# Annotation of the Clostridium Botulinum Genome At Locus Tags CLJ_0135, CLJ_0137, CLJ_0143, CLJ_0152

Ana Youngblut, Margaret Hoare, Erica Sayer, Miranda Pasquarette, Gleb Morshuk, Brian Cowburn and Christopher Benedict
Honeoye Falls- Lima High School and Western New York Genetics in Research and Health Care Partnership

SEPA Science Education Partnership Award
Supported by the National Institutes of Health

## Abstract

A group of four genes from the microorganism *Clostridium botulinum* (CLJ_0135, CLJ_0137, CLJ_0143, and CLJ_0152) were annotated through GENI-ACT, an online modular gene annotation program (http://geni-act.org/). 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 results for the protein sequences was consistent throughout the data.

## 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).

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" (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

| GEN-ACT 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_0135:** (assessed with Modules 1-3)
The initial proposed product of this gene by GENI-ACT was a transcriptional regulator BotR. The BLAST hits for the amino acid sequence was inconclusive , due to the fact that no results were shown in the UniProtKB/Swiss-Prot database. In the Non-redundant protein sequences database, CLJ_0135 matched closely with a transcription activator in *Clostridium tetani*. CLJ_0135 is a protein with no signal peptide is predicted nor are transmembrane helices predicted, which leads to the conclusion that it is cytoplasmic .

**CLJ_0137:** (annotated with Modules 1, 2 and 4)
The initial proposed product of this gene by GENI-ACT was a hemagglutinin component HA17. This gene product proposal was supported by the top BLAST hits for the amino acid sequence, the presence of well-curated functional domains within the amino sequences, cellular location of the amino acid sequences, and the enzymatic function of the amino acid sequences. As such, the proposed annotation is a lysogenic phage.

**CLJ_0143:** (annotated with Modules 1-4)
The initial proposed product of this gene by GENI-ACT was a DNA Directed RNA Polymerase. This gene product proposal was reinforced by the top BLAST hits for the amino acid sequence, sequence based similarity, the cellular location of the amino acid sequence, and the structure based evidence of the amino acid sequence. There fore, the proposed annotation is a DNA Directed RNA Polymerase.

**CLJ_0152:** (annotated with Modules 1- 4)
The initial proposed product of this gene was a phage shock protein A. This gene product proposal was supported by the top BLAST hits for the amino acid sequence, sequence based similarity, and the location in the gene within the cell. The BLAST results show similarity to 657 similar genes evolved from a common ancestor. The THMM results also show no signs of transmembrane helices within the gene, indicating that it works within the cell.

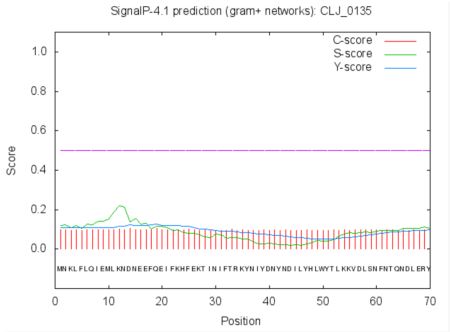
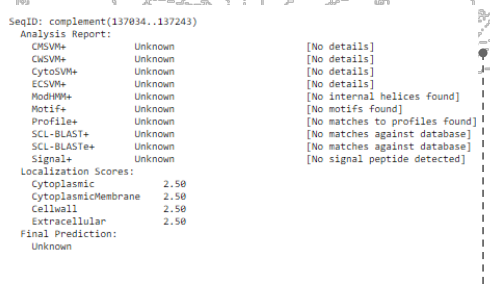



Figure 1 The SignalP-4.1 prediction for CLJ_0135, which shows that no signal peptide is predicted.




PSORT-B results of CLJ_0143. The results showed all the localization scores as the same score of 2.5. When that happens it means PSORT-B cannot find its location.

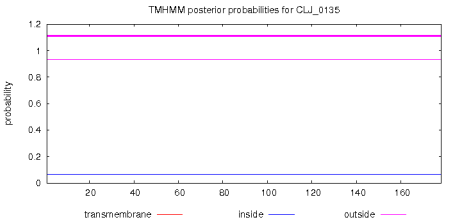



Figure 3 – CLJ_0135 TMHMM posterior probabilities show that no transmembrane helices are predicted, which supports the final prediction that the protein does not leave the cell. .

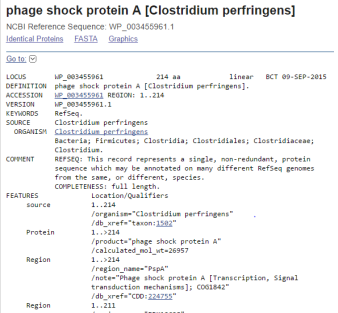



Figure 4 – Results from gene CLJ_0152 BLAST Alignment of the top hit and the query sequence, showing how it is a Phage Shock Protein A.

Figure 5 - WebLogo results from CLJ_0137. These results show The sequence being well conserved form 3-26

## 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.

| Gene Locus | GENI-ACT Product | Proposed Annotation |
|---|---|---|
| 0135 | Transcriptional regulator BotR | Insufficient Data to Propose |
| 0137 | Hemagglutinin component HA17 | Lysogenic Phage Protein |
| 0143 | DNA directed RNA polymerase | DNA directed RNA polymerase |
| 0152 | Phage Shock A | Phage Shock A |

## References

Sims et al. (2009). Complete genome sequence of *Kytococcus sedentarius* type strain (541T). Standards Genomic Sciences,12 - 20.
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.
5. 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