Annotation of *Helicobacter pylori* at Locus Tags HP0082, HP1067, HP0392, and HP0815

Zackery Conrad, Lauren DeYoung, Nicholas Downs, Samantha Nusstein, and Laura O'Donnell

Attica High School- Attica, 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

University at Buffalo

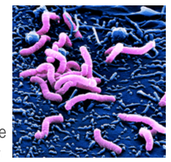
Abstract

A group of genes from the microorganism *Helicobacter pylori* were annotated using the collaborative genome annotation website GENI-ACT. The Genbank proposed gene product name for each gene was assessed in terms of the general genomic information, amino acid sequence-based similarity data, structure-based evidence from the amino acid sequence, and cellular localization data. 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 the computer database.

Introduction

Helicobacter pylori is gram-negative species of bacteria most closely related to the cause of ulcers in the stomach and small intestine, an increase in the risk of stomach cancer, and found in those with chronic gastritis at higher rates. It is transmitted through the oral-fecal route or oral-oral route. The bacteria lives best in an environment with a lesser oxygen concentration in comparison to the typical concentration of the air, also known as a microaerophilic bacteria. They use the oxygen to oxidize hydrogen in order to make energy, this ideal environment fits the upper digestive tract.

H. pylori contains two to seven flagella and averages 3.5 microns long and 0.5 microns wide. Its helix shape and flagella help to propel it through the tough mucus linings of the upper digestive tract or stomach and small intestine. In order to survive the environment of the stomach and small intestine they find their way into the mucus lining to get to the epithelial cells underneath which have a less acidic pH. They then bind with the epithelial cells through production of adhesins. This process exposes the lining behind the mucus to acids from the digestive tract. The acids and bacteria ultimately cause gastric erosions, forming ulcers and inflammation of the lining. Some antibiotics used to fight *H. pylori,* are now becoming resistant so new studies are researching blocking its adhesin production as an alternative treatment solution.



Colored scanning electron micrograph of *H pylori* on surface of gastric cells — *web.stanford.edu*

Methods

Modules of the GENI-ACT (http://www.geni-act.org/) were used to complete *Helicobacter pylori* 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- Structure-Based Evidence	TIGRfam, Pfam, PDB	Are there functional domains in my protein?
Module 4- Cellular Localization Data	Gram Stain, TMHMM, SignalP, PSORT, Phobius	Is my protein in the cytoplasm, secreted or embedded in the membrane?

Results

Gene Locus: HP0082		
Basic Information	DNA Coordinates	Complement (86656..88677)
	DNA Length	2022 nt
	Amino Acid Sequence Length	673 aa
Sequence-based Similarity	COGS	COG0840 Tar Methyl-accepting chemotaxis protein [Cell motility, Signal transduction mechanisms];
	T-Coffee	More conserved in the beginning. Less conserved towards the end.
	Web-Logo	Not well conserved
Structure-based Evidence	TIGRFAMs	TIGR01612: 235kDa-fam
	PFAMs	pfam00015 Methyl accepting chemotaxis
	HMM Logo Key Residues	R62 G64 G67 G69 F70
	PDB	Double CACHE (dCACHE) sensing domain of TlpC chemoreceptor
Cellular Localization	Transmembrane Helices	2
	Signal Peptide	none
	PSORTb Final Prediction	Unknown (may have multiple location sites)

Helicobacter pylori HP0082:

COG, TIGRFAM, PFAM, and PDB data supports HP0082 to be a methyl-accepting chemotaxis protein which senses changes in chemical concentrations in the environment in order to adjust its locomotion accordingly. Cellular localization data is conflicting but TMHMM predicts that this protein has two transmembrane helices.

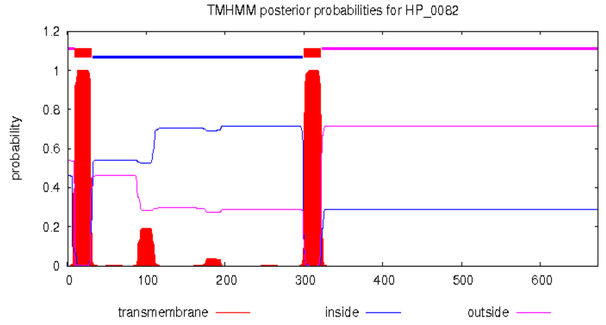


Figure 1 – TMHMM data for HP0082: The cellular localization data shows the presence of 2 transmembrane helices.

Gene Locus: HP1067		
Basic Information	DNA Coordinates	1126268..1126642
	DNA Length	375 nt
	Amino Acid Sequence Length	124 aa
Sequence-based Similarity	COGS	COG0784: CheY chemotaxis protein or a CheY-like REC (receiver) domain [Signal transduction mechanisms];
	T-Coffee	Well conserved throughout
	Web-Logo	Fairly well conserved
Structure-based Evidence	TIGRFAMs	TIGR0215: PhoB; phosphate regulon transcriptional regulatory protein
	PFAMs	pfam00072: Response regulator receiver domain
	HMM Logo Key Residues	D50, G58, K100
	PDB	CheY
Cellular Localization	Transmembrane Helices	None
	Signal Peptide	None
	PSORTb Final Prediction	Cytoplasmic

Helicobacter pylori HP1067:

COG, TIGRFAM, PFAM, and PDB data supports HP1067 to be a cytoplasmic signaling protein. It is used to transmit signals from chemoreceptors to flagellar motors, controlling the direction that the flagella rotate. Due to the presence of many orthologs, it may be hypothesized that this protein is commonly found in bacteria that possess a flagellum and respond to chemicals in their environment. Cellular localization data predicts that HP1067 is a cytoplasmic membrane protein.

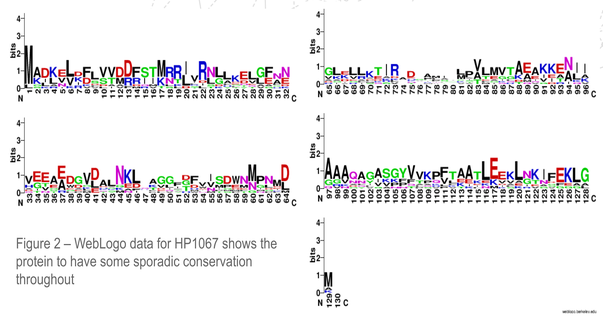


Figure 2 – WebLogo data for HP1067 shows the protein to have some sporadic conservation throughout

Gene Locus: HP0392		
Basic Information	DNA Coordinates	Complement (400546..402957)
	DNA Length	2412 nt
	Amino Acid Sequence Length	803 aa
Sequence-based Similarity	COGS	COG0784, CheY, CheY chemotaxis protein or a CheY-like REC (receiver) domain [Signal transduction mechanisms]
	T-Coffee	Well conserved in beginning, not well conserved in the end
	Web-Logo	Not well conserved
Structure-based Evidence	TIGRFAMs	TIGR02154: PhoB:phosphate regulon transcriptional regulatory protein
	PFAMs	pfam01584: CheW-like domain
	HMM Logo Key Residues	G46, V85, D6, D50, G58, K100
	PDB	Bacterial Chemotaxis Signaling
Cellular Localization	Transmembrane Helices	None
	Signal Peptide	None
	PSORTb Final Prediction	Cytoplasmic

Helicobacter pylori HP0392:

COG, TIGRFAM, PFAM, and PDB data suggests that HP00392 is protein involved with chemotaxis signaling which monitors the outside environment then sends signals to the flagellar motor to move. Cellular localization data predicts that HP0392 functions in the cytoplasm.

Gene Locus: HP0815		
Basic Information	DNA Coordinates	868381..869154
	DNA Length	774 nt
	Amino Acid Sequence Length	257 aa
Sequence-based Similarity	COGS	COG1291:MotA Flagellar motor component MotA [Cell motility]
	T-Coffee	Well conserved until the very end
	Web-Logo	Some sporadic areas of conservation
Structure-based Evidence	TIGRFAMs	TIGR03818 MotA
	PFAM	pfam01618, MotA_ExbB
	HMM Logo Key Residues	G60, G63, T64, T97, G100
	PDB	no results found
Cellular Localization	Transmembrane Helices	3
	Signal Peptide	none
	PSORTb Final Prediction	Cytoplasmic membrane

Helicobacter pylori HP0815:

COG, TIGRFAM, and PFAM data suggests that HP0815 is flagellar motor protein MotA, that assists with flagella movement and changes in direction, working alongside another flagellar motor protein, MotB. Cellular localization data predicts that HP0815 has 3 transmembrane helices, residing in the cytoplasmic side of the cell membrane.

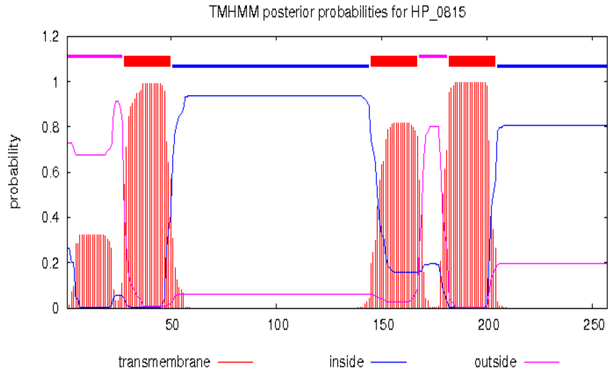


Figure 4 – TMHMM data for *Helicobacter pylori* HP0815. The cellular localization data shows the presence of 3 transmembrane helices

Conclusion

The GENI-ACT proposed gene product 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 the computer database. All proteins annotated seem to be involved with chemotaxis and flagellar movement in *H. pylori*.

Gene Locus	Geni-Act Gene Products	Proposed Annotation
HP0082	Methyl-accepting chemotaxis protein	Methyl-accepting chemotaxis protein
HP1067	CheY chemotaxis signaling protein	CheY chemotaxis signaling protein
HP0392	CheY chemotaxis signaling protein	CheY chemotaxis signaling protein
HP0815	MotA: Flagellar motor component protein	MotA: Flagellar motor component protein

References

Peptic Ulcers, Gastritis & Helicobacter Pylori. (n.d.). Retrieved from https://www.rileychildrens.org/health-info/peptic-ulcers-gastritis-helicobacter-pylori

New receptors discovered for stomach germ Helicobacter pylori. (2016, October 18). Retrieved from https://www.sciencedaily.com/releases/2016/10/161018094227.htm

Acknowledgments

Supported by an NIH Science Education Partnership (SEPA) Award - R25ODO10536

www.buffalo.edu