Annotation of the *Clostridium botulinum* Genome at Locus Tags CBO0003, CBO0004, CBO0007, CBO0011 and CBO0013

Aidan Cap, Lauren Cody, Fiona Gaynor, Julia Lewis, Marcis Scroger, Moshe Seluanov, Rachel VanderVen, and Chris Benedict
Honeoye Falls-Lima High School, Honeoye Falls, NY 14472 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 five genes from the microorganism *Clostridium botulinum* (CBO0003, CBO0004, CBO0007, CBO0011, & CBO0013) were annotated through GENI-ACT, an online gene annotation program. GENI-ACT expressed the genes in terms of their basic information, sequence- based similarly data, cellular localization data and structure- based evidence. To find the product of the gene, several programs like TIGRfam, Pfam, PDB, Blast, WebLogo, CDD and T-Coffee were used. Compared to GenBank, the final results for the protein sequences were similar.

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 (CLOSTRIDIUM BOTULINUM). 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 (Erbguth).

Unfortunately, *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" (Bertolasi 1997). *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.

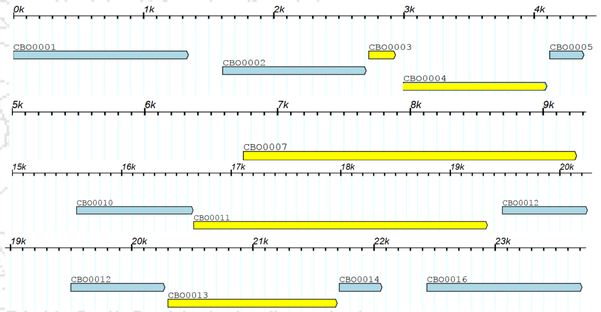

Figure 1. The locus tags and relative position of the genes under investigation in this research

Methods

Modules of the GENI-ACT (http://www.geni-act.org/) were used to complete *Clostridium botulinum* genome annotation. The modules used 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 5- Structure-Based Evidence	TIGRfam, Pfam, PDB	Are there functional domains in my protein?

Results

CBO0003:
The initial proposed product of this gene by GENI-ACT was an S4 domain-containing protein, which assists in initiating the assembly and binding of rRNA and stabilizes its tertiary structure. This gene product proposal was supported by the top BLAST hits for the amino acid sequence, the absence of transmembrane helices or signal peptides, and the cellular location of the amino acid sequence, which was predicted to be cytoplasmic. Therefore, the proposed annotation is an S4 RNA binding protein.

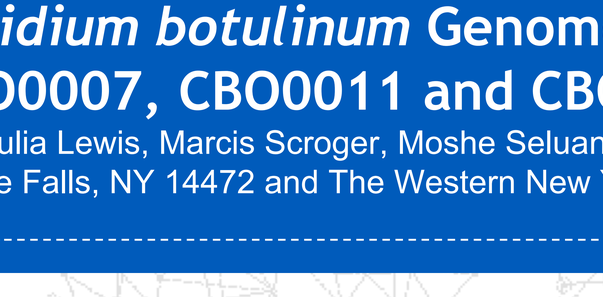

Figure 2 – SwissProt results for CBO0003, neither of which possessed an E-value below the cutoff of 0.001, indicating little prior research existed pertaining to the gene.

CBO0004:
The initial proposed product of this gene by GENI-ACT was a recombination product F. This gene product proposal was supported by the top BLAST hits from both swissprot as well as the nr database. Our top hit from the BLAST result was "~~Recombinational DNA repair ATPase RecF~~ [Replication, recombination and repair]". The protein from *C. botulinum* is structurally related to other proteins in other bacteria that do that same function. The protein is most likely cytoplasmic. The TMHMM graph showed that this protein has no helices. Therefore, the protein does not intersect with the membrane. The psortB program suggested high extracellular probability, indicating that the protein is secreted, but the other tools in the module all suggested cytoplasmic, so we treated this result as anomalous.

CBO0007:
The initial proposition of the product of the gene by GENI-ACT had been a DNA gyrase subunit A, which negatively supercoils closed, circular, double-stranded DNA. Such a gene product proposition was highly supported by SwissProt, which had depicted the gene product in a variety of organisms. Along with that, the TIGRfam and the CDD had supported the gene product to be of the DNA gyrase subunit A. With the lack of transmembrane helices and a PSORT-B value of 9.97n for cytoplasmic, the product would be likely located in the cytoplasm. Fitting the various parts together, the gene annotation would likely be of a DNA gyrase subunit A.

CBO0011:
The initial proposed product of these genes by GENI-ACT was a hypothetical protein. This gene product proposal was supported by the top BLAST hits for the amino acid sequences being under fifty percent identical leading to inconclusive matches. This is also supported through the inconsistencies among different collected results due to the protein not being in certain databases due to the lack of prior research of the protein. Not only inconsistencies, but lack of results seen through CBO0011 only coming up with one hit in the CDD search, that was also a hypothetical protein with an unknown function. CBO0011 did contain transmembrane helices which could indicate that it is a transmembrane protein. In conclusion, the proposed annotation is a hypothetical protein. with an unknown function.

```
# CBO011 Length: 893
# CBO011 Number of predicted TMHs:  7
# CBO011 Exp number of AAs in TMHs: 155.5923
# CBO011 Exp number, first 60 AAs:  35.47609
# CBO011 Total prob of N-in:        0.99999
# CBO011 POSSIBLE N-term signal sequence
CBO011  TMHMM2.0        inside     1   6
CBO011  TMHMM2.0        TMhelix    7   29
CBO011  TMHMM2.0        outside    30  48
CBO011  TMHMM2.0        TMhelix    49  68
CBO011  TMHMM2.0        inside     69  93
CBO011  TMHMM2.0        TMhelix    94  116
CBO011  TMHMM2.0        outside    117 150
CBO011  TMHMM2.0        TMhelix    151 173
CBO011  TMHMM2.0        inside     174 201
CBO011  TMHMM2.0        TMhelix    202 224
CBO011  TMHMM2.0        outside    225 243
CBO011  TMHMM2.0        TMhelix    244 266
CBO011  TMHMM2.0        inside     267 272
CBO011  TMHMM2.0        TMhelix    273 295
CBO011  TMHMM2.0        outside    296 893
```

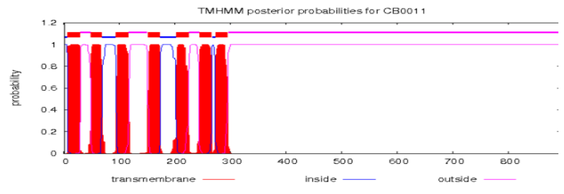

Figure 3 – The transmembrane topology graph from CBO0011 indicating that the protein has transmembrane helices. This means that the protein is in the membrane of the cell.

CBO0013:
The initial proposed product of this gene by GENI-ACT was a dehydrogenase. This gene product proposal was supported by the top BLAST hits from SwissProt, as well as the nr database, showing this gene exists in similar forms outside the world of *C. botulinum*. These both gave results with an e-value under 0.001. The WebLogo results showed large conservation sporadically-almost in thirds-throughout the image, specifically with the color green and letters G,Y,C, and T. This shows that CBO0013 contains many polar amino acids.

Conclusion

The GENI-ACT proposed gene product did not differ significantly from the proposed gene annotation for each of the genes in the group and as such, the genes appear to be correctly annotated by the computer database. Important note: research presented represents four GENI-ACT modules.

Gene Locus	GENI-ACT Gene Products:	Proposed Annotation:
CBO0003	S4 domain-containing protein	S4 domain-containing protein
CBO0004	Recombination product F	Recombination product F
CBO0007	DNA gyrase subunit A	DNA gyrase subunit A
CBO0011	hypothetical protein	hypothetical protein
CBO0013	dehydrogenase	

References

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.

*Clark, Marler. "Clostridium Botulinum (Botulism)." Clostridium Botulinum (Botulism) Food Poisoning. N.p., 2010. Web. 13 May 2017.

* "CLOSTRIDIUM BOTULINUM, TYPE A – Los Angeles, California." Morbidity and Mortality 17.48 (1968): 446. Web. 13 May 2017.

* 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.

* Grazko, 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