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: 23564..24478, 25533..26435, 31770..32393

Lily Johnson, Mi Rasa, June Fortner, Latavia Thompson, Mantaqaa Oheen, Safayath Rafat, Conrad Kegler and Mrs. Pryor-Moncrieffe

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

Abstract

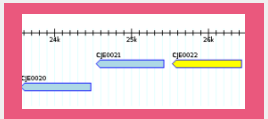
A group of three genes from the microorganism *Campylobacter jejuni* (CJE0020, CJE0022, and CJE0026) were annotated using the collaborative genome annotation website GENI-ACT (Genomics Education National Initiative - Annotation Collaboration Toolkit). Each gene was assessed in terms of the general genomic information, amino acid sequence-based similarity data, structure-based evidence, cellular localization data, and potential alternative open reading frames. 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 cytochrome c551 peroxidase,thymidylate synthase, and Pseudouridine Synthase Family Protein. The genes are well discovered but we still have much to learn about them. While we did not finish all of the GENI-ACT modules, we did learn a lot about these genes and what they do.

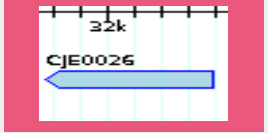
Introduction

Campylobacter jejuni is a gram negative, spiral shaped bacterium that is directly correlated food poisoning. The bacterium is most commonly transferred when raw or improperly cooked poultry is eaten. It is also found in animal feces, such as those from wombats and kangaroos. 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 affects human's immune system and can also lead to illness in cats, dogs, and other animals (Sean F. Altekruse, Norman J. Stern, Patricia I. Fields, and David L. Swerdlow, 2010).

Campylobacter jejuni was first founded as a bacteria in 1886 by Escherich. The specimen was contained within a chicken's diarrhea. The first full examination of the genome occurred in 1957, in which it was discovered to be a major contributor to illness and infection. There is a lack of information on *Campylobacter jejuni* genes. Questions on the causes of illnesses like Guillain-Barre Syndrome, which is known to carry some of the *Campylobacter jejuni* gene, will be able to be treated as information of the genome advances.

Three groups from Frederick Olmsted High School annotated three different gene sequences from the *Campylobacter jejuni* genome. After analyzing the sequences, we have determined the gene products: cytochrome c551 peroxidase, Pseudouridine Synthase Family Protein and thymidylate synthase to have been named correctly by the GENI-ACT.





The graph to the left provides a visual image of the DNA coordinates of our gene sequences. CJE0020 starts at about 23564 and ends at around 24478. CJE0022 starts at around 255333 and ends at around 26435 CJE0026 starts at 31770 and ends at 32393

Methods and Materials

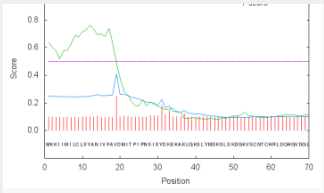
We used GENI-ACT (http://www.geni-act.org/) modules to complete our annotation of the *Campylobacter jejuni* genome. The modules we completed are described below.

Modules	***Activities***	***Questions***
1: Basic Information	DNA Coordinates DNA Sequence Protein Sequence	What are the protein sequence and DNA sequence of the gene? What is the gene's DNA coordinates?
2: Sequence-Based Similarity Data	Blast, CDD, T-Coffee, WebLogo	Does my gene share any similarities with other genes in Genbank?
3: Structure Based Evidence	TIGRfam, Pfam, PDB	Are any of the domains in my gene's protein functioning?
4: Cellular Localization Data	Gram Stain, TMHMM, SignalP, PSORT, Phobius	What is the Gram stain of the microbe? Where is the protein located?
5: Alternative Open Reading Frame	IMG Sequence Viewer for Alternate ORF Search	Was the protein sequence given by the computer correct?

Results

Campylobacter jejuni 0020:

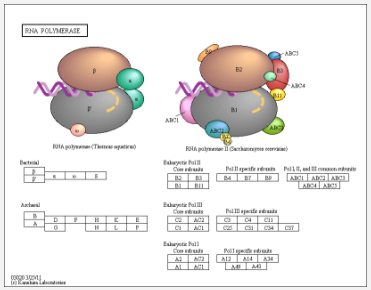
The initial proposed product of this gene by GENI-ACT was a cytochrome c551 peroxidase. The first BLAST hit using the amino acid sequence of the gene and the presence of well-curated functional domains within the amino acid sequence supported this gene product proposal. Based on all the Cellular Localization data, it was determined that the gene is strongly predicted to be a single peptide that not located in the cell membrane. Given these results, the gene product is most likely cytochrome c551 peroxidase.



This image shows the probability of a signal peptide present in the membrane of the gene which is 0.508. The graph shows a cleavage point between 18 and 19 showing the single peptide. Although not all results showed a singnal peptide present, the transmembrane topology graph showed no cleavage sites and a signal peptide was not present. This showed that the singnal peptide is most likely close to the cytoplasmic membrane.

Campylobacter Jejuni 0022:

The initial proposed product of this gene by GENI-ACT was a Pseudouridine Synthase 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. This gene product proposal was also supported by the KEGG pathway ID search. The structure of our proposed gene shows that it is not an enzyme. The KEGG results also predicted that the gene may carry "Treacher Collins Syndrome". Given the results we believe the Protein was correctly predicted as a Pseudouridine Synthase Family Protein.



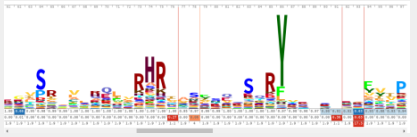
As shown by the KEGG website, this organism is not perfect, as it carries a disease called "Treacher Collins syndrome."

This image shows that the Query and Subject of the Protein Data Bank search have a few alignments. This is shown by the middle section of the picture.

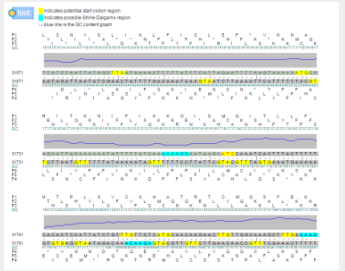
Campylobacter Jejuni 0026:

The initial proposed product of these genes by GENI-ACT was a thymidylate synthase. This gene product proposal was supported by the top BLAST hits for the amino acid sequences from curated Swiss-Prot database, 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 thymidylate synthase, flavin-dependent.

The image above shows the extent to which our protein sequence is aligning with the ones from the BLAST databases. This supports the conclusion that the protein product is thymidylate synthase, flavin-dependent.



The graphic above is an HMM logo, which shows the conservation of the protein family of our protein sequence in a visual manner. The larger the letter, the more conserved it is across different organisms. This image shows that since this family is highly conserved, our protein conclusion is true.



According to the IMG Sequence Viewer, since the predicted Shine-Dalgarno region is upstream from the predicted start codon (in red), the protein sequence was most likely called correctly, and this supports our conclusion.

Conclusion

The GENI-ACT proposed gene product did not change at all from the proposed gene annotation for each of the genes in the group. The genes appear to be correctly annotated by the computer database and GENI-ACT.

Gene Locus	Geni-Act Gene Products	Proposed Annotation
CJE0020	cytochrome c551 peroxidase	cytochrome c551 peroxidase
CJE0022	Pseudouridine Synthase Family Protein	Pseudouridine Synthase Family Protein
CJE0026	thymidylate synthase	thymidylate synthase

References

Altedruse SF, Stern NJ, Fields PI, Swerdlow DL. *Campylobacter jejuni*—An Emerging Foodborne Pathogen. Emerg Infect Dis. 1999, Feb, [cited May 9 2017].

Acheson, David, and Ban M. Allos. *Campylobacter jejuni Infections: Update on Emerging Issues and Trends*. Oxford: Oxford University Press, 2001. Web. 9 May 2017.

Acknowledgments

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

Dr. Stephen Koury and Dr. Rama Dey-Rao