ACT proposed gene products did not differ significantly from the proposed gene annotation for each of the genes in the group and as such, the genes annotated by the computer database.

Gene Locus Tag	GENI-ACT Gene Product	Proposed Annotation
BDPT_0032	Glycyl-tRNA sythetase	Glycyl-tRNA sythetase
BDPT_0034	Phosphatase	Phosphatase
BDPT_0035	Cytoplasmic trans-membrane/(possible signal peptide)	Cytoplasmic trans-membrane
BDPT_0037	Glyoxalase	Glyoxalase

The purpose of this project was to introduce high school students to STEM and bioinformatics. Students involved in our first project agreed that this was a valuable enrichment project and they learned a lot more about genetics and the power of bioinformatics than they would have in a typical biology classroom setting. Overall, the project resulted in students gaining a basic understanding of bioinformatics and STEM.

References

Guiso et al. (2009). *Bordatella pertussis and Purtussis vaccines.* Clinical Infectious Disease v99 (10) p1565-1569 Oxford Journal.

http://www.CDC.gov/pertussis/lab.html

Acknowledgments

This project was supported by NIH SEPA Award R25ODO10536
Special thanks to:
Dr. Rama Dey-Rao Clinical Assistant Professor, University at Buffalo
Dr. Stephen Koury Research Associate Professor, University at Buffalo