University at Buffalo *The State University of New York* | REACHING OTHERS

Annotation of the *Helicobacter pylori* strain 26695 at Locus Tags C694_01655, C694_01460, C694_04745, C694_00340, C694_00315 and C694_02685

Sabrina Donohue, Claudia Hardie, Jackie Keicher, Jackie Kibler, Megan Travis, Harmony Worden, Laura O'Donnel
and Elizabeth Winters
Attica High School and The Western New York Genetics in Research Partnership

Abstract

A group of 6 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 helix shaped, gram negative bacteria that lives in human digestive tracts. An infection of the *Helicobacter pylori* bacteria is the leading cause of gastric cancer as well as gastritis and peptic ulcers in humans. The bacteria living in the digestive tract can cause ulcers in either the stomach lining or the upper section of the small intestine. These ulcers form when the bacteria wear down the stomach or intestinal lining allowing the body's natural acid in.

H. pylori has five major outer membrane families of protein. The largest family are adhesins which help the bacteria bind to lipids in the epithelial lining of the stomach. The other four include porins, which allow molecules to diffuse through the membrane, iron transporters, flagellum associated proteins and proteins of unknown function

The *Helicobacter pylori* bacteria can be received through food, water, or utensils, or through the transfer of bodily fluids from an infected person. Although this bacteria is found worldwide, it is most commonly found in developing countries that don't have sanitary water or sewage systems. In these countries, ten percent of children and eighty percent of adults are expected to have or have had a *H. pylori* infection. Recent research shows that since mobility is critical for the survival of *H. pylori*, they can create a vaccine will target the flagella and help increase protection from *H. pylori.*

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

Methods and Materials

Modules of the GENI-ACT (http://www.geni-act.org/) were used to complete *Helicobacter pylori* genome annotation. The modules are described below:

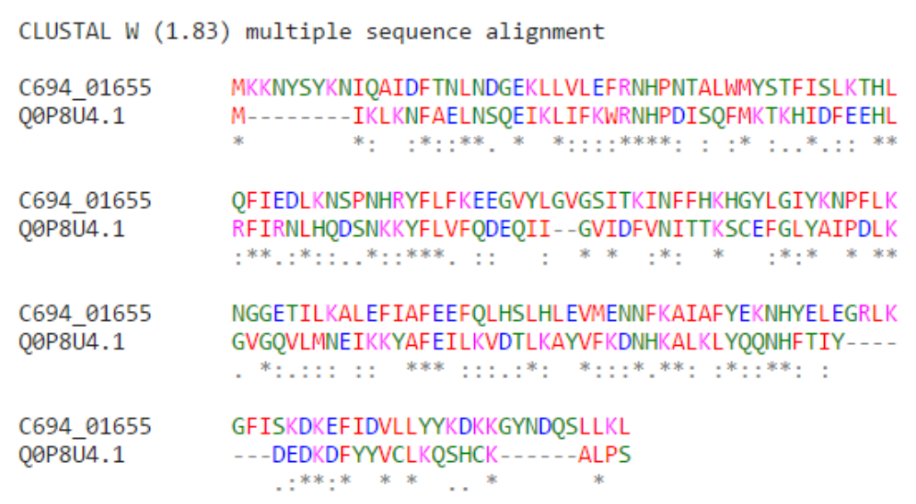
Modules	Activities	Questions Investigated
Module 1- Basic Information Module	DNA Coordinates and Sequence, Protein Sequence	What is the sequence of my gene and protein? Where is it located in the genome?
Module 2- Sequence-Based Similarity Data	Blast, CDD, T-Coffee, WebLogo	Is my sequence similar to other sequences in Genbank?
Module 3- 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?

Results

C694_01655		
Basic Information	**DNA Coordinates**	343118..343660
	DNA Length	543 bp
	Amino Acid Sequence Length	180 aa
Sequence Based Similarity	**COGs**	COG 1670 RimL Protein N-acetyltransferase, RimJ/RimL family [Translation, ribosomal structure and biogenesis]
	TCOFFEE	Too few orthologs to show conservation
	WEBLOGO	Too few orthologs to show conservation
Cellular Localization	**Transmembrane Helices**	none
	Signal Peptide	none
	PSORTB Final Prediction	Cytoplasmic
Structure Based Evidence	**TIGRFAMs**	PseH: pseudaminic acid biosynthesis N-acetyl
	PFAMs	Acetyltransferase
	HMM Logo Key Residues	D8, G131, F132
	PDB	Crystal structure of *H. pylori* pseudaminic acid biosynthesis N -acetyltransferase PseH complex with acetyl-coA

Helicobacter pylori C694_01655:

From the BLAST search there were not a lot of orthologs, which means the protein may be specific to this strain of bacteria. Data from the COG, TIGRFAM, PFAM, and PDB databases indicate this protein may be an acetyl transferase responsible in transferring an acetyl group, important in the creation of pseudaminic acid. Pseudaminic acid is a sugar that can be used to power flagellum movement through the thick mucus lining of the stomach.



```
CLUSTAL W (1.83) multiple sequence alignment

C694_01655      MKKNYSYKNIQAIDFTNLNDGEKLLVLEFRNHPNTALWMYSTFISLKTHL
Q0P8U4.1        M--------IKLKNFAELNSQEIKLIFKWRNHPDISQFMKTKHIDFEEHL
                *        *:  :*::**. *  *::::****: : :* :..*.:: **

C694_01655      QFIEDLKNSPNHRYFLFKEEGVYLGVGSITKINFFHKHGYLGIYKNPFLK
Q0P8U4.1        RFIRNLHQDSNKKYFLVFQDEQII--GVIDFVNITTKSCEFGLYAIPDLK
                :**.:*::..*::***. ::    :  * *  :*:  *    :*:*  * **

C694_01655      NGGETILKALEFIAFEEFQLHSLHLEVMENNFKAIAFYEKNHYELEGRLK
Q0P8U4.1        GVGQVLMNEIKKYAFEILKVDTLKAYVFKDNHKALKLYQQNHFTIY----
                . *:.::: ::  *** ::::.:*:  *:::*.**: :*::**: :

C694_01655      GFISKDKEFIDVLLYYKDKKGYNDQSLLKL
Q0P8U4.1        ---DEDKDFYYVCLKQSHCK------ALPS
                  .:**:*  * *  .. *       *
```

Figure I – T-Coffee Data for *Helicobacter pylori C694_01655*
The lack of orthologs to show protein conservation indicates this protein may be unique to this strain of *Helicobacter pylori.*

C694_01460		
Basic Information	**DNA Coordinates**	298029..306710
	DNA Length	8682 bp
	Amino Acid Sequence Length	2893 aa
Sequence Based Similarity	**COGs**	COG 5651 PPE-repeat protein (unknown function)
	TCOFFEE	Too few orthologs to show conservation
	WEBLOGO	Too few orthologs to show conservation
Cellular Localization	**Transmembrane Helices**	one
	Signal Peptide	none
	PSORTB Final Prediction	Outermembrane
Structure Based Evidence	**TIGRFAMs**	Auto-barl:outermembrane autotransporter
	PFAMs	Putative vacuolating cytotoxin
	HMM Logo Key Residues	G4, G24, G30, G31, F37
	PDB	no significant hits

Helicobacter pylori C694_01460:

PFAM data supports C694_01460 to be a putative vacuolating cytotoxin which breaks down mucus membranes in the stomach allowing *H. pylori* to reach the stomach lining, causing an ulcer. This function is further supported by Cellular Localization Data which predicted its location to be in the outermembrane. The BLAST Database did not show any orthologs for this protein, indicating that it may be unique to this strain of *Helicobacter pylori*.

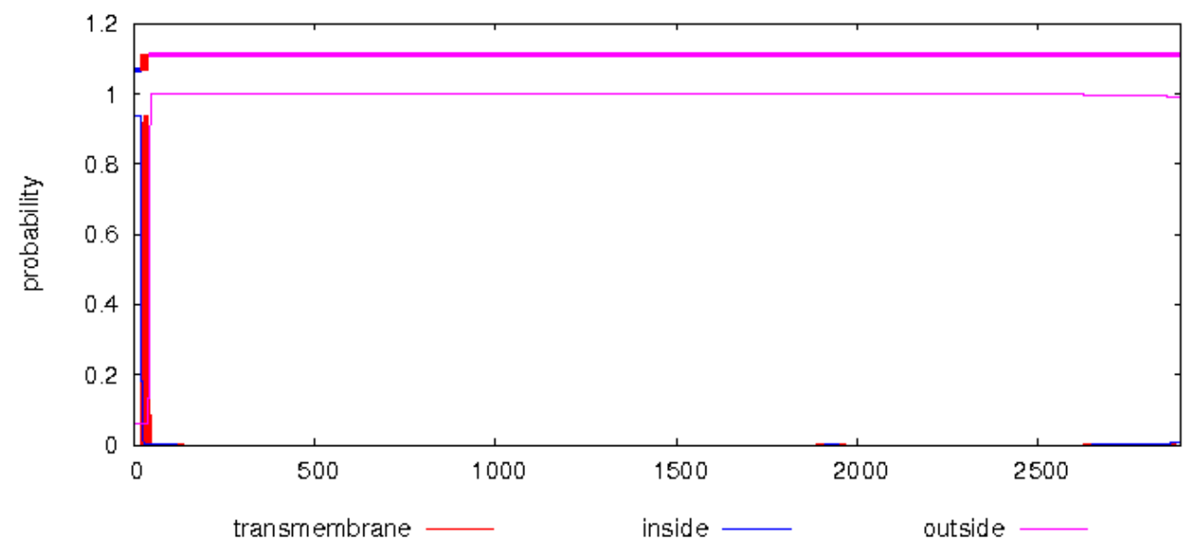


Figure II – TMHMM Data for *Helicobacter pylori* C694_01460.
The TMHMM image to the right shows the presence of one transmembrane helix which is predicted to be in the outermembrane of the bacteria.

C694_04745		
Basic Information	**DNA Coordinates**	979446..987035
	DNA Length	7590 bp
	Amino Acid Sequence Length	2529 aa
Sequence Based Similarity	**COGs**	COG 5651 PPE-repeat protein (unknown function)
	TCOFFEE	Too few orthologs to show conservation
	WEBLOGO	Too few orthologs to show conservation
Cellular Localization	**Transmembrane Helices**	none
	Signal Peptide	none
	PSORTB Final Prediction	Outermembrane
Structure Based Evidence	**TIGRFAMs**	No significant TIGRFAMs
	PFAMs	Putative vacuolating cytotoxin
	HMM Logo Key Residues	G4, G24, G30, G31, F37
	PDB	no significant hits

Helicobacter pylori C694_04745:

PFam data supports C694_04745 to be a VacA toxin. Secretion of this toxin enhances the ability of the bacteria to colonize the stomach and contributes to peptic ulcer disease. There is conflicting information from the Cellular Localization Data. The TMHMM database shows no transmembrane helices, while Phobius shows the presences of one. The PSORTB final prediction is outermembrane, but also extracellular.

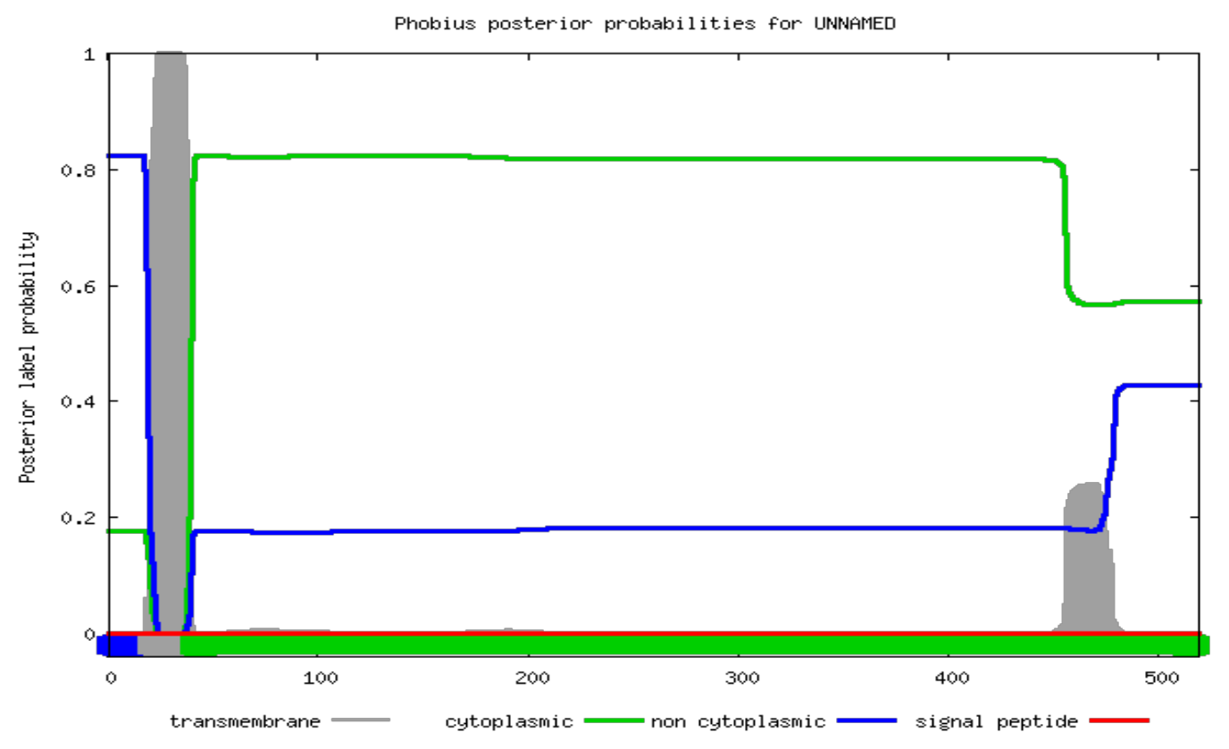


Figure III – Cellular Localization from the Phobius Database for *Helicobacter pylori* C694_04745.
The image to the right shows the presence of one transmembrane helix, most likely present in the outermembrane of the bacteria.

C694_00340		
Basic Information	**DNA Coordinates**	74747...75334
	DNA Length	588 bp
	Amino Acid Sequence Length	195 aa
Sequence Based Similarity	**COGs**	no significant hits
	TCOFFEE	Sporadically conserved throughout
	WEBLOGO	Too few orthologs to show conservation.
Cellular Localization	**Transmembrane Helices**	6
	Signal Peptide	none
	PSORTB Final Prediction	Cytoplasmic membrane
Structure Based Evidence	**TIGRFAMs**	no significant hits
	PFAMs	AmiS/Urel family transport
	HMM Logo Key Residues	F67, Y71, W118, W121, W125, P158
	PDB	Crystal structure of urea channel from the human gastric pathogen *Helicobacter pylori*

Helicobacter pylori C694_00340:

PFAM and PDB data support C694_00340 to be a urea channel found in the membrane. The urea channel allows rapid access of urea to the urease *H. pylori* produces. Urease breaks down urea into ammonia and bicarbonate to neutralize stomach acid. TMHMM data shows it has six transmembrane helices.

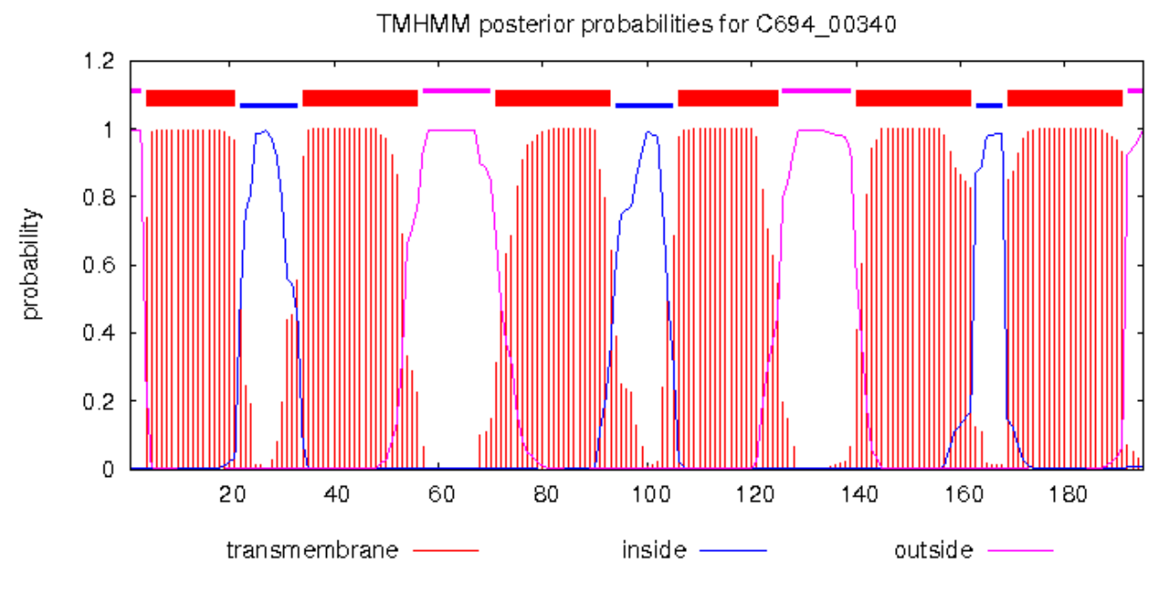


Figure IV – Cellular Localization from the TMHMM Database for *Helicobacter pylori* C694_00340.
The image to the left shows the presence of six transmembrane helices.

C694_00315		
Basic Information	**DNA Coordinates**	69537..71963
	DNA Length	2427 bp
	Amino Acid Sequence Length	808 aa
Sequence Based Similarity	**COGs**	COG 1674 FtsK DNA segregation ATPase FtsK/SpoIIIE and related proteins [Cell cycle control, cell division]
	TCOFFEE	Too few orthologs to show conservation.
	WEBLOGO	Too few orthologs to show conservation.
Cellular Localization	**Transmembrane Helices**	none
	Signal Peptide	none
	PSORTB Final Prediction	Cytoplasmic membrane
Structure Based Evidence	**TIGRFAMs**	no significant hits
	PFAMs	FtsK/SpoIIIE
	HMM Logo Key Residues	G46, G51, K52, D154, H233
	PDB	*P. aeruginosa* FTSK motor domain hexamer

Helicobacter pylori C694_00315:

COG and PFam data supports C694_00340 to be FtsK, a double stranded DNA translocase which acts as a motor protein that converts the chemical energy of ATP into movement of DNA during replication in cell division.

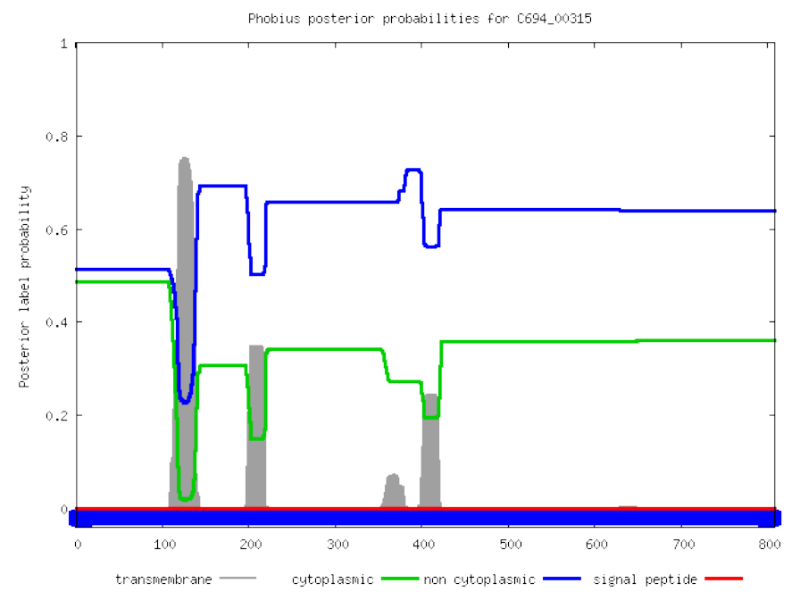


Figure V – Cellular Localization from the Phobius Database for *Helicobacter pylori* C694_00340.
The image shows the possible presence of one transmembrane helix, most likely present in the cytoplasmic membrane of the bacteria. It was also found to be able to move around the cytoplasm.

C694_02685		
Basic Information	**DNA Coordinates**	548035..548151
	DNA Length	117 bp
	Amino Acid Sequence Length	38 aa
Sequence Based Similarity	**COGs**	no significant hits
	TCOFFEE	no orthologs to show conservation
	WEBLOGO	no orthologs to show conservation
Cellular Localization	**Transmembrane Helices**	one
	Signal Peptide	none
	PSORTB Final Prediction	Unknown
Structure Based Evidence	**TIGRFAMs**	no significant hits
	PFAMs	no significant hits
	HMM Logo Key Residues	n/a
	PDB	no significant hits

Helicobacter pylori C694_02685:

Although there were no orthologs found, BLAST data indicates this protein may be found only in this strain of *H. pylori* and could be a cag pathogenicity island protein. The function of this protein may be to help the bacteria invade host epithelial cells. TMHMM data does suggest this protein might be found in the membrane.

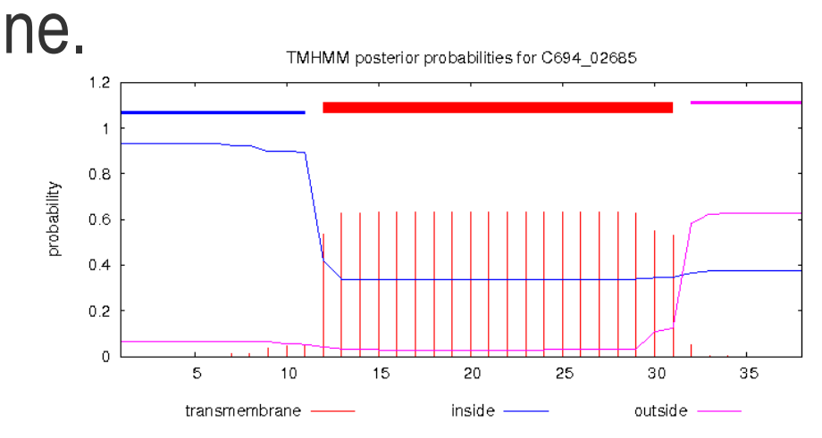


Figure VI – Cellular Localization from the TMHMM Database for *Helicobacter pylori* C694_02865.
The image to the right shows the presence of one transmembrane helix.

Conclusion

The Geni-Act proposed gene product did not differ significantly from the proposed gene annotation for each of the genes and as such, the genes appear to be correctly annotated by the computer database.

References

Foegeding, N. J., Caston, R. R., McClain, M. S., Ohi, M. D., & Cover, T. L. (2016, June). An Overview of *Helicobacter pylori* VacA Toxin Biology. Retrieved May 19, 2017, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4926140/

Acknowledgments

Supported by NSF ITEST Strategies Award Number 1311902

Gene Locus	Proposed Annotation
01655	Acetyltransferase involved with pseudaminic acid biosynthesis
01460	Putative vacuolating cytotoxin
04745	Putative vacuolating cytotoxin
00340	AmiS/UreI family transport
00315	FtsK, DNA translocase motor protein
02685	Cag pathogenicity island protein