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 *Campylobacter jejuni* Genome

DNA Coordinates: 24542..25420, 32513..34144, 30265..31650

Niles Gonzalez, Kayla Harwell, Rafia Maryum, Nusrat Khan, Christopher Murphy, Nayema Chowdhury, Nabiha Ukaj, Aaron Ettestad, Dana Pryor Moncrieffe

Frederick Law Olmsted Science Club and The Western New York Genetics in Research and Health Care Partnership

Abstract

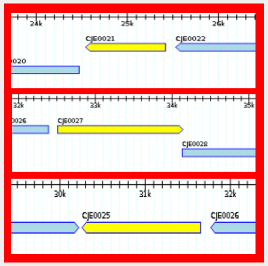
Three different genes from the microorganism *Campylobacter Jejuni* (CJE0021, CJE0025, and CJE0027) were annotated using the collaborative genome annotation website GENI-ACT (Genomics Education National Initiative - Annotation Collaboration Toolkit). The Genbank proposed a gene product name for each one of these genes. Each gene was analyzed in terms of the general genomic information, amino acid sequence-based similarity data, structure-based evidence, cellular localization data, potential alternative open reading frames, and enzymatic function. While the Genbank proposed that our genes were very similar, they also had some variations in certain aspects of the research, such as in the Structured Based Evidence module. The gene products are a fumarylacetoacetate hydrolase family protein, a CTP Synthase, and a sodium/dicarboxylate symporter. The genes have been researched but there is still more to discover, as not all of the modules were completed.

Introduction

Campylobacter Jejuni is a gram negative, helical shaped bacterium most commonly related to food poisoning. The *Campylobacter* genome is most commonly obtained when poultry is eaten raw, but it can be obtained through almost any other raw meats.The genome is most commonly related to bacterial enteritis in the United States. *Campylobacter jejuni* was first identified as a bacterial pathogen in 1973. The pathogen is not just known to react in a human's immune system and can lead to illness in cats, dogs, and other animals (Altekruse et al., 2010).

Campylobacter Jejuni was identified as bacterial caused infection and was first examined in 1886 by Escherich in a chicken's diarrhea. The genome was further examined and was first researched in humans in 1957 where we were informed of the bacteria which causes the illness. Guillain-Barre Syndrome, which is known to carry some of the *Campylobacter Jejuni* genes, still has a lot of unknown questions which could be answered with more information about the genome. For example, we might be able to answer the unknown question of why this syndrome hits some but not others.

Our team annotated three different gene sequences from the *Campylobacter Jejuni* genome. After analyzing the sequences, we have confirmed that the gene products: a fumarylacetoacetate hydrolase family protein, a sodium/dicarboxylate symporter, and a CTP Synthase were named correctly by the GENI-ACT.



The graph to the left provides a visual image of the DNA coordinates of our gene sequences. CJE0021 starts at about 24500 and ends at around 25400. CJE0025 starts at around 30300 and ends around 31600. CJE0027 starts at around 32500 and ends at around 34000.

Methods and Materials

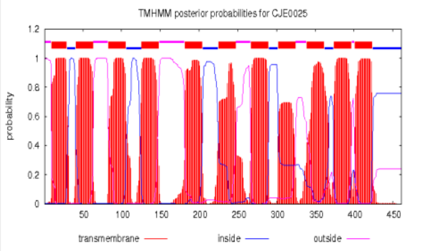
We used modules from GENI-ACT (http://www.geni-act.org/) in order to conduct our annotation of the *Campylobacter Jejuni* genome. The modules we completed are described below.

Modules	***Activities***	***Questions***
1: Basic Information	DNA Coordinates DNA Sequence Protein Sequence	What is this protein's DNA sequence? What is the gene's DNA sequence? What are the DNA coordinates of the gene?
2: Sequence-Based Similarity Data	Blast CDD T-Coffee WebLogo	Are there similarities that are shared between our genes and other genes in Genbank?
3: Structure Based Evidence	TIGRFAM PFAM PDB	What are the functioning domains in our genes?
4: Cellular Localization Data	Gram Stain TMHMM SignalP LipoP PSORT-B Phobius	What is the microbe's gram stain? What is the location of the protein?
5: Alternative Open Reading Frame	IMG Sequence Viewer for Alternate ORF Search	Are the protein's coordinates called correctly by the computer?
6: Enzymatic Function	KEGG MetaCyc	What process do our proteins take part in?

Results

Campylobacter Jejuni 0025:

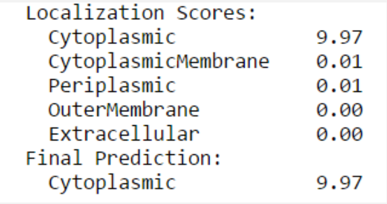
The initial proposed product of this gene by GENI-ACT was a sodium/dicarboxylate symporter. The proposal of this gene product was supported by both the first BLAST hit using the amino acid sequence of the gene in the nr database and the SwissProt database. Given these results, it is reasonable to conclude that the gene is most likely a sodium/dicarboxylate symporter, as suggested by GENI-ACT. This gene's predicted DNA coordinates were 30265..31650. The gene's DNA sequence length was predicted to be 1386 and its protein sequence length was predicted to be 461 amino acids.



This image shows the number of transmembrane helices that can be found in the protein. As shown in the TMHMM results, there are 11 transmembrane helices that are predicted to be found within the protein.

Campylobacter Jejuni 0021:

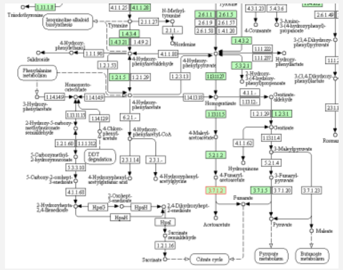
The initial proposed product of this gene by GENI-ACT was a fumarylacetoacetate hydrolase family protein. This gene product proposal was supported by the first blast hit for the amino acid sequence in the non-redundant protein sequences database. According to the LipoP and PSORT-B predictions, it is likely that this gene's protein is located in the cytoplasm.



```
Localization Scores:
  Cytoplasmic                9.97
  CytoplasmicMembrane        0.01
  Periplasmic                0.01
  OuterMembrane              0.00
  Extracellular              0.00
Final Prediction:
  Cytoplasmic                9.97
```

According to the PSORTB prediction, the protein location for CJE0021 is most likely to be cytoplasmic.

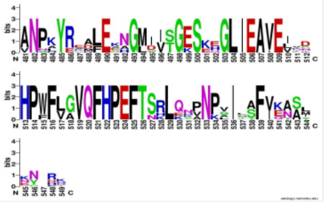
This gene's KEGG Pathway ID is hsa00350: Tyrosine Metabolism, which is found in homo sapiens (humans).



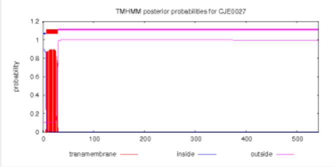
Pathway map for hsa00350

Campylobacter Jejuni 0027:

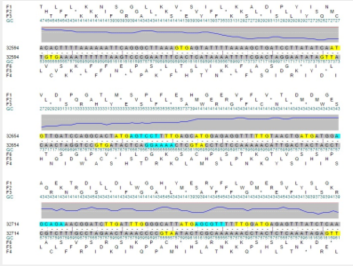
The initial proposed product of these genes by GENI-ACT was a CTP Synthase. This gene product proposal was supported by the top BLAST hits for the amino acid sequences from curated Swiss-Prot database, the cytoplasmic location of the amino acid sequences, and structural evidence. This proposal is also supported by the IMG sequence viewer, which determined the protein sequence to be correct. As such the proposed annotation is a CTP Synthase.



The above image is from the WebLogo representation of the sequences found among different organisms. This demonstrates that the gene is well conserved across different organisms.



The chart above shows the TMHMM graph, which predicts one transmembrane helix



According to the IMG Sequence Viewer, since the predicted Shine-Dalgarno region is upstream from the predicted start codon, the protein sequence was most likely called correctly

Conclusion

We have concluded that the genes appear to have been correctly annotated by the computer database and GENI-ACT, since the GENI-ACT proposed gene product did not differ at all from our proposed gene annotation for each of the genes in the group.

Gene Locus	***Geni-Act Gene Products***	***Proposed Annotation***
CJE0021	fumarylacetoacetate hydrolase family protein	fumarylacetoacetate hydrolase family protein
CJE0025	sodium/dicarboxylate symporter	sodium/dicarboxylate symporter
CJE0027	CTP Synthase	CTP Synthase

References

Altedruse SF, Stern NJ, Fields PI, Swerdlow DL. *Campylobacter jejuni*—An Emerging Foodborne Pathogen. Emerg Infect Dis. 1999, Feb.

Acheson, David, and Ban M. Allos. *Campylobacter jejuni Infections: Update on Emerging Issues and Trends.* Oxford: Oxford University Press, 2001.

Acknowledgments

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

Dr. Stephen Koury, SUNY Buffalo, Dr. Rama Dey-Rao, SUNY Buffalo
Mrs. Pryor-Moncrieffe, Frederick Law Olmsted H.S.