Annotation of *Helicobacter pylori* at Locus Tags HP0615, HP1022, HP1245, HP0012, and HP1362

Lily Baratta, Emily Gadd, Chase Harding, Morgan Hofheins, Elizabeth Peters, 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 a species of gram-negative bacteria that causes peptic ulcers and gastric cancer. The bacterium obtained its name from the root word "helico" meaning spiral. The spiral shape of the bacterium as well as its many flagella allows them to penetrate the stomach's lining where the immune cells are not able to reach them due to the mucus coating. When *H. pylori* enters the body it will grow in the mucus layer of the stomach lining because the bacteria are well adapted to survive in the harsh acidic environment of the stomach.

Many people obtain *Helicobacter pylori* by drinking unsanitary water, eating improperly prepared food, or coming in direct mouth-to-mouth contact with a person who has *H. pylori.* Children living in crowded or unsanitary conditions are most likely to be infected with the bacterium. *H. pylori* infection rates are much higher in developing nations rather than developed countries. Approximately two-thirds of the world's population is infected with H. pylori even though not all infected people will experience illness. The bacterium will live for many years before symptoms will start to show.

H. pylori cannot be fully prevented because there is no specific vaccine to against the bacteria. Drinking safe water, eating proper foods, and maintaining good hygiene can help prevent the spread of *H. pylori*.

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: HP0615		
Basic Information	DNA Coordinates	complement (659069..661039)
	DNA Length	1971 nt
	Amino Acid Sequence Length	656 aa
Sequence-based Similarity	COGS	no significant hits
	T-Coffee	Well conserved until the end
	Web-Logo	M1, L456
Structure-based Evidence	TIGRFAMs	No significant hits
	PFAMs	pfam01653: NAD-dependent DNA ligase adenylation domain
	HMM Logo Key Residues	P27, P49, P57, H74, G117, Y124, G136, G138, G141, P155, G171, F180, G204, D285, G286, G301, P307
	PDB	DNA Ligase
Cellular Localization	Transmembrane Helices	none
	Signal Peptide	none
	PSORTb Final Prediction	cytoplasmic

Helicobacter pylori HP0615:

PFAM and PDB data supports HP0615 to be a ligase protein. It joins strands of DNA through the catalyzing of the formation of phosphodiester bonds common in single strand breaks in duplex DNA during DNA replication. Cellular localization data predicts that this protein functions in the cytoplasm.

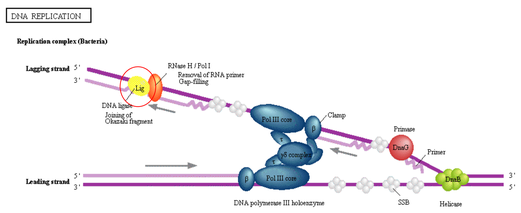


Figure 1 shows DNA ligase (circled in red) during the process of DNA replication
www.genome.jp/kegg/

Gene Locus: HP1022		
Basic Information	DNA Coordinates	1085284..1086120
	DNA Length	837 nt
	Amino Acid Sequence Length	278 aa
Sequence-based Similarity	COGS	COG0258: 5'-3' exonuclease [Replication, recombination and repair]
	T-Coffee	Well conserved at the beginning; not at the end
	Web-Logo	Well conserved at the beginning; not at the end. Few orthologs
Structure-based Evidence	TIGRFAMs	TIGR00593: pola; DNA polymerase I
	PFAMs	pfam02739: 5'-3' exonuclease, N-terminal resolvase-like domain
	HMM Logo Key Residues	D6, R61, Y68, R72, D107
	PDB	TAQ DNA POLYMERASE
Cellular Localization	Transmembrane Helices	none
	Signal Peptide	none
	PSORTb Final Prediction	Cytoplasmic

Helicobacter pylori HP1022:

COG TIGRFAM, PFAM, and PDB data supports HP1022 to be a protein involved with the exonuclease activity of DNA polymerase, replicating, recombining, and repairing DNA. Cellular localization data predicts that this protein is found in the cytoplasm. Few orthologs were found.

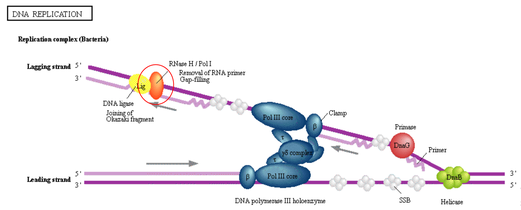


Figure 2 shows the exonuclease (circled in red) during the process of DNA replication
www.genome.jp/kegg/

Gene Locus: HP1245		
Basic Information	DNA Coordinates	Complement (1320618..1321157)
	DNA Length	540 nt
	Amino Acid Sequence Length	179 aa
Sequence-based Similarity	COGS	COG0628 Ssb- Single-stranded DNA-binding protein [Replication, recombination, and repair]
	T-Coffee	well conserved at the beginning, not well conserved at the end
	Web-Logo	most conserved at the beginning, not well conserved at the end
Structure-based Evidence	TIGRFAMs	TIGR00621: single-stranded DNA binding proteins
	PFAMs	pfam00436: SSB: single-stranded binding protein
	HMM Logo Key Residues	G8, G69 ,V74, G76
	PDB	Single stranded DNA binding protein complex
Cellular Localization	Transmembrane Helices	none
	Signal Peptide	none
	PSORTb Final Prediction	cytoplasmic

Helicobacter pylori HP1245:

COG, TIGRFAM, PFAM, and PDB data predict HP1245 to be a DNA binding protein. This protein replicates, recombines, and repairs DNA in the cell. Cellular localization data predicts that this protein functions in the cytoplasm. Due to the presence of many orthologs, it can be inferred that HP1245 is a common protein found in many other species of bacteria.

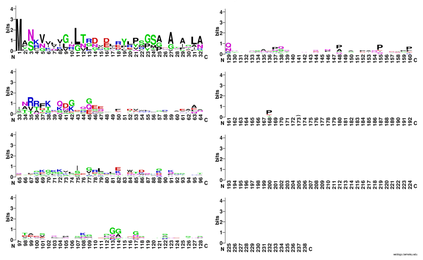


Figure 3 – WebLogo data for HP1245 shows some conservation of the protein in the beginning, minimal conservation in the middle and no conservation at the end.

Gene Locus: HP0012		
Basic Information	DNA Coordinates	9911..11590
	DNA Length	1680 nt
	Amino Acid Sequence Length	559 aa
Sequence-based Similarity	COGS	COG0358;DnaG- DNA primase [Replication, recombination and repair]
	T-Coffee	Not as well conserved at the end; well conserved at the beginning
	Web-Logo	Not as well conserved towards the end; well conserved at the beginning
Structure-based Evidence	TIGRFAMs	TIGR01391: dnaG; DNA primase
	PFAMs	pfam01807: zf-CHC2: DNA binding in DNA primases
	HMM Logo Key Residues	C36, P37, C57, C60,
	PDB	DnaG Primase C terminal domain
Cellular Localization	Transmembrane Helices	none
	Signal Peptide	none
	PSORTb Final Prediction	cytoplasmic

Helicobacter pylori HP0012:

COG, TIGRFAM, PFAM, and PDB data supports HP0012 to be a DNA primase and aids in the replicating of DNA. 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 HP0012 functions in the cytoplasm.

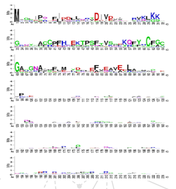


Figure 4 (left) – WebLogo data for HP0012 shows some conservation of the protein from position 1 to 89. There is no conservation from position 90 to 664.

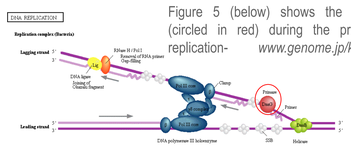


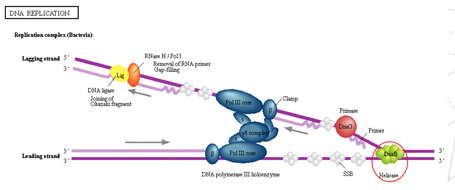
Figure 5 (below) shows the DnaG primase (circled in red) during the process of DNA replication- *www.genome.jp/kegg/*

Gene Locus: HP1362		
Basic Information	DNA Coordinates	complement (1422915..1424381)
	DNA Length	1467 nt
	Amino Acid Sequence Length	488 aa
Sequence-based Similarity	COGS	COG0305: Replicative DNA helicase [Replication, recombination and repair]
	T-Coffee	Well conserved until the end
	Web-Logo	Little conservation throughout
Structure-based Evidence	TIGRFAMs	TIGR00665: helicase DnaB
	PFAMs	pfam03796: DnaB-like helicase C terminal domain
	HMM Logo Key Residues	G1, G31, D136, P190, K233, G237
	PDB	DNA helicase
Cellular Localization	Transmembrane Helices	none
	Signal Peptide	none
	PSORTb Final Prediction	cytoplasmic

Helicobacter pylori HP1362:

COG, TIGRFAM, PFAM, and PDB data support HP1362 to be a helicase protein necessary for the unwinding and separating of DNA before replication. Cellular localization data predicts that this protein functions in the cytoplasm.

Figure 6 shows the DNA helicase enzyme (circled in red) during the process of DNA replication
www.genome.jp/kegg/



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 DNA replication, recombination, and repair.

Gene Locus	Geni-Act Gene Products	Proposed Annotation
HP0615	DNA ligase	DNA ligase
HP1022	DNA polymerase I (exonuclease)	DNA polymerase I (exonuclease)
HP1245	Single-stranded DNA binding protein	Single-stranded DNA binding protein
HP0012	dnaG: DNA primase	dnaG: DNA primase
HP1362	DNA helicase	DNA helicase

References

Helicobacter pylori and Cancer. (n.d.). Retrieved from https://www.cancer.gov/about-cancer/causes-prevention/risk/infectious-agents/h-pylori-fact-sheet

Helicobacter pylori (H. pylori) infection. (2017, May 17). Retrieved from https://www.mayoclinic.org/diseases-conditions/h-pylori/symptoms-causes/syc-20356171

Acknowledgments

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

www.buffalo.edu