Annotation of the *Helicobacter pylori 83* Genome from DNA Coordinates (82463 to 83719), (89946 to 90623) and (90605 to91042)

Wesley Dorow and Lawrence Hohl. East Rochester High School, East Rochester, New York and The Western New York Genetics in Research and Health Care Partnership

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

University at Buffalo

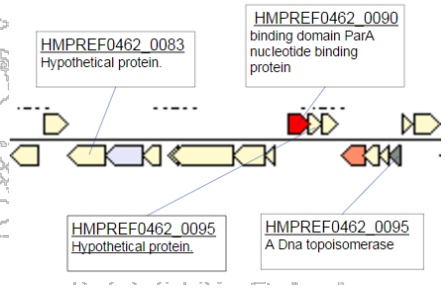
Abstract

This year, I annotated 4 genes from *Helicobacter pylori 83* using the modules given by GENI-ACT. Following the instructions given in our GENI-ACT manual, we went to many different sites that aided us in annotating our genes. These included BLAST, CDD, T-coffee, WebLogo, Gram Stain, TMHMM, SignalP, Psort, Phobius, IMG EDU, TIGRfam, Pfam, PDB, KEGG, MetaCyc and Rfam. I recorded these findings on the online GENI-ACT lab notebook. My results were established using all 4 annotations. Finallyl am presenting our results on this trifold poster. I added pictures and diagrams to help show what our genes do, and the processes used to annotate them. The most frustrating part about this year, was that three of my four genes were hypothetical proteins. This meant that most websites provided little info on the gene. My 4 genes were HMPREF0462_0083, HMPREF0462_0090, HMPREF0462_0091 and HMPREF0462_0095. Only 83's annotation remained the same, while the others changed slightly, and sometimes dramatically. In 90, I learned it contained a nucleotide binding protein domain, 91 was hypothetical, but I learned it was transmembrane, and 95 was annotated as a hypothetical, but I learned it may be a DNA topoisomerase.

Introduction

The gram negative *Helicobacter pylori* 83 bacterium can alter the human regulatory mechanisms for gastric acid production. During initial, *Helicobacter pylori* 83 can decrease acid secretion levels in the stomach. This can result in ulcers in the stomach, and even in the duodenum. For many years, people believed that smoking, stress, spicy food, and other lifestyle caused stomach ulcers. It was later found that *Helicobacter pylori* 83 was ingested through unclean food. The pylorus is the sphincter muscles between the duodenum and the stomach. This research is very important because it will help us learn new ways to fight diseases and bacteria, such as *Helicobacter pylori* 83.

Of the four genes that I was given, 2 of them were hypothetical proteins. This unfortunately meant that little information was known about them. This made it extremely difficult to complete the task of annotating them. These hypothetical proteins are proteins that have had very little work in research done on them. I used the resources given to on the geni-act website, I tried to validate the initial proposed annotation of the genomes. Through this, I also learned the new things about the genes that I annotated. Below shows the ortholog neighborhood of the genes I annotated.



Methods

Modules of the GENI-ACT (http://www.geni-act.org/) were used to complete Helicobacter pylori 83 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?
Module 6- Enzymatic Function	KEGG, MetaCyc, E.C. Number,	In what process does my protein take part?
Module 7- Gene Duplication/ Gene Degradation	Paralog, Pseudogene	Are there other forms of my gene in the bacterium? Is my gene functional?
Module 8- Evidence for Horizontal Gene Transfer	Phylogenetic Tree,	Has my gene co-evolved with other genes in the genome?
Module 9- RNA	RFAM	Does my gene encode a functional RNA?

Results

HMPREF0462_0083

This gene was initially proposed as Hypothetical Proteins, and this remained consistent throughout. This gene have had very little information on them. Few results came up on most sources.

HMPREF0462_0090

The proposed annotation of this gene was that it was a ParA nucleotide binding domain protein. This gene is the gene that separates and partitions the two chromosomes in a chromosome pair. The ParA protein helps with chromosomal replication in helicobacter pylori 83. The initial proposition of the gene remained constant through the study.

HMPREF0462_0091

This was initially proposed as Hypothetical Proteins, and this remained consistent throughout. This gene have had very little information on them. Few results came up on most sources. I did learn that HMPREF4062_0091 was transmembrane.

HMPREF0462_0095

The initial proposed annotation of HMPREF0462_0095 was actually a hypothetical protein. Using Blast, IMG, MetaCyc and ExPASy ENZYME, determined that it actually was a DNA Topoisomerase. The DNA topoisomerase in an enzyme that aids in the breakage and rejoining of single stranded DNA.

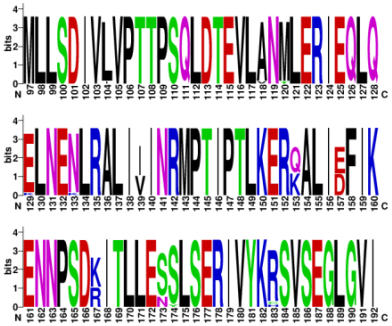


Figure 1- a snippet of the weblogo of HMPREF0462_0090. This weblogo is extremely well conserved. This shows tht the ParAprotein is a crucial part of Helicobacter pylori 83

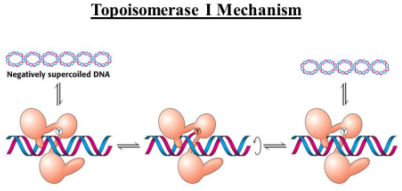


Figure 2- This shows the actual function of a Dna topoisomerase. It cleaves a strand of DNA, then it realigns them, making a relaxed strand of dna.

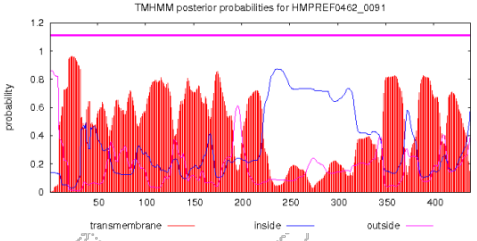


Figure 3, even though HMPREF4062_0091 was a Hypothetical protein, Tmhmm said that it very likely is transmembrane.

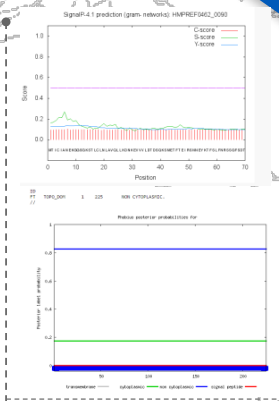


Figure 3- example of the results given, when I attempted to annotate HMPREF4062_0083 and HMPREF4062_0091 showing negative results for the presence of a signal peptide by TMHMM and negative results for a signal peptide and transmembrane helixes by Phobius

Conclusion

	Initial proposed product	Product after
HMPREF4062_0083	Hypothetical protein	Hypothetical protein
HMPREF4062_0090	ParA protien	ParA nucleotide binding domain protein
HMPREF4062_0091	Hypothetical protien	Still hypothetical, but most likely transmembrane.
HMPREF4062_0095	Hypothetical protien	Dna topsemirase

References

2005-2016 WebMD, LLC.
Geni-act.org

Acknowledgments

Supported by an NIH Science Education Partnership (SEPA) Award - 5R25OD010536
Thank you to Stephen T. Koury, Ph.D. University at Buffalo
Thank you to Rama Dey-Rao, Ph.D. University at Buffalo.
E.R. Tech department.
Jennifer R.Cooke,MPS,CSA (Program coordinator) WNY R-AHEC
www.buffalo.edu