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

SEPA SCIENCE EDUCATION PARTNERSHIP AWARD Supported by the National Institutes of Health

Annotation of the *Helicobacter pylori* Genome from DNA Coordinates 45733 to 50762

Rebekah Boy, Samatha Freiberger, Lindsey Nudds, Sorcha Lester and Lawrence Hohl

East Rochester Junior/Senior High School and the Western New York Genetics in Research and Healthcare Partnership

Abstract

We were given a segment of the *Helicobacter pylori 83* genome to analyze whether the particular genes 0045, 0046, 0049, and 0050 were correctly annotated. We used the GENI-ACT modules to analyze these genes, and the data indicated that the proposed annotations for all four of these genes were correct. The proposed annotations are that gene 0045 codes for a competence protein, which is involved in trans-membrane transport; gene 046 codes for 4-hydroxybenzoateoctaprenyltransferase, which is involved in ubiquinone-8 synthesis; gene 0049 codes for phosphatidylserine decarboxylase, which is involved in aminophospholipid synthesis; and gene 0050 codes for quinolinate synthase, which is involved in NAD synthesis.

Introduction

Helicobacter pylori 83 is a gram-negative bacteria, which burrows into the epithelial cells of the stomach (Amieva et al., 2008). It contains a hydrogenase, allowing it to gain energy through the oxidation of molecular hydrogen produced in the intestine (Olson et al., 2002). It also produces urease to neutralize the gastric acid around itself, which produces ammonia and damages the surrounding epithelial cells, inflaming the cells and often causing gastric and duodenal ulcers. Other chemicals it produces, such as protease and lipase, have a similar effect (Smoot, 1997). H. pylori has been shown to be the main direct cause of distal gastric adenocarcinoma and gastric lymphoma, but only 15% of the bacteria develop these diseases. Research into this bacterium and how it becomes cancerous may lead to a cure for these forms of gastric cancer (Atherton, 2006).

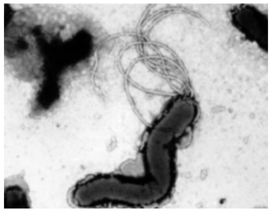


Figure I – This picture shows the helical shape of *Helicobacter pylori* 83, and shows the flagella that help it to burrow into the membrane of the stomach.

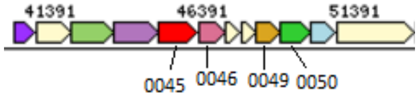


Figure II – This diagram shows the locus tags of the genes analyzed in this annotation relative to each other and the DNA coordinates.

Methods and Materials

Modules of the GENI-ACT (http://www.geni-act.org/) were used to complete *Helicobacter pylori 83* 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?
Module 5- Structure-Based Evidence	TIGRfam, Pfam, PDB	Are there functional domains in my protein?
Module 6- Enzymatic Function	KEGG, MetaCyc, E.C. Number,	In what process does my protein take part?
Module 7- Gene Duplication/ Gene Degradation	Paralog, Pseudogene	Are there other forms of my gene in the bacterium? Is my gene functional?
Module 8- Evidence for Horizontal Gene Transfer	Phylogenetic Tree,	Has my gene co-evolved with other genes in the genome?
Module 9- RNA	RFAM	Does my gene encode a functional RNA?

Results

Helicobacter pylori 83 0045:

The function of this gene is that it codes for a protein of the competence family that assists with transport across the cell membrane.

Helicobacter pylori 83 0046:

The function of this gene is that it aids in the attachment of a polyprenyl group to para-hydroxybenzoate as a step in the biosynthesis of ubiquinone-8.

Helicobacter pylori 83 0049:

The function of this gene is that it plays a central role in the biosynthesis of aminophospholipids by converting phosphatidylserine to phosphatidylethanolamine.

Helicobacter pylori 83 0050:

The function of this gene is that it codes for quinolinate synthase, an enzyme that aids in NAD synthesis.

The gene 0045 in *Helicobacter pylori 83* codes for a comE protein of the competence family. This protein assists with transport across the cell membrane and provides the materials needed for the synthesis reactions of genes 0046, 0049, and 0050. This protein is present in the cytoplasmic membrane of the cell.

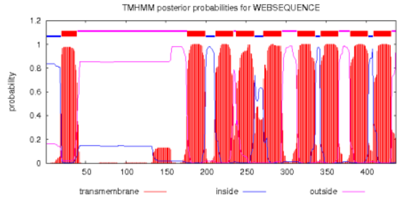


Figure III – This graph shows that the gene *Helicobacter 83 0045* which codes for a protein of the competence family possesses 9 transmembrane helices, proving that it assists with transmembrane transportation.

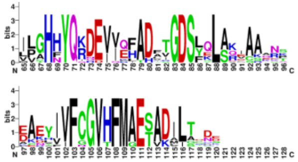
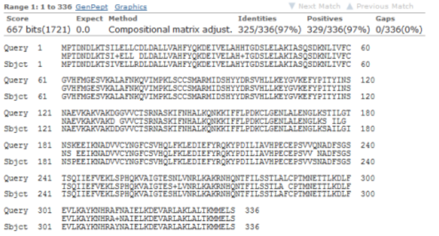


Figure IV -This diagram shows a portion of a WebLogo of gene 0050, showing the similarity between the gene sequence in *Helicobacter pylori* 83 and the same gene in other species. This segment shows that this gene is well conserved since many of the species have the same amino acids in the same places

```
Range 1: 1 to 267 GenPept  Graphics                              ▼ Next Match  ▲ Previous
Score           Expect  Method                       Identities     Positives      Gaps
539 bits(1388)  0.0     Compositional matrix adjust. 262/267(98%)   265/267(99%)   0/267(0%)

Query  1    MIALSNALSRLFGSVAGYEFPSFIQKGINALYVKIFKIDLSEFEPLENYKSLNALFTRSL  60
            MIALSNALSR+FGSVAGYEFPSFIQKGINALYVKIFKIDLSEFEPLENYKSLNALFTRSL
Sbjct  1    MIALSNALSRVFGSVAGYEFPSFIQKGINALYVKIFKIDLSEFEPLENYKSLNALFTRSL  60

Query  61   KKERPFDKAPNACIAPCDALITECAFLDNDSALQIKGMPYKAHELVGEINPLSPSFFYVN  120
            KKERPFDK PNACIAPCDALITECAFLDNDSALQIKGMPYKAHELVGEINPLSPSFFYVN
Sbjct  61   KKERPFDKTPNACIAPCDALITECAFLDNDSALQIKGMPYKAHELVGEINPLSPSFFYVN  120

Query  121  FYLSPKDYHHYHAPCDLEILEARCFAGKLLPVNKPSLYKNKNLFVGNERVALVAKDIQGN  180
            FYLSPKDYHHYHAPCDLEILEARCFAGKLLPVNKPSL+KNKNLFVGNERVALVAKDIQGN
Sbjct  121  FYLSPKDYHHYHAPCDLEILEARCFAGKLLPVNKPSLHKNKNLFVGNERVALVAKDIQGN  180

Query  181  RLYFVAVGALNVGKMRFNFDKNIQTNAKAHLTQTYSYNPPIKVKKGDNLGNFEMGSTIVL  240
            RLYFVAVGALNVGKMRFNFDKNIQTNAKAHLTQ YSYNPPIKVKKGDNLGNFEMGSTIVL
Sbjct  181  RLYFVAVGALNVGKMRFNFDKNIQTNAKAHLTQAYSYNPPIKVKKGDNLGNFEMGSTIVL  240

Query  241  FIQNTAFKDLKEKSVKFGESIGEFHAN  267
            F+QNTAFKDLKEKSVKFGESIGEFHAN
Sbjct  241  FVQNTAFKDLKEKSVKFGESIGEFHAN  267
```

Figure V – This diagram shows a BLAST search comparing the protein sequence of gene 0049 to a sequence known to convert phosphatidylserine to phosphatidylethanolamine. The sequences are exceptionally similar to each other, which proves that they most likely are the same.



```
Range 1: 1 to 336 GenPept  Graphics                              ▼ Next Match  ▲ Previous Match
Score           Expect  Method                       Identities     Positives      Gaps
667 bits(1721)  0.0     Compositional matrix adjust. 325/336(97%)   329/336(97%)   0/336(0%)

Query  1    MPTDNDLKTSILELLCDLDALLVAHFYQKDEIVELAHHTGDSLELAKIASQSDKNLIVFC  60
            MPTDNDLKTSI+ELL DLDALLVAHFYQKDEIVELAH+TGDSLELAKIASQSDKNLIVFC
Sbjct  1    MPTDNDLKTSIVELLRDLDALLVAHFYQKDEIVELAHYTGDSLELAKIASQSDKNLIVFC  60

Query  61   GVHFMGESVKALAFNKQVIMPKLSCCSMARMIDSHYYDRSVHLLKEYGVKEFYPITYINS  120
            GVHFMGESVKALAFNKQVIMPKLSCCSMARMIDSHYYDRSVHLLKEYGVKEFYPITYINS
Sbjct  61   GVHFMGESVKALAFNKQVIMPKLSCCSMARMIDSHYYDRSVHLLKEYGVKEFYPITYINS  120

Query  121  NAEVKAKVAKDGGVVCTSRNASKIFNHALKQNKKIFFLPDKCLGENLALENGLKSTILGT  180
            NAEVKAKVAKD GVVCTSRNASKIFNHALKQNKKIFFLPDKCLGENLALENGLKS ILG
Sbjct  121  NAEVKAKVAKDDGVVCTSRNASKIFNHALKQNKKIFFLPDKCLGENLALENGLKSAILGI  180

Query  181  NSKEEIKNADVVCYNGFCSVHQLFKLEDIEFYRQKYPDILIAVHPECEPSVVQNADFSGS  240
            NS EEIKNADVVCYNGFCSVHQLFKLEDIEFYRQKYPDILIAVHPECEPSVV NADFSGS
Sbjct  181  NSPEEIKNADVVCYNGFCSVHQLFKLEDIEFYRQKYPDILIAVHPECEPSVVSNADFSGS  240

Query  241  TSQIIEFVEKLSPHQKVAIGTESNLVNRLKAKRNHQNTFILSSTLALCPTMNETTLKDLF  300
            TSQIIEFVEKLSPHQKVAIGTES+LVNRLKAKRNHQNTFILSSTLA CPTMNETTLKDLF
Sbjct  241  TSQIIEFVEKLSPHQKVAIGTESHLVNRLKAKRNHQNTFILSSTLAFCPTMNETTLKDLF  300

Query  301  EVLKAYKNHRAFNAIELKDEVARLAKLALTKMMELS  336
            EVLKAYKNHRA+NAIELKDEVARLAKLALTKMMELS
Sbjct  301  EVLKAYKNHRAYNAIELKDEVARLAKLALTKMMELS  336
```

Figure VI – This diagram shows a BLAST search comparing the protein sequence of gene 0050 to a sequence known to code for the enzyme quinolinate synthase. The two sequences are extremely similar, proving that this gene very likely codes for quinolinate synthase.

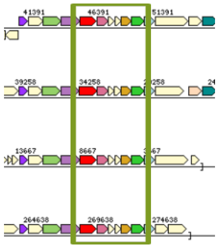


Figure VII – The loci under investigation in this diagram are *Helicobacter pylori 83* genes 0045, 0046, 0049, and 0050, as shown in Figure II. This diagram shows the annotated genes on the DNA of multiple organisms. It is evident that the gene neighborhoods for these loci are very well conserved.

Conclusion

The GENI-ACT proposed gene product did not differ significantly from the proposed gene 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	Gene Function
0045	Competence protein	Assists with transport across cell membrane
0046	4-hydroxybenzoate octaprenyltransferase	Aids in the attachment of polyprenyl group to para-hydroxybenzoate as a step in the biosynthesis of ubiquinone-8
0049	Phosphatidylserine decarboxylase	Plays central role in biosynthesis of aminophospholipids
0050	Quinolinate synthetase	Aids in NAD synthesis

References

1. Amieva, M., and El-Omar, E. 2008. Host-bacterial interactions in *Helicobacter pylori* infection. http://www.gastrojournal.org/article/S0016-5085(07)02016-1/abstract
2. Olson, J., and Maier, R. 2002. Molecular hydrogen as an energy source for Helicobacter pylori. Science 298 : 1788-1790
3. Smoot, D. 1997. How does Helicobacter pylori cause mucosal damage? Direct mechanisms. http://www.gastrojournal.org/article/S0016-5085(97)80008-X/abstract
4. Atherton, J. The pathogenesis of Helicobacter pylori-induced gastro-duodenal diseases. http://www.ncbi.nlm.nih.gov/pubmed/18039108
5. Google images

Acknowledgments

Supported by an NIH Science Education Partnership (SEPA) Award R25OD010536-01A1
Stephen T. Koury, Ph.D. University at Buffalo
Jennifer R. Cooke, MPS, CSA R-AHEC
East Rochester Tech Department