Annotation of Helicobacter pylori at Locus Tags HP0001, HP1585, HP0019, HP0243, HP0597, and HP0110

Regen Orlowski, Nora Parkhurst, Emily Tait, Kayla Warwick, Autumn Wildey, Abby Zymowski, and Laura O'Donnell Attica High School-Attica, NY and The Western NY Genetics in Research and Health Care Partnership

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 a gram-negative bacteria usually found in the stomach. It can usually be found in people dealing with chronic gastritis problems and/or gastric ulcers that can lead to forms of stomach cancers. Around fifty percent of the world's population has Helicobacter pylori in them. Every year, there are approximately 500 to 850 thousand cases of reported ulcers due to *H. pylori*.

The shape of the organism is known to help penetrate the walls of the stomach. The bacteria is mobile, having a total of six flagella all in the same place to generate power into the stomach lining. The organism does need oxygen but only at a low concentration, otherwise known as microaerobic. H. pylori can sense the pH in the mucus and can then move towards more non-acidic regions to avoid being destroyed by the acidic conditions. When Helicobacter pylori cause an ulcer, they dig through the thick mucus lining of the stomach that protects it from the acid it holds. This hole in the lining results in the stomach tissue being exposed to the acidity within it at a pH ranging from 1.5 to 3.5, causing what we call an ulcer. *H. pylori* can usually be treated with antibiotics.

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:

Questions Investigated
What is the sequence of my gene and protein? Where is it located in the genome?
Is my sequence similar to other sequences in Genbank?
Are there functional domains in my protein?
Is my protein in the cytoplasm, secreted or embedded in the membrane?



Results

Gene Locus: HP0001			
	DNA Coordinates	complement(217633)	
Basic Information	DNA Length	417 bp	
	Amino Acid Sequence Length	138 aa	
Saguanaa bacad	COGS	COG0781: NusB-Transcription termination factor NusB [Transcription];	
Sequence-Daseu Similarity	T-Coffee	Well conserved throughout	
Similarity	Web-Logo	Very well conserved until about amino acid 131, not well conserved at the end	
	TIGRFAMs	nusB: DNA Replicator	
Structure-based	PFAMs	pfam01029: NusB	
Evidence	HMM Logo Key Residues	N 126, L 129,	
	PDB	Solution structure of NusB from Aquifex aeolicus	
	Transmembrane Helices	none	
Cellular Localization	Signal Peptide	none	
	PSORTb Final Prediction	Unknown	

Helicobacter pylori HP0001:

Data from the COG, PFAM, TIGRFAM, and PDB databases support the prediction that NusB is a transcription anti-termination factor. Cellular localization data was inconclusive because the cytoplasmic, cell wall, periplasmic, outer membrane, and extracellular scores all shared the same value of 2.00. Due to the numerous orthologs, it can be inferred that many species of bacteria contain a similar protein with a similar function.

Figure 1 – WebLogo data for *Helicobacter pylori* HP0001 showing the protein is mostly well conserved with less conservation at the N terminus.

Gene Locus: HP1585				
	DNA Coordinates 16635901664378			
Basic Information	DNA Length	789 bp		
	Amino Acid Sequence Length	262		
Sequence-based	COGS	COG4786-FlgG- Flagellar basal body rod protein FlgG [Cell motility]		
Similarity	T-Coffee	Well conserved in the beginning and end		
	Web-Logo	Well conserved in the beginning		
	TIGRFAMs	FIgEFG_subfam and flgG_G_neg		
Structure bood	PFAMs	Flg_bbr_C		
Structure-Daseu	HMM Logo Key Residues	Y 54, V 38,		
Evidence	PDB	FlgG structure based on the CryoEM map of the bacterial flagellar polyrod		
	Transmembrane Helices	none		
Cellular Localization	Signal Peptide	none		
	PSORTb Final Prediction	Extracellular		

Helicobacter pylori HP1585:

COG, TIGRFAM, PFAM, and PDB data predict that HP1585 is a flagellar basal body rod protein involved with flagellar movement in the bacteria. Due to presence of many orthologs, the data supports the conclusion that this protein is common in many different species of bacteria that contain flagella. Cellular localization data predicts that this protein is extracellular.

TIOI	IE
Extrace	ellular
	4 2 2 2 2 2 2 2 2 2 2 2 2 2
	4- 3- 3- 2- 1- 0 0 88 8 3 80 5 3 80 5 80 5 80 5 80 5 80 5 80 5 80 5 80 5
	0 N N N N N N N N N N N N N N N N N N N
3- 2- 1-	

Figure 2 – WebLogo data for Helicobacter pylori HP1585 showing the protein is well conserved at the beginning from location 1-127, the middle is not well conserved with sporadic conserved spots, and then more conservation towards the end.

Gene Locus: HP0019			
	DNA Coordinates	1838019345	
Basic Information	DNA Length	966 bp	
	Amino Acid Sequence Length	321 aa	
	COGS	COG0835-CheW-Chemotaxis signal transduction protein	
Soguonco-basod	0005	[Cell motility, Signal transduction mechanisms]	
Sequence-based Similarity	T-Coffee	Not well conserved	
Sinnanty	Web-Logo	Poorly conserved from 1 to 717. Well conserved from 718	
		to 854. Sporadically conserved at the end.	
	TIGRFAMs	TIGR01818- ntrC nitrogen regulation protein NR(I)	
	PFAMs	PF01584-CheW-like domain; CheW proteins are part of	
Structure-based		the chemotaxis signaling mechanism	
Evidence	UMM Logo Koy Posiduos	L-3 F-5 Y-12 V-19 I-22 R-45 G-46 L-54 V-73 G-82 V-	
	HIVIWI LOGO KEY RESIDUES	85 G-1115 D-129	
	PDB	4JA2: Entity 1 containing Chain A	
	Transmembrane Helices	none	
Cellular Localization	Signal Peptide	none	
	PSORTb Final Prediction	Cytoplasmic	

lelicobacter pylori HP0243: COG, PFAM, and PDB data predict that HP0243 is an ironstoring protein found in the cytoplasm that protects DNA rom oxidative stress and starvation. Cellular localization lata predicts that this protein functions in the cytoplasm.

Figure 4 – WebLogo data for *Helicobacter pylori* HP0243 showing the protein is poorly conserved at the beginning, then well conserved at locations 39-112, and sporadically conserved towards the end.



transduction protein. This protein allows the bacteria to respond to chemicals in their environment by either moving towards or away from them. Cellular localization data predicts this protein to reside in the cytoplasm.

4	**		1
	* 3 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5		
**************************************			4
**************************************	* # # # #	g.	*
		1 000000000000000000000000000000000000	1
5- - 	* 1 **		, ,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,
			9-
・ - 5 - 5 - 1 - 1 - 1 - 1 - 1 - 1 - 1 - 1			A-

Figure 3 – WebLogo data for Helicobacter pylori HP0019 showing that the protein is poorly conserved from to 717, well conserved from 718 to 854 and contains areas with minimal conservation towards the end.

Gene Locus: HP0243			
ic Information	DNA Coordinates	complement (252272252706)	
	DNA Length	435 bp	
	Amino Acid Sequence Length	144 aa	
quence-based Similarity	COGS	Dps-DNA-binding ferritin-like protein (oxidative damage protectant)	
	T-Coffee	Poorly conserved in beginning, well conserved in middle, and sporadically conserved at the end	
	Web-Logo	Not conserved from 1-38, well conserved in the middle from 39-112, and sporadically conserved at the end from 119-182	
	TIGRFAMs	no TIGRFAM hit	
ucture-based	PFAMs	ferritin-like superfamily	
Evidence	HMM Logo Key Residues	L4, Y18, G59, D111	
	PDB	NAP protein from Helicobacter pylori	
lar Localization	Transmembrane Helices	none	
	Signal Peptide	none	
	PSORTb Final Prediction	cytoplasmic	

Gene Locus: HP0597			
Information	DNA Coordinates	complement (630840632819)	
	DNA Length	1980 bp	
	Amino Acid Sequence Length	659 aa	
	COGS	no significant hits	
anco-basod	T-Coffee	Not well conserved in the beginning but more conserved	
milarity	I-COILEE	throughout the middle and end	
iiiiaiity	Web-Logo	Not well conserved in the beginning but sporadically conserved	
	Web-Logo	from 225-383. Many hydrophobic and polar amino acids.	
	TIGRFAMs	no significant hit	
turo-basod	PFAMs	pfam00912-Transglycosylase	
vidence	HMM Logo Key Residues	G65, Q72, E34, Y111	
	PDB	3VMA: Entity 1 containing Chain A-Crystal Structure of the Full-	
		Length Transglycosylase PBP1b from Escherichia coli	
⁻ Localization	Transmembrane Helices	1	
	Signal Peptide	no signal peptide	
	PSORTb Final Prediction	Cytoplasmic membrane	

Helicobacter pylori HP0597

PFAM, PDB and BLAST data supports HP0597 to be a transglycosylase penicillin binding protein which is an enzyme that aids in synthesis of the bacterial cell wall. Due to the presence of many orthologs, it can be inferred that this a common protein in many different species of bacteria. Cellular localization data predicts that HP0597 resides in the cytoplasmic membrane

with 1 transmembrane helix. TMHMM posterior probabilities for HP0593

Figure 5 – WebLogo and 🗿 TMHMM data for HP0597 The cellular localization data on the left shows the oresence transmembrane helix. The Weblogo image on the some conservation halfway through the protein with conservation at the C and N terminals.

		TAYA A THE AR	* #F.G.TPx:
3 5 ***********************************	. As FALLARE ALLARPEARS	S. B. R. Y. ALG	
K.L.F.L.A.L.Q.	₩ ₽		
P			
	8		
2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2		S. SP. MARSAN	
S LYO PG LIA		Revel Police	* * * * * * * * * * * * * * * * * * *
		St. St. Ar. John	* * * * * * * * * * * * * * * * * * *
		SIG A	e N
LJ L KL	Part v. the		
			A NEW CONTRACTOR CONTE

Basic Infor

Sequence-ba Similarity

Structure-ba Evidence

Cellular Locali

COG, I chapero by help localizat cytoplasr

Figure 6 – WebLogo data for Helicobacter pylori HP0110 minimal conservation of the protein amongst different species of bacteria.

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.

Gene Locus	Geni-Act Gene Products	Proposed Annotation
HP0001	Transcription anti-termination protein NusB	Transcription anti-termination protein NusB
HP1585	Flagellar basal body rod protein FlgG	Flagellar basal body rod protein FlgG
HP0019	Chemotaxis signal transduction protein	Chemotaxis signal transduction protein
HP0243	Dps-DNA-binding ferritin-like protein	Dps-DNA-binding ferritin-like protein
HP0597	Transglycosylase penicillin binding protein	Transglycosylase penicillin binding protein
HP0110	GrpE-Molecular chaperone protein	GrpE-Molecular chaperone protein

References

Gastritis and Peptic Ulcer Disease Caused by Helicobacter pylori. (n.d.). Retrieved from https://microbewiki.kenyon.edu/index.php/Gastritis_and_Peptic_Ulcer_Dise ase_Caused_by_Helicobacter_pylori

Acknowledgments Supported by an NIH Science Education Partnership (SEPA) Award - R250D010536

	DNA Coordinates	complement(118254118823)
ation	DNA Length	570 bp
	Amino Acid Sequence Length	189 aa
	COCS	GrpE-Molecular chaperone GrpE (heat shock protein)
sed	COGS	[Posttranslational modification, protein].
,	T-Coffee	Not well conserved throughout
	Web-Logo	Not well conserved throughout
	TIGRFAMs	no TIGRFAM hit
	PFAMs	PF01025-GrpE
and	HMM Logo Key Residues	D73, G94, F121, D122, G149, Y150, R158, V162, V164
sea		4ANI:Entity 1 containing Chain A, B, E, F
	PDB	Structural basis for the intermolecular communication between
		DnaK and GrpE in the DnaK chaperone system from Geobacillus
		kaustophilus HTA426
	Transmembrane Helices	noné
zation	Signal Peptide	none
1	PSORTb Final Prediction	Cytoplasmic

Helicobacter pylori HP0110:

PFAM, and PDB data support HF	
ne protein that assists DnaK with pro	otein folding
ing to release ADP from DnaK	
ion data predicts that this protein fun	
m.	N 886888558862552222222288588888558858858655858586 4- 3-

Conclusion

www.buffalo.edu