Annotation of the *Helicobacter pylori* Genome from DNA Coordinates 48459 to 60086

Rebekah Boy and Lawrence Hohl
East Rochester Junior/Senior High School, East Rochester 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

I was given a segment of the *Helicobacter pylori 83* genome to annotate, to attempt to find the probable products of these genes, as there was no predicted product for these four in GENI-ACT. Using the modules of GENI-ACT to analyze them, I found evidence that gene 0048 likely contains a helix-turn-helix domain, gene 0052 is likely an adenine-specific DNA methyltransferase, gene 0055 is likely a type II restriction endonuclease, and gene 0057 is likely located in the periplasmic membrane of the bacterium. As type II restriction endonucleases often work in tandem with methyltransferases, and genes 0048 and 0055 are close together, it may be plausible that the products of these two genes work together. Gene 0057 may also be related to these genes, as helix-turn-helix domains are often involved in regulation of DNA.

Introduction

Helicobacter pylori 83 is a gram-negative bacteria, which burrows into the epithelial cells of the stomach (Amieva et al., 2008).It 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).

Like many other bacteria, *Helicobacter pylori* contains type II restriction endonucleases, which cleave foreign DNA a set number of nucleotides after a particular recognition sequence, as a defense mechanism. Often, methyltransferases work in tandem with these enzymes, recognizing the same sequence in its own DNA that an endonuclease does, to add a methyl group to it, thus protecting it from being cleaved by the endonuclease (New England BioLabs, 2017).

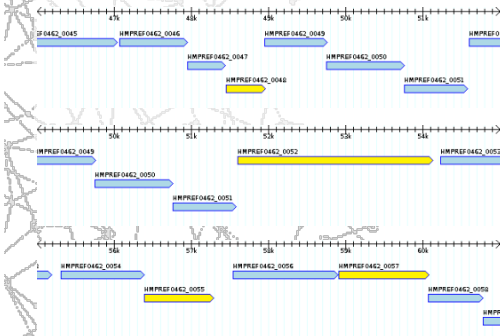


Figure 1. The locus tags and relative position of the genes under investigation in this research

Methods

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

The proposed products of these four genes by GENI-ACT were all hypothetical proteins.

HMPREF0462_0048:

The proposed annotation of this gene is a helix-turn-helix domain-containing protein because BLAST hits showed that this amino acid sequence a had a very strong correlation with sequences known to code for them.

HMPREF0462_0052:

BLAST hits and structure-based evidence, as well as its localization within the cytoplasm support its categorization as a DNA methyltransferase. Thus, the proposed annotation of this gene is an adenine-specific DNA methyltransferase.

HMPREF0462_0055:

The proposed annotation of this gene is a type II restriction endonuclease, based on strong BLAST results comparing the gene with ones known to code for restriction enzymes.

HMPREF0462_0057:

This gene codes for a protein that is most likely located in the periplasmic membrane, due to the presence of transmembrane helices.

Due to the proximity of genes 0052 and 0055, and their annotations being a restriction endonuclease and methyltransferase, which work closely together, these two genes may have related functions.

As gene 0048 is also located near these genes, and likely contains a helix-turn-helix domain, which is also associated with gene regulation, this gene's function may also be related with that of genes 0052 and 0055.

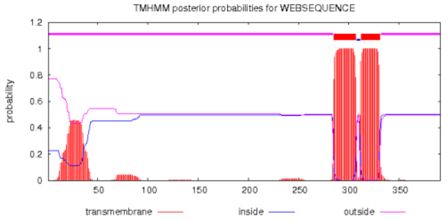
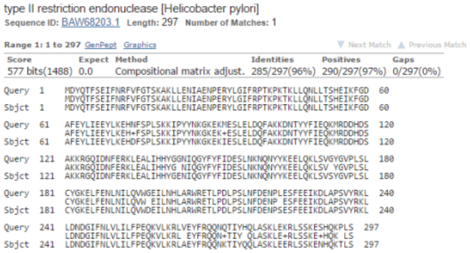
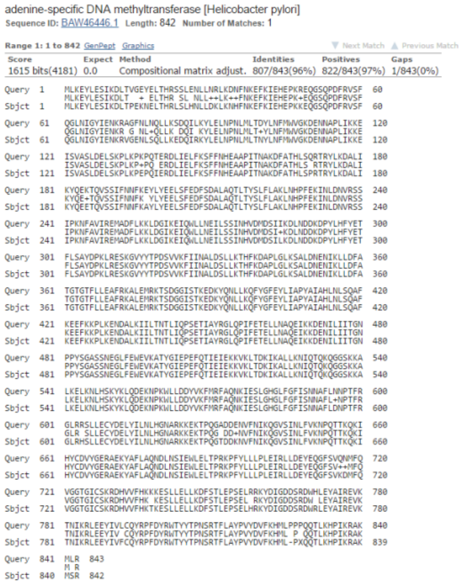


Figure 2 – TMHMM graph for gene 0057 shows 2 transmembrane helices, supporting its localization in the periplasmic membrane.



```
type II restriction endonuclease [Helicobacter pylori]
Sequence ID: BAW68203.1  Length: 297  Number of Matches: 1

Range 1: 1 to 297 GenPept  Graphics                    ▼ Next Match  ▲ Previous Match
Score            Expect  Method                        Identities      Positives       Gaps
577 bits(1488)   0.0     Compositional matrix adjust.  285/297(96%)    290/297(97%)    0/297(0%)

Query  1    MDYQTFSEIFNRFVFGTSKAKLLENIAENPERYLGIFRPTKPKTKLLQNLLTSHEIKFGD  60
            MDYQTFSEIFNRFVFGTSKAKLLENIAENPERYLGIFRPTKPKTKLLQNLLTSHEIKFGD
Sbjct  1    MDYQTFSEIFNRFVFGTSKAKLLENIAENPERYLGIFRPTKPKTKLLQNLLTSHEIKFGD  60

Query  61   AFEYLIEEYLKEHNFSPLSKKIPYYNKGKEKMESLELDQFAKKDNTYYFIEQKMRDDHDS  120
            AFEYLIEEYLKEH+FSPLSKKIPYYNKGKEK+ESLELDQFAKKDNTYYFIEQKMRDDHDS
Sbjct  61   AFEYLIEEYLKEHDFSPLSKKIPYYNKGKEKIESLELDQFAKKDNTYYFIEQKMRDDHDS  120

Query  121  AKKRGQIDNFERKLEALIHHYGGNIQGYFYFIDESLNKNQNYYKEELQKLSVGYGVPLSL  180
            AKKRGQIDNFERKLEALIHHYG NIQGYFYFIDESLNKNQNYYKEELQKLSV YGVPLSL
Sbjct  121  AKKRGQIDNFERKLEALIHHYGENIQGYFYFIDESLNKNQNYYKEELQKLSVSYGVPLSL  180

Query  181  CYGKELFENLNILQVWGEILNHLARWRETLPDLPSLNFDENPLESFEEIKDLAPSVYRKL  240
            CYGKELFENLNILQVW EILNHLARWRETLPDLPSLNFDENP ESFEEIKDLAPSVYRKL
Sbjct  181  CYGKELFENLNILQVWDEILNHLARWRETLPDLPSLNFDENPSESFEEIKDLAPSVYRKL  240

Query  241  LDNDGIFNLVLILFPEQKVLKRLVEYFRQQNQTIYHQLASKLEKRLSSKESHQKPLS  297
            LDNDGIFNLVLILFPEQKVLKRL EYFRQQN+TIY QLASKLE+RLSSKE+HQK LS
Sbjct  241  LDNDGIFNLVLILFPEQKVLKRLAEYFRQQNKTIYQQLASKLEERLSSKENHQKTLS  297
```

Figure 3 – BLAST comparing the sequences of gene 0055 and a type II restriction endonuclease, showing a very strong correlation.



```
adenine-specific DNA methyltransferase [Helicobacter pylori]
Sequence ID: BAW46446.1  Length: 842  Number of Matches: 1

Range 1: 1 to 842 GenPept  Graphics                     ▼ Next Match  ▲ Previous Match
Score            Expect  Method                        Identities      Positives       Gaps
1615 bits(4181)  0.0     Compositional matrix adjust.  807/843(96%)    822/843(97%)    1/843(0%)

Query  1    MLKEYLESIKDLTVGEYELTHRSSLENLLNRLKDNFNKEFKIEHEPKREQGSQPDFRVSF  60
            MLKEYLESIKDLT  + ELTHR SL NLL++LK++FNKEFKIEHEPK+EQGSQPDFRVSF
Sbjct  1    MLKEYLESIKDLTPEKNELTHRLSLHNLLDKLKNHFNKEFKIEHEPKKEQGSQPDFRVSF  60

Query  61   QGLNIGYIENKRAGFNLNQLLKSDQILKYLELNPNLMLTDYLNFMWVGKDENNAPLIKKE  120
            QGLNIGYIENKR G NL+QLLK DQI KYLELNPNLMLT+YLNFMWVGKDENNAPLIKKE
Sbjct  61   QGLNIGYIENKRVGENLSQLLKEDQIRKYLELNPNLMLTNYLNFMWVGKDENNAPLIKKE  120

Query  121  ISVASLDELSKPLKPKPQTERDLIELFKSFFNHEAAPITNAKDFATHLSQRTRYLKDALI  180
            ISVASLDELSKPLKP+PQ ERDLIELFKSFFNHEAAPITNAKDFATHLS RTRYLKDALI
Sbjct  121  ISVASLDELSKPLKPEPQIERDLIELFKSFFNHEAAPITNAKDFATHLSPRTRYLKDALI  180

Query  181  KYQEKTQVSSIFNNFKEYLYEELSFEDFSDALAQTLTYSLFLAKLNHPFEKINLDNVRSS  240
            KYQE+TQVSSIFNNFK YLYEELSFEDFSDALAQTLTYSLFLAKLNHPFEKINLDNVRSS
Sbjct  181  KYQEETQVSSIFNNFKAYLYEELSFEDFSDALAQTLTYSLFLAKLNHPFEKINLDNVRSS  240

Query  241  IPKNFAVIREMADFLKKLDGIKEIQWLLNEILSSINHVDMDSIIKDLNDDKDPYLHFYET  300
            IPKNFAVIREMADFLKKLDGIKEIQWLLNEILSSINHVDMDSI+KDLNDDKDPYLHFYET
Sbjct  241  IPKNFAVIREMADFLKKLDGIKEIQWLLNEILSSINHVDMDSILKDLNDDKDPYLHFYET  300

Query  301  FLSAYDPKLRESKGVYYTPDSVVKFIINALDSLLKTHFKDAPLGLKSALDNENIKLLDFA  360
            FLSAYDPKLRESKGVYYTPDSVVKFIINALDSLLKTHFKDAPLGLKSALDNENIKLLDFA
Sbjct  301  FLSAYDPKLRESKGVYYTPDSVVKFIINALDSLLKTHFKDAPLGLKSALDNENIKLLDFA  360

Query  361  TGTGTFLLEAFRKALEMRKTSDGGISTKEDKYQNLLKQFYGFEYLIAPYAIAHLNLSQAF  420
            TGTGTFLLEAFRKALEMRKTSDGGISTKEDKYQNLLKQFYGFEYLIAPYAIAHLNLSQAF
Sbjct  361  TGTGTFLLEAFRKALEMRKTSDGGISTKEDKYQNLLKQFYGFEYLIAPYAIAHLNLSQAF  420

Query  421  KEEFKKPLKENDALKIILTNTLIQPSETIAYRGLQPIFETELLNAQEIKKDENILIITGN  480
            KEEFKKPLKENDALKIILTNTLIQPSETIAYRGLQPIFETELLNAQEIKKDENILIITGN
Sbjct  421  KEEFKKPLKENDALKIILTNTLIQPSETIAYRGLQPIFETELLNAQEIKKDENILIITGN  480

Query  481  PPYSGASSNEGLFEWEVKATYGIEPEFQTIEIEKKVKLTDKIKALLKNIQTQKQGGSKKA  540
            PPYSGASSNEGLFEWEVKATYGIEPEFQTIEIEKKVKLTDKIKALLKNIQTQKQGGSKKA
Sbjct  481  PPYSGASSNEGLFEWEVKATYGIEPEFQTIEIEKKVKLTDKIKALLKNIQTQKQGGSKKA  540

Query  541  LKELKNLHSKYKLQDEKNPKWLLDDYVKFMRFAQNKIESLGHGLFGFISNNAFLNNPTFR  600
            LKELKNLHSKYKLQDEKNPKWLLDDYVKFMRFAQNKIESLGHGLFGFISNNAFL+NPTFR
Sbjct  541  LKELKNLHSKYKLQDEKNPKWLLDDYVKFMRFAQNKIESLGHGLFGFISNNAFLDNPTFR  600

Query  601  GLRRSLLECYDELYILNLHGNARKKEKTPQGADDENVFNIKQGVSINLFVKNPQTTKQKI  660
            GLR SLLECYDELYILNLHGNARKKEKTPQG DD+NVFNIKQGVSINLFVKNPQTTKQKI
Sbjct  601  GLRHSLLECYDELYILNLHGNARKKEKTPQGTDDKNVFNIKQGVSINLFVKNPQTTKQKI  660

Query  661  HYCDVYGERAEKYAFLAQNDLNSIEWLELTPRKPFYLLLPLEIRLLDEYEQGFSVQNMFQ  720
            HYCDVYGERAEKYAFLAQNDLNSIEWLELTPRKPFYLLLPLEIRLLDEYEQGFSV++MFQ
Sbjct  661  HYCDVYGERAEKYAFLAQNDLNSIEWLELTPRKPFYLLLPLEIRLLDEYEQGFSVKDMFQ  720

Query  721  VGGTGICSKRDHVVFHKKKESLLELLKDFSTLEPSELRRKYDIGDDSRDWHLEYAIREVK  780
            VGGTGICSKRDHVVFHK KESLLELLKDFSTLEPSEL RKYDIGDDSRDW LEYAIREVK
Sbjct  721  VGGTGICSKRDHVVFHKTKESLLELLKDFSTLEPSELHRKYDIGDDSRDWRLEYAIREVK  780

Query  781  TNIKRLEEYIVLCQYRPFDYRWTYYTPNSRTFLAYPVYDVFKHMLPPPQQTLKHPIKRAK  840
            TNIKRLEEYIV CQYRPFDYRWTYYTPNSRTFLAYPVYDVFKHML P QQTLKHPIKRAK
Sbjct  781  TNIKRLEEYIVFCQYRPFDYRWTYYTPNSRTFLAYPVYDVFKHML-PXQQTLKHPIKRAK  839

Query  841  MLR  843
            M R
Sbjct  840  MSR  842
```

Figure 4 – BLAST comparing the sequences of gene 0052 and an adenine-specific methyltransferase, showing a strong correlation.

```
helix-turn-helix domain-containing protein [Helicobacter pylori]
Sequence ID: WP_075711202.1  Length: 167  Number of Matches: 1
▶ See 1 more title(s)

Range 1: 1 to 167 GenPept  Graphics                     ▼ Next Match  ▲ Previous Match
Score            Expect  Method                        Identities      Positives       Gaps
326 bits(835)    4e-115  Compositional matrix adjust.  165/167(99%)    166/167(99%)    0/167(0%)

Query  1    MLSSNDLFMVVLGAILLVLVCLVGYLYLKEKEFYHKMRRLEKTLDESYQENYIYSKRLKE  60
            MLSSNDLFMVVLGAILLVLVCLVGYLYLKEKEFYHKMRRLEKTLDESYQENYIYSKRLKE
Sbjct  1    MLSSNDLFMVVLGAILLVLVCLVGYLYLKEKEFYHKMRRLEKTLDESYQENYIYSKRLKE  60

Query  61   LEGRLEGISLEKSAKEDSSLRTTLSHLYNQLQEIQKSMDKERDYLEEKIIILENKFKDMG  120
            LEGRLEGISLEKS KEDSSLRTTLSHLYNQLQEIQKSMDKERDYLEEKIIILENKFKDMG
Sbjct  61   LEGRLEGISLEKSTKEDSSLRTTLSHLYNQLQEIQKSMDKERDYLEEKIIILENKFKDMG  120

Query  121  HYAANDEINEKQVLKMYQEGYSVDSISKEFKVSKGEVEFVLNMAGLK  167
            HYAA+DEINEKQVLKMYQEGYSVDSISKEFKVSKGEVEFVLNMAGLK
Sbjct  121  HYAASDEINEKQVLKMYQEGYSVDSISKEFKVSKGEVEFVLNMAGLK  167
```

Figure 5 – BLAST comparing the sequences of gene 0048 and a helix-turn-helix domain-containing protein, showing a very strong correlation.

Conclusion

GENI-ACT had not predicted specific products for any of these four genes, but after completing the GENI-ACT modules, likely products became evident.

Gene Locus	GENI-ACT Gene Products	Proposed Annotation
HMPREF0462_ 0048	hypothetical protein	helix-turn-helix domain-containing protein
HMPREF0462_ 0052	hypothetical protein	adenine-specific DNA methyltransferase
HMPREF0462_ 0055	hypothetical protein	type II restriction endonuclease
HMPREF0462_ 0057	hypothetical protein	periplasmic protein

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. Smoot, D. 1997. How does Helicobacter pylori cause mucosal damage? Direct mechanisms. http://www.gastrojournal.org/article/S0016-5085(97)80008-X/abstract
3. New England BioLabs. 2017. Everything You Ever Wanted to Know About Type II Restriction Enzymes. https://www.neb.com/tools-and-resources/feature-articles/everything-you-ever-wanted-to-know-about-type-ii-restriction-enzymes
4. Brennan, R., and Matthews, B. 1989. The Helix-Turn-Helix DNA Binding Motif. The Journal of Biological Chemistry 264:4 1903-1906

Acknowledgments

Rama Dey-Rao, Ph.D. University at Buffalo
Stephen T. Koury, Ph.D. University at Buffalo
Supported by an NIH Science Education Partnership (SEPA) Award - R25ODO10536

www.buffalo.edu