

Annotation of *Helicobacter pylori* at Locus Tags HP1253, HP1126, HP1329, and HP1478

Reagan Bifarella, Megan Karalus, Breanna Mest, Michael Peresan, and Laura O’Donnell
Attica High School- Attica, NY and The Western NY Genetics in Research and Health Care Partnership



SEPA

SCIENCE EDUCATION PARTNERSHIP AWARD

Supported by the National Institutes of Health

UB

University at Buffalo

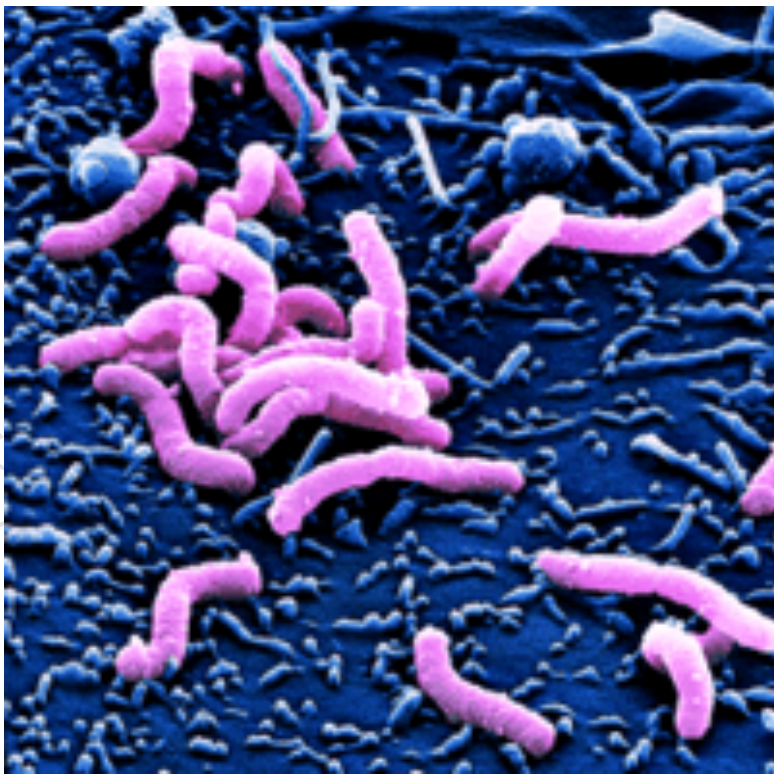
Abstract

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 common type of bacteria that grows in the digestive tract and has a tendency to attack the stomach lining. It is the main cause of peptide ulcers, and can cause gastritis and stomach cancer. This bacteria may be present in more than half of the world’s population. Helico in the word helicobacter means spiral which means that the bacteria are spiral shaped. It is this shape that allows *H. pylori* use their flagellum to burrow into the stomach lining in an attempt to reach the epithelial cells underneath, where it is less acidic and unable to be reached by immune cells.

H. pylori infections are thought can be spread through fecal matter, and contaminated food and water. Symptoms of an *H. pylori* infection are excessive burping, feeling bloated, heartburn, fever, nausea, and unexplained weight loss. Medications are available to help. For most people however, *Helicobacter pylori* causes almost no symptoms.



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: HP1253		
Basic Information	DNA Coordinates	1328909...1329928
	DNA Length	1020 bp
	Amino Acid Sequence Length	339aa
Sequence-based Similarity	COGS	Tryptophanyl-tRNA synthetase [Translation, ribosomal structure and biogenesis]; Sporadically conserved
	T-Coffee	Sporadically conserved
	Web-Logo	Sporadically conserved
	TIGRFAMs	trpS: tryptophan--tRNA ligase
Structure-based Evidence	PFAMs	tRNA-synt_1b: tRNA ligase class I (W and Y)
	HMM Logo Key Residues	G11, P14, T15, H19, G21, Y138, D145, G157, D159, Q160, K199, I209
	PDB	Crystal structure of tryptophanyl-tRNA synthetase from <i>Yersinia pestis</i> CO92
Cellular Localization	Transmembrane Helices	none
	Signal Peptide	none
	PSORTb Final PRediction	cytoplasmic

Helicobacter pylori HP1253:

COG, TIGRFAM, PFAM, and PDB databases support HP1253 to be a tryptophan-tRNA ligase which is an enzyme that catalyzes chemical reactions. It belongs to a family of ligases, specifically those that form carbon-oxygen bonds. Cellular localization data supports the prediction that this protein is most likely found in the cytoplasm because its job is to join larger molecules by forming a chemical bond during translation in protein synthesis. Due to the presence of many orthologs, it can be inferred that this is a common protein found in many different species of bacteria.

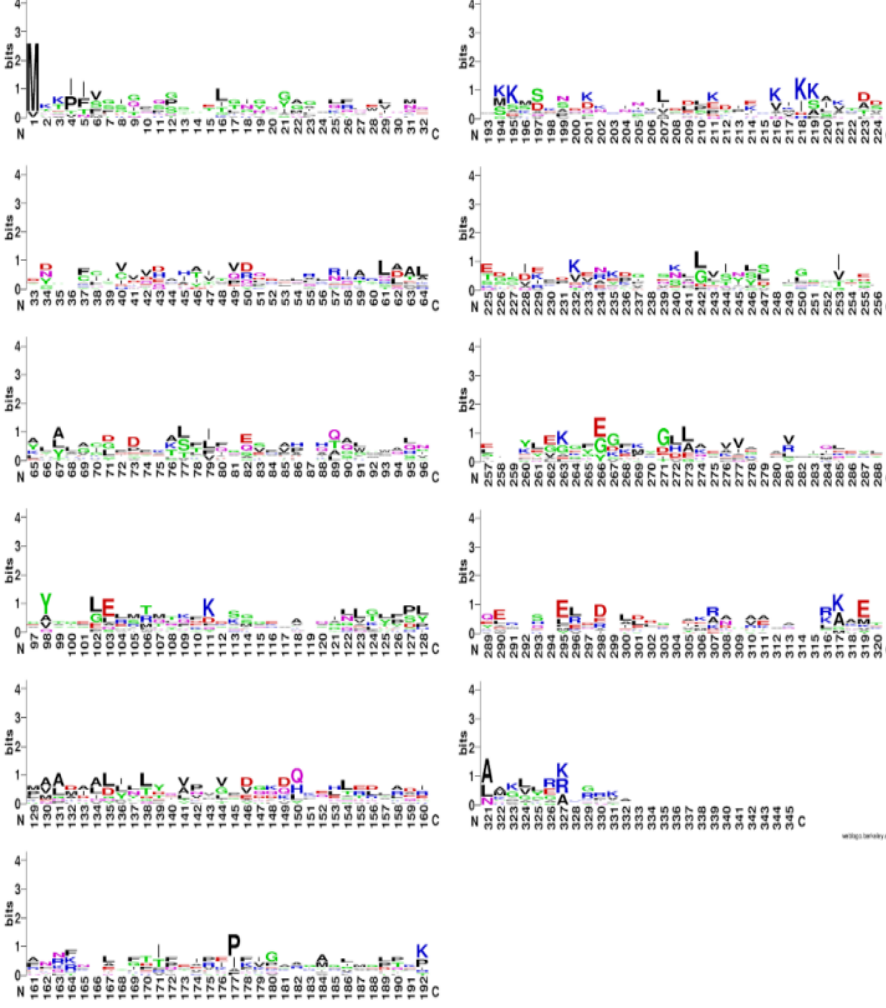


Figure 1 – WebLogo data for *Helicobacter pylori* HP1253 showing the protein is sporadically conserved throughout with no conservation at the end.

Gene Locus: HP1126		
Basic Information	DNA Coordinates	complement(1190763..1192016)
	DNA Length	1254 bp
	Amino Acid Sequence Length	417 aa
Sequence-based Similarity	COGS	COG0823
	T-Coffee	Periplasmic component of the Tol biopolymer transport system [Intracellular trafficking]
	Web-Logo	Sporadically conserved in the beginning but well conserved towards the middle and the end
	TIGRFAMs	Tol-Pal system beta propeller repeat protein TolB
Structure-based Evidence	PFAMs	TolB_N: TolB amino-terminal domain
	HMM Logo Key Residues	W70, I5, D40, F47
	PDB	4JML: Entity 1 containing Chain A
Cellular Localization	Transmembrane Helices	0
	Signal Peptide	0
	PSORTb Final Prediction	periplasmic

Helicobacter pylori HP1126:

COG, TIGRFAM, PFAM, and PDB data supports HP1126 to be a Tol-Pal system beta propeller repeat protein. Cellular localization data predicts this protein to be found within the periplasm of the bacteria. This corresponds to the protein's function of helping to maintain the outer membrane. Due to the presence of many orthologs, it can be inferred that this is a common protein in many different species of bacteria.

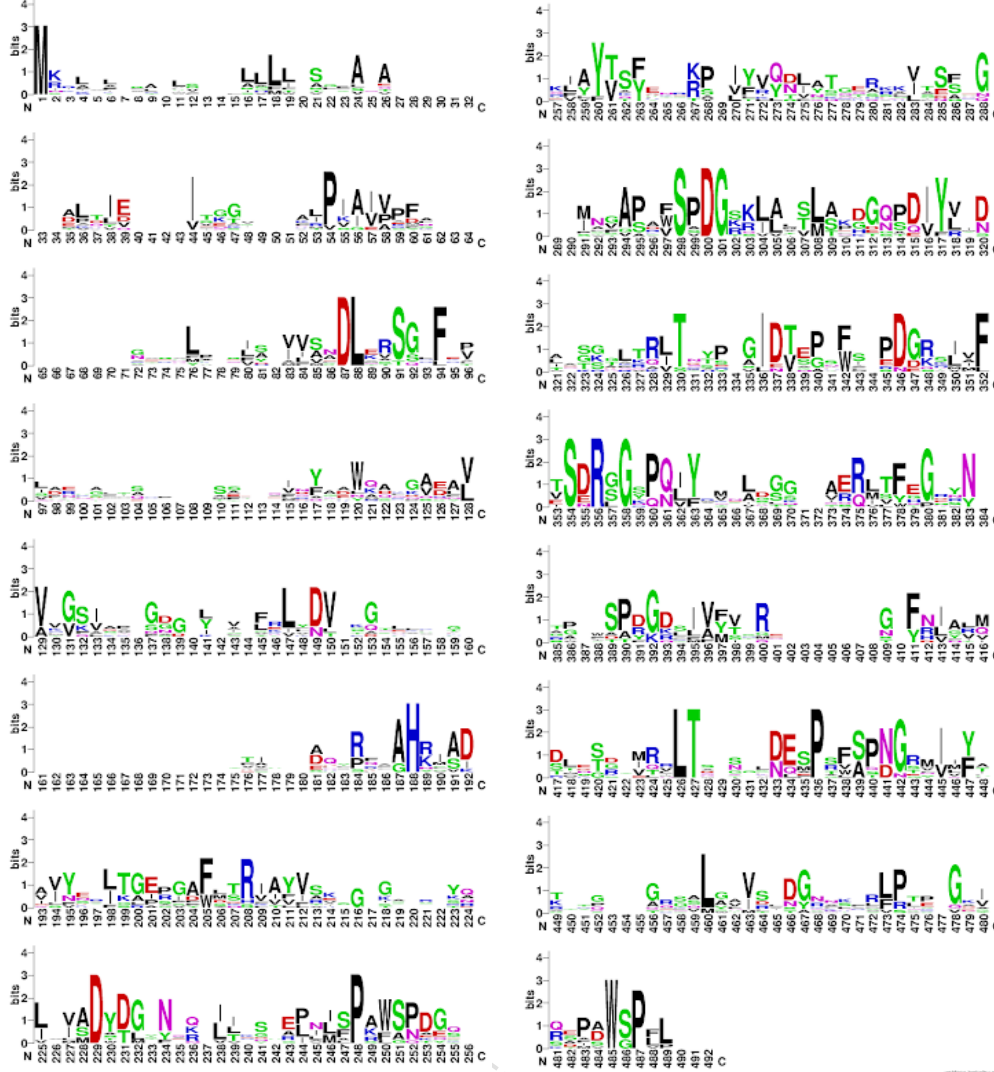


Figure 2 – WebLogo data for *Helicobacter pylori* HP1126 showing the protein is sporadically conserved in the beginning, but well conserved towards the middle and the end.

Gene Locus: HP1329		
Basic Information	DNA Coordinates	1388417..1391524
	DNA Length	3108 bp
	Amino Acid Sequence Length	1035aa
Sequence-based Similarity	COGS	no significant COG hit
	T-Coffee	well conserved throughout
	Web-Logo	well conserved from amino acid 1 to 1063 , not well conserved from 1064 to 2112
	TIGRFAMs	2A0601: heavy metal efflux pump, CzcA family
Structure-based Evidence	PFAMs	ACR_tran: RND transporter, Hydrophobe/Amphi
	HMM Logo Key Residues	P36, P336, P448, P463, Y516, P867, P979, P1014
	PDB	no significant hit
Cellular Localization	Transmembrane Helices	12
	Signal Peptide	none
	PSORTb Final Prediction	cytoplasmic membrane

Helicobacter pylori HP1329:

BLAST, TIGRFAM, and PFAM data support that HP1329 to be a heavy metal efflux pump. Cellular localization data predicts that this pump resides in the membrane with 12 transmembrane helices. This location prediction corresponds with the protein's function of pumping out harmful toxins and heavy metals from the cell.

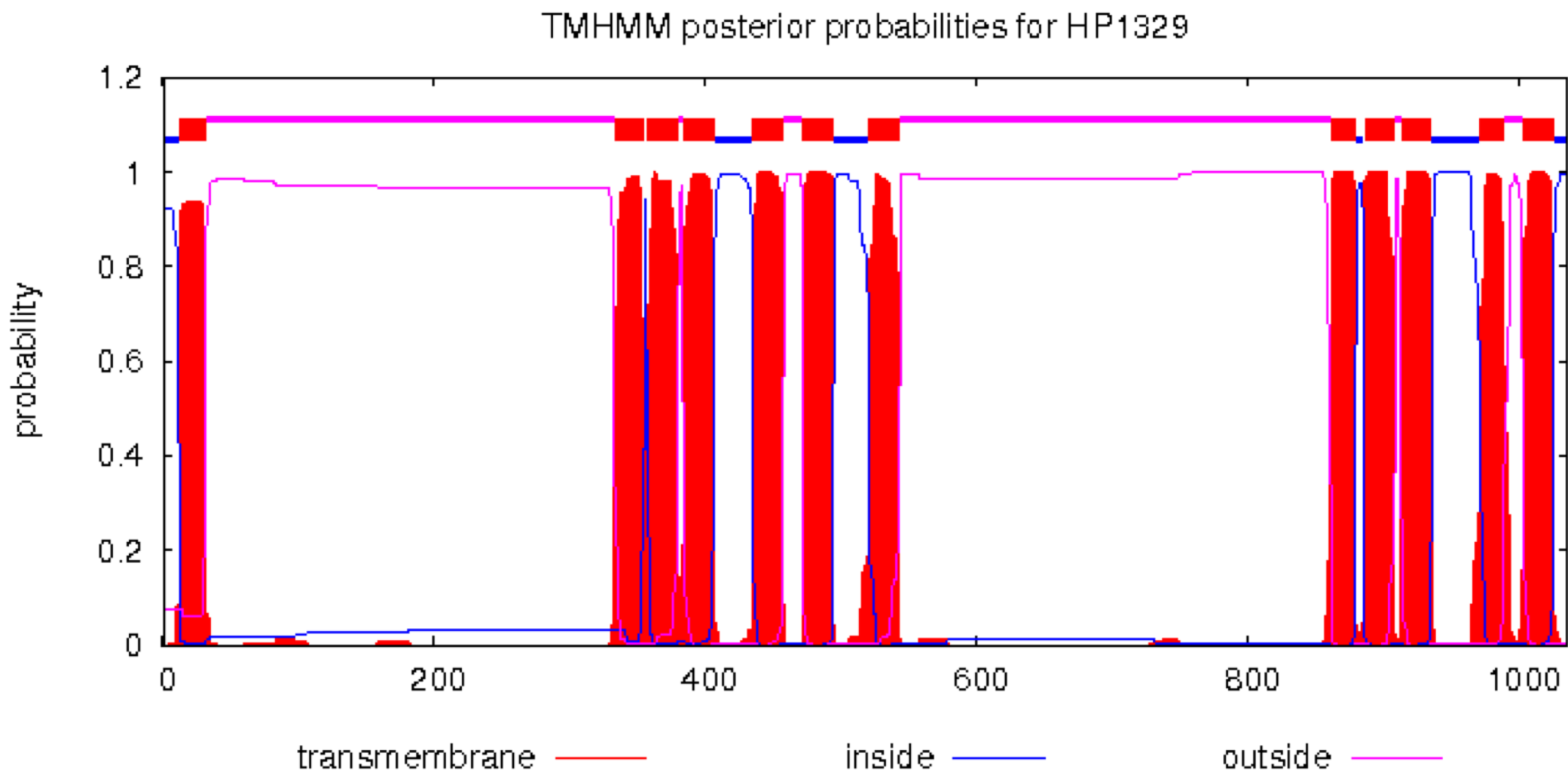


Figure 3 – WebLogo data and Cellular location data from the TMHMM database for *Helicobacter pylori* HP1329. The WebLogo at the top shows that the protein is well conserved from amino acid 1 to 1063 , but then not well conserved from 1064 to the end. The TMHMM graph above shows the presence of 12 transmembrane helices.

Gene Locus: HP1478		
Basic Information	DNA Coordinates	complement(1549611..1551659)
	DNA Length	2049 bp
	Amino Acid Sequence Length	682 aa
Sequence-based Similarity	COGS	No significant COG hit
	T-Coffee	Well conserved throughout
	Web-Logo	Well conserved until 705-768
	TIGRFAMs	pcrA: ATP-dependent DNA helicase PcrA
Structure-based Evidence	PFAMs	UvrD-helicase: UvrD/REP helicase
	HMM Logo Key Residues	Q266, Q294, D291, Y297
	PDB	1PJR: Entity 1 containing Chain A- Structure of DNA helicase
Cellular Localization	Transmembrane Helices	none
	Signal Peptide	none
	PSORTb Final Prediction	cytoplasmic

Helicobacter pylori HP1478:

TIGRFAM, PFAM, and PDB data supports HP1478 to be an ATP-dependent DNA helicase PcrA protein. The protein is an enzyme responsible for the separation of bases in DNA when the bacterial cell is undergoing replication. Cellular localization data predicts this protein to function in the cytoplasm where DNA replication would occur. Due to the presence of many orthologs, it can be inferred that his is a common protein in many different species of bacteria.

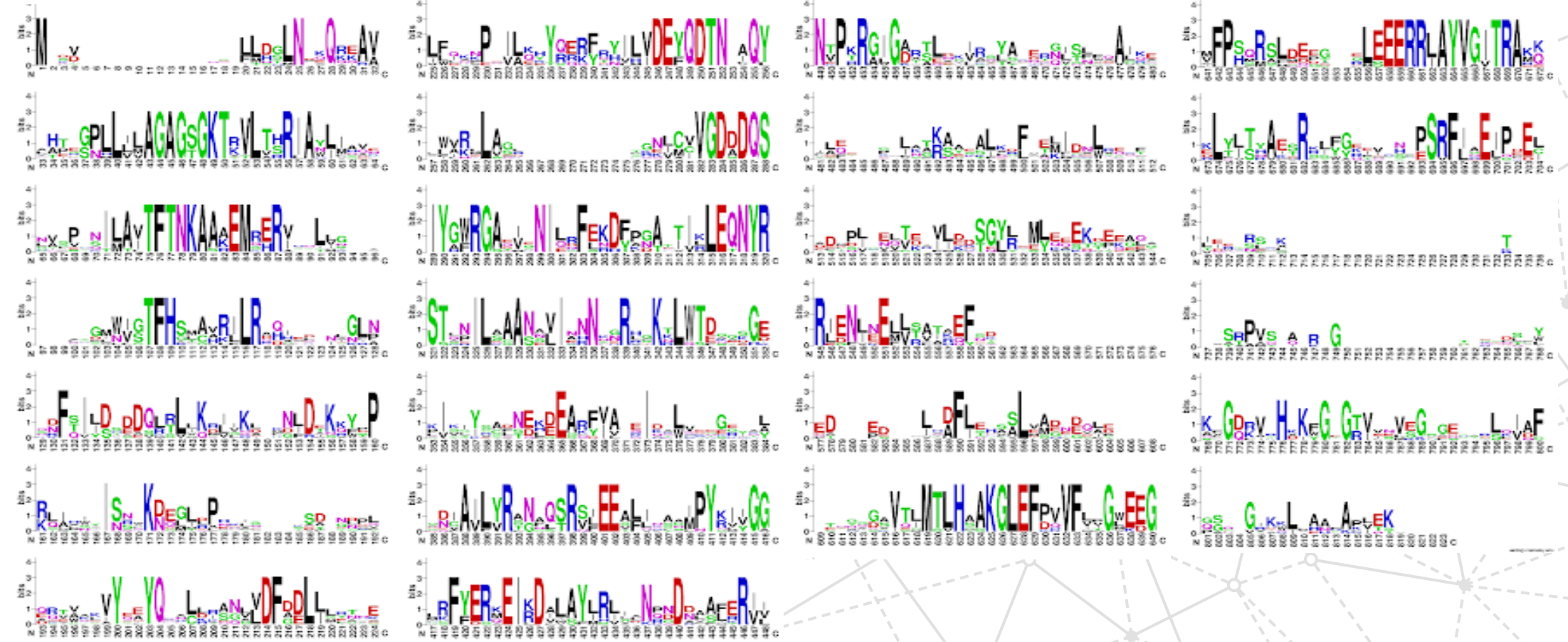


Figure 4 – WebLogo data for *Helicobacter pylori* HP1478 showing the protein is well conserved until amino acid 705-768 where it is poorly conserved.

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.

Gene Locus	Geni-Act Gene Products	Proposed Annotation
HP1253	Tryptophanyl-tRNA synthetase	Tryptophanyl-tRNA synthetase
HP1126	Tol-Pal system beta propeller repeat protein TolB	Tol-Pal system beta propeller repeat protein TolB
HP1329	Heavy metal efflux pump	Heavy metal efflux pump
HP1478	ATP-dependent DNA helicase protein	ATP-dependent DNA helicase 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_Disease_Caused_by_Helicobacter_pylori

Acknowledgments

Supported by an NIH Science Education Partnership (SEPA) Award - R25ODO10536
www.buffalo.edu