Annotation of *the Clostridium botulinum* Genome at Locus Tags CLJ_0136, CLJ_0138, CLJ_0144 and CLJ_0147

Max Magee, Isabel Ross, Shira Silverstein, Anastasia Van Ark, Gleb Morshuk, Brian Cowburn, and Christopher Benedict
Honeoye Falls- Lima High School, Honeoye Falls, New York 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

A group of genes from the bacteria *Clostridium botulinum* were annotated using the website GENI-act. The function and location of each gene was predicted through analysis of the structure based evidence and genomic amino acid sequences of the genes. The function was predicted based on amino acid sequence similarity comparisons between the test gene and genes whose functions were already known. Additionally, determining the location of the gene, such as the gene being located in the membrane or cytoplasm, helped determine function. The BLAST program used in initial stages of analysis also helped determine how well conserved the genes' amino acid sequences were.

Introduction

Clostridium botulinum is a gram-positive bacterium well known for its ability to form endospores. Endospores act as an internal armor of the bacteria, protecting only the most vital parts needed for it to survive. It is a rod-shaped, non-encapsulated anaerobic bacterium which is commonly found in soil and marine environments. However, there have been instances where it originated from the gastrointestinal tracts of wildlife (1). The bacterium is motile and can move quickly due to its peritrichous flagella. It was successfully isolated in 1895 by Emile Pierre van Ermengem, a Professor of bacteriology at the University of Ghent (2).

Clostridium botulinum produces an extremely deadly neurotoxin capable of killing its host. This bacterium starts its destruction by blocking the host's neurotransmitters responsible for motor control, which can leave the host in a paralyzed state. However, in certain situations, this toxin can be advantageous. For example, if a person is suffering from muscle spasms, he can get a dose of Botulinum as treatment. This is done by giving a localized injection of Botulinum to the affected muscle(s), which coincidentally is released by *Clostridium botulinum.* The neurotoxin is also useful for people with muscle cramps through "reducing presynaptic cholinergic stimulation of motor nerve terminals and by impairing the input/output function of intrafusal and extrafusal motor end plates" (3). *Clostridium botulinum* is vital to modern science due to both its healing and destructive abilities. The more we study this bacterium, the more likely we are to find other relating bacteria which can hold other medicinal uses.

Methods

Modules of the GENI-ACT (http://www.geni-act.org/) were used to complete *Clostridium botulinum* genome annotation.

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- Structure-Based Evidence	TIGRfam, Pfam, PDB	Are there functional domains in my protein?
Module 4- Cellular Localization Data	Gram Stain, TMHMM, SignalP, PSORT, Phobius	Is my protein in the cytoplasm, secreted or embedded in the membrane?

Results

CLJ_0136 (completed through module 3)

This gene of *Clostridium botulinum* is most likely a hemagglutinin component, as suggested by BLAST results and supported by WebLogo. There were no COG results or CDD results which suggest this gene has not been studied much or is unique to Clostridium botulinum. The Weblogo showed mostly a lack of conservation until there was a section that had too many similarities to dismiss (Figure 2). These similarities suggest that through evolution, that protein sequence was preserved. CLJ_0136 is likely to contribute to the production of Hemagglutinin which is a protein that causes red blood cells to clump and stick together in an action called agglutination

CLJ_0138 (completed through module 4)

The results of the gene sequencing programs suggest that gene CLJ-0138 is unique. The BLAST program yielded very few results, showing that the amino acid sequence of the gene was similar only to very few genes which had already been studied or logged. The results of the T-Ccoffee program had a similar issue: very few significant results were found which were not of the same genus and species as CLJ-0138. The web logo revealed that large portions of the gene were very well conserved, having many amino acids in common and in the same sequence. Given that many of the comparative genes used in the web logo were phages or toxins, it was strongly suggested by the level of conservation present in the web logo that the CLJ-0138 gene was of a similar nature. A Pfam search yielded highly significant results with the family clenterotox, which are clostridium enterotoxins. Of these toxins, *Clostridium botulinum* produces one of the most potent. The data did not predict any transmembrane helices or signaling peptides. Additionally, although the results were tentative, the Phobius program suggested that the gene is cytoplasmic, which is consistent with what would be expected given the lack of transmembrane helices or signaling peptides. In all, it is highly likely that gene CLJ-0138 is a potent enterotoxin located in the cytoplasm of the bacteria.

CLJ_0144 (Completed through module 3)

CLJ_0144 has the coordinates of 137264 to 137512 on the reverse strand. The initial product proposed by GENI-ACT was a sporulation transcriptional regulator. This proposal is supported by the nr and Swiss-Prot database using the amino acid sequence as these BLAST programs show similarities between *Clostridium botulinum*, Stage III sporulation protein D; AltName: Full=14 kDa transcription factor, and sporulation transcriptional regulator SpoIIID [Clostridium lundense]. There were no significant COG hits found in the CDD search results. The multiple sequence alignment T-COFFEE results show that there are many similarities in the amino acid sequence of Clostridium, Alkaliphilus, Caloramator and eight more species all with the same or similar function of sporulation transcriptional regulator. The web logo revealed that significant portions of the gene are highly conserved. To definitively determine the function of Clostridium botulinum_0144 more extensive research would need to be performed.

CLJ_0147 (Completed through module 4)

The gene sequencing of CLJ_0147 suggests that this particular gene has very seldom been studied. The nr BLAST program resulted in multiple significant matches to the amino acid sequence of this gene, however all of these matches were hypothetical *Clostridium botulinum* proteins meaning there was no actual information indicating the function of the gene that was annotated. The Swiss-Prot BLAST results were resulted in even less information due to the smaller database. There were only two Swiss-Prot results and both had a significantly high E-Value meaning their match to the Clostridium Gene is insignificant. The lack of results in both BLAST runs also contained no CDD result. For the T-Coffee program there were no results under an E-value threshold of 0.001 that were not of the clostridium genus and they were all hypothetical proteins resulting in a T-Coffee that suggests that these results are very similar to the amino acid sequence of the gene being annotated. Therefore overall the WebLogo shows consistently high conservation. This information is irrelevant because the significant proteins being compared were of the same genus and many were of the same species. The Pfam results were also very insignificant, seen in the pairwise alignment of the top hit. (Figure 5) However highly improbable, the gene could possibly be found to the N-terminus of bacterial signal peptidases of the S49 family or be a regulator of G protein signaling based on Pfam results however these matches are very insignificant. In Conclusion, this gene has either not been seen in any of the other species sequenced thus far or that it may not really be a functional gene at all. According to Dr. Stephen T. Koury of the University at Buffalo these results would further be the basis for wet lab experiments to determine of the gene is expressed as a mRNA in Clostridium and whether the predicted protein sequence can be detected.

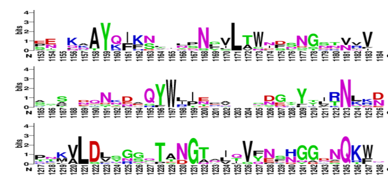

Figure 2- CLJ_0136 WebLogo which shows at first, there are not many similarities, but at around position 961, more similarities are shown until position 1266. Prior to 961 and after 1266, there are little or no letters, showing a lack of conservation.

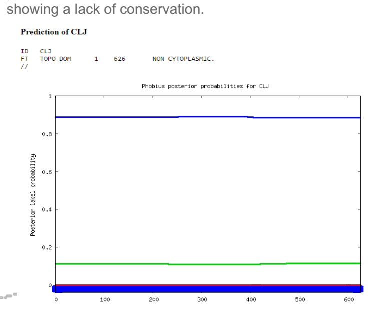

Figure 3 – CLJ_0138 this graph shows that the most likely location of enterotoxin gene is in the cytoplasm

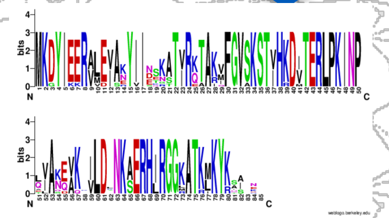

Figure 4 – CLJ_0144 Positions 1-85 show moderate conservation. Positions 83 and 85 show little conservation with these positions having no letters at all. Positions 30-50 and 63-80 show a high degree of conservation, with many capital letters and wide letters.

```
Insignificant match:
                .........*.........*.........*.........*.......
Model       41  keqeelkkkakeIynkylapsspkel.nidkklreelkknlekaqps 86
                ++++++++k+k+I+ +y+++++ p+e  +i+++l e+++++++ ++p+
Query       31  TTKRNRSNKIKDILIEYFNEKYPEEFnEIKNRLSENMEQKITPKKPT 77
PP              678999*********************99*********9999986444
```

Figure 5- CLJ_0147 . The Pfam results are very insignificant, seen in the pairwise alignment of the top hit

Conclusion

The gene tagged as CLJ_0136 is most likely a hemagglutinin component, which is a protein that causes red blood cells to clump and stick together in an action called agglutination. The gene tagged as CLJ_0138 is very unique however it is highly likely that gene CLJ-0138 is a potent enterotoxin located in the cytoplasm of the bacteria. Evidence points to the gene locus CLJ_0144 as a transcription regulator. The gene tagged CLJ_0147 has either not been seen in any of the other species sequenced thus far or that it may not really be a functional gene at all.

References

1. "CLOSTRIDIUM BOTULINUM, TYPE A – Los Angeles, California." Morbidity and Mortality 17.48 (1968): 446. Web. 13 May 2017.
2. Erbguth, F. J. "Historical Notes on Botulism, Clostridium Botulinum, Botulinum Toxin, and the Idea of the Therapeutic Use of the Toxin." Movement Disorders : Official Journal of the Movement Disorder Society. U.S. National Library of Medicine, 19 Mar. 2004. Web. 13 May 2017.
3. Bertolasi, L., A. Priori, G. Tomelleri, L. G. Bongiovanni, E. Fincati, A. Simonati, D. De, and N. Rizzuto. "Botulinum Toxin Treatment of Muscle Cramps: A Clinical and Neurophysiological Study." Annals of Neurology. U.S. National Library of
Medicine, Feb. 1997. Web. 13 May 2017.
4. Clark, Marler. "Clostridium Botulinum (Botulism)." Clostridium Botulinum (Botulism) Food Poisoning. N.p., 2010. Web. 13 May 2017.
5Grazko, M. A., K. B. Polo, and B. Jabbari. "Botulinum Toxin A for Spasticity, Muscle Spasms, and Rigidity." Neurology. U.S. National Library of Medicine, Apr. 1995. Web.

Acknowledgments

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

www.buffalo.edu