Annotation of the *Helicobacter pylori* Genome from DNA Coordinates 78358 to 82441 (or Locus Tags HMPREF0462_0079 to HMPREF0462_0082)

Samantha Freiberger and Lawrence Hohl
East Rochester High School, East Rochester, New York East Rochester High School, East Rochester, New York

University at Buffalo

Abstract

I was given a segment of the *Helicobacter* pylori *83* genome to to determine whether or not the genes HMPREF0462_0079, HMPREF0462_0080, HMPREF0462_0081 and HMPREF0462_0082 were correctly annotated, I used the GENI-ACT modules to analyze these genes, and the data indicated that the proposed annotations of these genes were not all correct. The proposed annotations for HMPREF0462_0079 and HMPREF0462_0082 were correct, coding for a hypothetical protein with an unknown function. The proposed annotation for HMPREF0462_0080 may have been correct as a tyrosine recombinase XerH, although it may instead be an integrase/recombinase, which is involved in the integration of bacteria into infected cells. The proposed annotation of HMPREF0462_0081 may have been correct as a serine/threonine-protein kinase, although it may instead be a cell translocating kinase A N-terminus , functioning as a proinflammatory protein, producing and promoting inflammation.

Introduction

Helicobacter pylori is a helix shaped, gram-negative, microaerophilic bacterium that is usually found in the stomach and the small intestine. *H. pylori* is known for causing peptic ulcers in the stomach lining and the upper part of the small intestine. *H. pylori* was discovered in 1982 by Dr. Barry J. Marshall and Dr. J. Robin Warren as a cause of gastritis and peptic ulcer disease (Canadian Society of Intestinal Research, 2005).

One of the genes that I had annotated was an integrase/recombinase. An integrase allows the bacteria to be integrated into the infected cell. Another of the genes that I had annotated appeared to be a cell translocating kinase that appears to function as a proinflammatory protein, that produces and promotes inflammation. The other two genes that I had annotated were hypothetical proteins whose functions are unknown.

The current study that is being done on *H. pylori* is attempting to learn how different genes are working together within *H. pylori* to make it function the way it does.

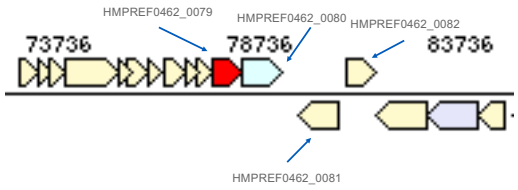

Figure 1. The locus tags and relative position of the genes under investigation in this research

Methods

Modules of the GENI-ACT (http://www.geni-act.org/) were used to complete Kytococcus sedentarius 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_0079:
The annotation showed that this gene is a cytoplasmic enzyme. The absence of transmembrane helices and signal peptides show that the enzyme is not located on the cell membrane nor outside the cell membrane. The function of this gene is unknown due to the lack of information on structure and lack of information and results overall. The proposed annotation of this gene by GENI-ACT was a hypothetical protein. Based on the results, it can be concluded that this gene is a hypothetical protein.

HMPREF0462_0080:
The annotation showed that this gene is a cytoplasmic enzyme as well, based on cellular localization data and that absence of transmembrane helices and signal peptides, showing that the gene is not located on the cell membrane nor outside the cell membrane. The function of this gene appears to be acting as an enzyme that helps in recombination. The proposed annotation of this gene by GENI-ACT was a tyrosine recombinase XerH, but based on the results, it can be concluded that this gene may instead be an integrase/recombinase.

HMPREF0462_0081:
The annotation showed that this gene is a cytoplasmic enzyme due to the absence of transmembrane helices and signal peptides. This cellular localization data shows that the gene is not located on the cell membrane nor outside the cell membrane. The function of this gene appears to be as a proinflammatroy protein. The proposed annotation of this gene by GENI-ACT was a serine/threonine-protein kinase, however, based on the results, this gene may be a cell translocating kinase A N-terminus .

HMPREF0462_0082:
The annotation of this gene showed that the location of this gene is unknown and that the protein may have multiple localization sites. The presence of one transmembrane helix shows that the gene could be located on the cell membrane, but the absence of signal peptides shows that the gene is not located outside the cell membrane. The function of this gene is unknown. The proposed annotation of this gene by GENI-ACT was a hypothetical protein. Based on the results, it can be concluded that this gene is a conserved hypothetical protein.

```
:QUENCE Length: 250
:QUENCE Number of predicted TMHs:  0
:QUENCE Exp number of AAs in TMHs: 0.00035
:QUENCE Exp number, first 60 AAs:  0
:QUENCE Total prob of N-in:        0.01712
JENCE     TMHMM2.0       outside     1   250
```

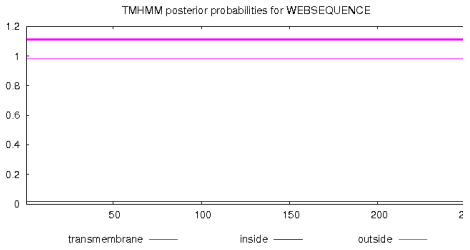

Figure 2. Shows lack of transmembrane helices present in HMPREF0462_0079, proving that it does not exist on the cell membrane.

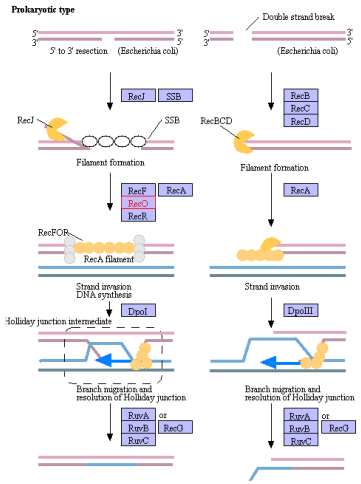

Figure 3. Shows the recombination/restructuring of DNA as the function of HMPREF0462_0080.

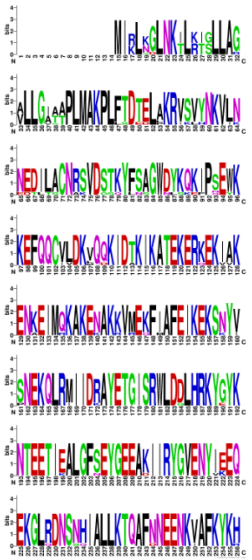

Figure 4. -This diagram shows a portion of a WebLogo of HMPREF0462_0082, showing the similarity between the gene sequence in Helicobacter pylori 83 and the same gene in other species. This segment shows that this gene is well conserved since many of the species have the same amino acids in the same places.

Conclusion

The GENI-ACT proposed gene annotations were similar to the proposed gene annotation for each of the genes in the group and as such, the genes may not be correctly annotated by the computer database.

References

1. GENI-ACT Training Manual, 2016
2. Canadian Society of Intestinal Research, 2005. Nobel Prize for *H. pylori* Discovery. http://www.badgut.org/information-centre/a-z-digestive-topics/nobel-prize-for-h-pylori-discovery/

Acknowledgements

Supported by an NIH Science Education Partnership (SEPA) Award R25OD010536-01A1
Stephen T. Koury, Ph.D. University at Buffalo
Rama Dey-Rao, Ph.D. University at Buffalo

www.buffalo.edu