Annotation of the *Proteus mirabilis* HI4320 Genome at Locus Tag PMI_RS10005

Peyton Martin, and Brian Lewis
Mount Morris Central School, NY and the Western New York Genetics in Research and Health Care Partnership

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

Abstract

Proteus mirabilis is a bacterium that is gram negative. This organism has a gene called PMI_RS10005 that I have been studying. This gene is most similar to *"HlyD family type I secretion periplasmic adaptor subunit -product"*. This gene is responsible for secreting proteins into the periplasm. I have learned that the DNA sequence in this gene is very similar to that of other bacteria that secrete proteins.

Introduction

I was given the gene PMI_RS10005, along with some basic information about said gene. I was given information such as, its coordinates, length, products along with its amino acid and nucleotide sequences. With this information I was able to use some bioinformatics tools which gave me even more information about the gene. Some of the tools that were used were, Blast, T-Coffee, and Pfam.

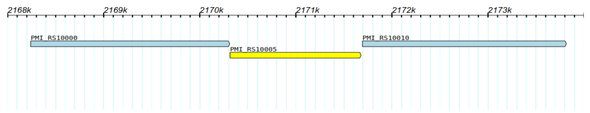

Figure 1. The locus tags and relative position my gene and two similar genes.

Methods

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

Modules	Activities	Questions Investigated
Basic Information	DNA Coordinates and Sequence, Protein Sequence	What is the sequence of the gene and protein? Where is it located in the genome?
Sequence-Based Similarity	Blast, CDD, T-Coffee, WebLogo	How similar is the protein under investigation to other proteins in GenBank?
Structure-Based Similarity	TIGRfam, Pfam, PDB	What functional domains are present in the protein under investigation?
Cellular Localization	Gram Stain, TMHMM, SignalP, LipoP, Psortb, Phobius	Is the protein under investigation located in the cytoplasm, secreted, in the periplasm or embedded in the cell membrane or cell wall?
Enzymatic Function	KEGG, MetaCyc, E.C. Number	In what process or structure is the protein under investigation involved?
Duplication and Degradation	Paralog, Pseudogene	Are there other forms of the protein under investigation in the same genome? Is it functional?
Horizontal Gene Transfer	Phylogenetic Tree, Gene Neighborhood, GC Content	Has the protein under investigation co-evolved with the rest of the genome or has it been obtained in a different way?
RNA family	Rfam	Does the gene under investigation encode a functional RNA?
Final Annotation	Evaluate data from all modules	Has the gene been correctly called by the pipeline annotation?

Results

From my research, I have found out a lot about my gene. My gene has the Pfam number PF13437 and is HlyD family secretion protein. Further information was found using Pfam: "HlyD is a component of the prototypical alpha-haemolysin (HlyA) bacterial type I secretion system, along with the other components HlyB and TolC. HlyD and HlyB are inner-membrane proteins and specific components of the transport apparatus of alpha-haemolysin. HlyD is anchored in the cytoplasmic membrane by a single transmembrane domain and has a large periplasmic domain within the carboxy-terminal 100 amino acids HlyB and HlyD form a stable complex that binds the recombinant protein bearing a C-terminal HlyA signal sequence and ATP in the cytoplasm. HlyD, HlyB and TolC combine to form the three-component ABC transporter complex that forms a trans-membrane channel or pore through which HlyA can be transferred directly to the extracellular medium. Cutinase has been shown to be transported effectively through this pore."(pfam.xfam.org). Pfam also provided the HMM logo as seen in figure 3.

Figure 2. The alignment scores from the Blast simulation using the nr database.

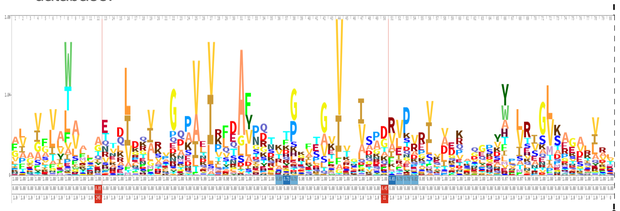

Figure 3. HMM logo for PF13437 generated by Pfam.

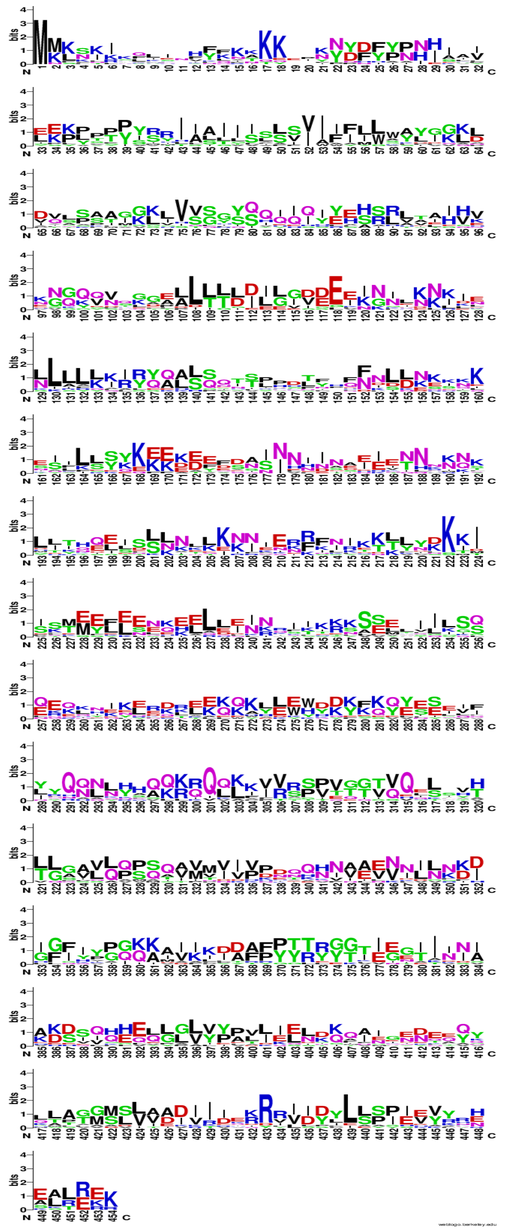

Figure 4. Comparison of my gene and other genes using WebLogo.

The simulations such as Pfam and T-Coffee (with WebLogo used to graph the results in an easier to read way) have provided a lot of insight as to how my gene compares to other genes.

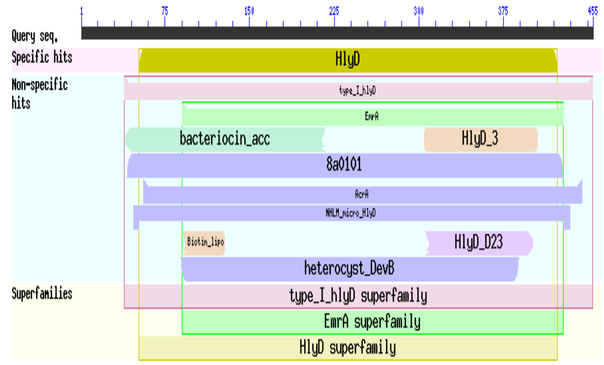

Figure 5. CDD results from Blast simulation using nr database.

Conclusion

I have learned that my gene, PMI_10005 from *Proteus mirabilis*, is a very common gene. A lot of the sequences in the gene are similar to that in other genes whose purpose is secreting proteins into periplasm. I derived this conclusion from my results from simulations such as Blast, T-Coffee and Pfam.

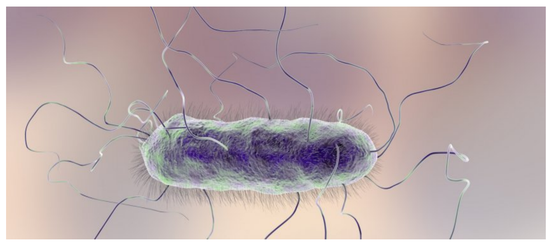

Figure 6. Picture of the bacteria *Proteus mirabilis*.

References

http://www.ncbi.nlm.nih.gov/blast
http://www.ebi.ac.uk/Tools/msa/tcoffee/
http://WebLogo.berkeley.edu/
http://www.pfam.xfam.org/

University at Buffalo The State University of New York
Jacobs School of Medicine and Biomedical Sciences
buffalo.edu

Acknowledgments
Supported by an NIH Science Education Partnership (SEPA) Award - R25GM129209