University at Buffalo *The State University of New York* | **REACHING OTHERS**

Annotation of the *Clostridium botulinum* A2 strain Kyoto at Locus Tags CLM_0114, CLM_0572, CLM_0623 and CLM_0002

Theresa Durzewski, Andrea Halsey, Mason Hinze, Elliot Schmidt, Zack Kowalewski, Laura O'Donnell and Elizabeth Winters

Attica High School and The Western New York Genetics in Research Partnership

Abstract

A group of 5 genes from the microorganism *Clostridium botulinum* were annotated using the collaborative genome annotation website GENI-ACT. The Genbank proposed gene product name for each gene was assessed in terms of the general genomic information, amino acid sequence-based similarity data, structure-based evidence from the amino acid sequence, and cellular localization data. The Genbank proposed gene product name 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.

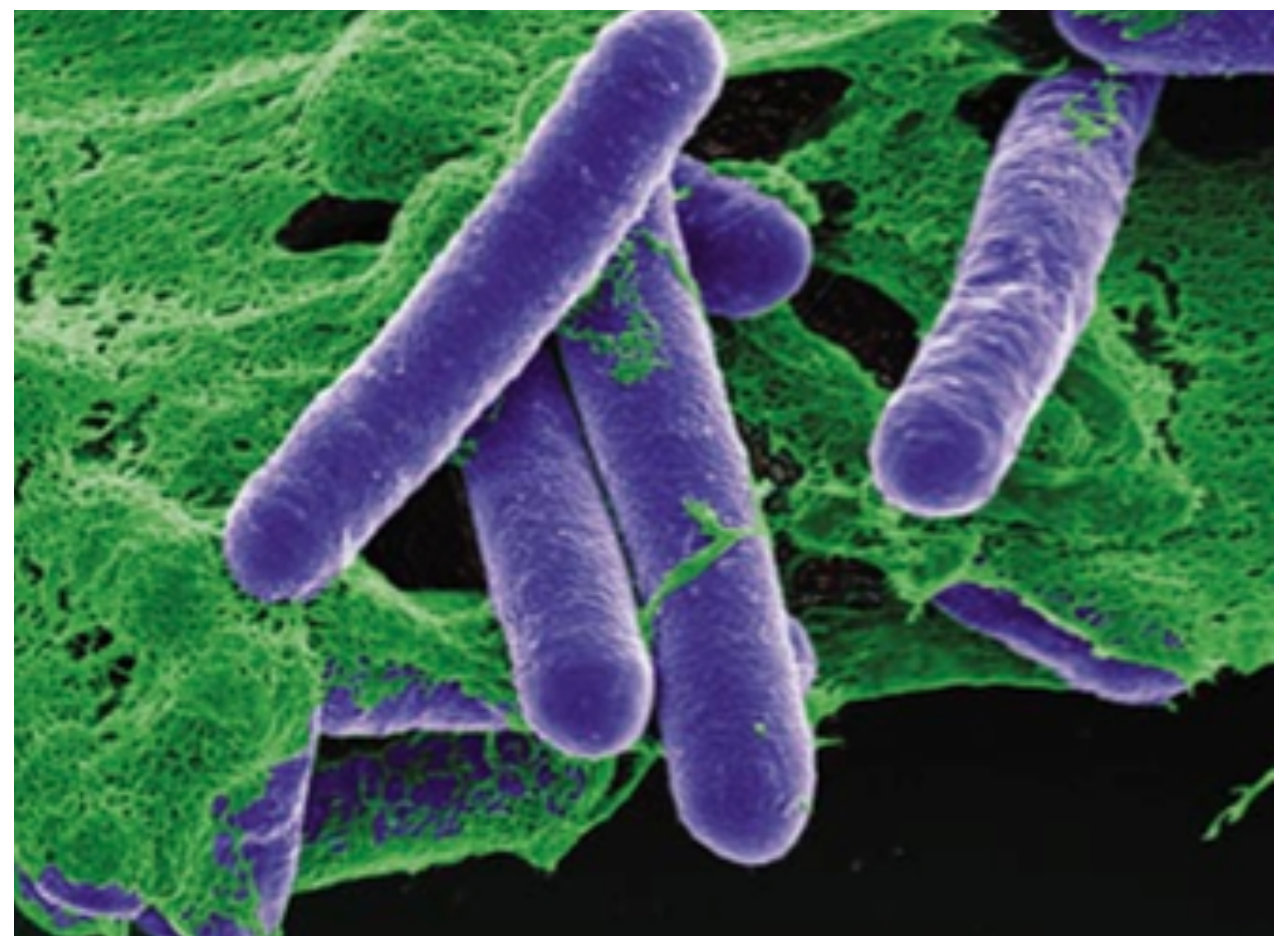
Introduction

Clostridium botulinum is a Gram-positive, rod-shaped, anaerobic, spore-forming, motile bacterium. It can produce the heat-resistant neurotoxin, botulinum. Its key parts are encased by its many protective layers and hard coating. It is able to produce endospores, mostly found in soil that can survive adverse conditions. *Clostridium botulinum* can only produce neurotoxins during sporulation, which happens only in anaerobic environments.

Clostridium botulinum is known as a foodborne illness because it can occur in preserved or home-canned foods from incorrect preservation procedures. As the bacteria grows on food, it releases its toxin, that when ingested, has a possibility to cause paralysis, maybe death.

For the most part botulinum poisoning is rare, but still dangerous. Some symptoms of botulism poisoning include, nausea, vomiting, fatigue, double vision, dizziness, difficulty swallowing, dry mouth, throat, or skin, muscle weakness, slurred speech, body aches, and in some cases, paralysis. Symptoms of botulism generally appear 12 to 72 hours after eating contaminated food. With treatment, illness lasts from 1 to 10 days. Full recovery from botulism poisoning can take weeks to months. Some people never fully recover.

Botulism poisoning is so dangerous that each case is considered a public health emergency. Studies have shown that there is a 35 to 65 percent chance of death for patients who are not treated immediately and effectively with botulism antitoxin.



Colorized Scanning Electron Micrograph of *Clostridium botulinum*.
microbewiki.kenyon.edu

Methods and Materials

Modules of the GENI-ACT (http://www.geni-act.org/) were used to complete *Clostridium botulinum* genome annotation. The modules 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 4- Alternative Open Reading Frame	IMG Sequence Viewer For Alternate ORF Search	Has the amino acid sequence of my protein been called correctly by the computer?

Results

CLM_0114		
Basic Information	**DNA Coordinates**	95018..95593
	DNA Length	576 bp
	Amino Acid Sequence Length	191 aa
Sequence Based Similarity	**COGs**	COG2715 SpmA Spore maturation protein SpmA
	TCOFFEE	Very well conserved throughout
	WEBLOGO	Very well conserved throughout
Cellular Localization	**Transmembrane Helices**	5
	Signal Peptide	none
	PSORTB Final Prediction	Cytoplasmic membrane
Structure Based Evidence	**TIGRFAMs**	No significant TIGRFAMs
	PFAMs	Nucleoside recognition
	HMM Logo Key Residues	I8, L18, P33, G53, Y68, F91, C98
	PDB	No results

Clostridium botulinum CLM_0114

COG data supports CLM_0114 to be a spore maturation protein. Maturation in spores would provide proper and fast reproduction for the bacterium. Reproduction in *Clostridium botulinum* is important to spread the illness. The BLAST database shows many orthologs exist which mean the protein is probably crucial to the survival of many different bacterial species.

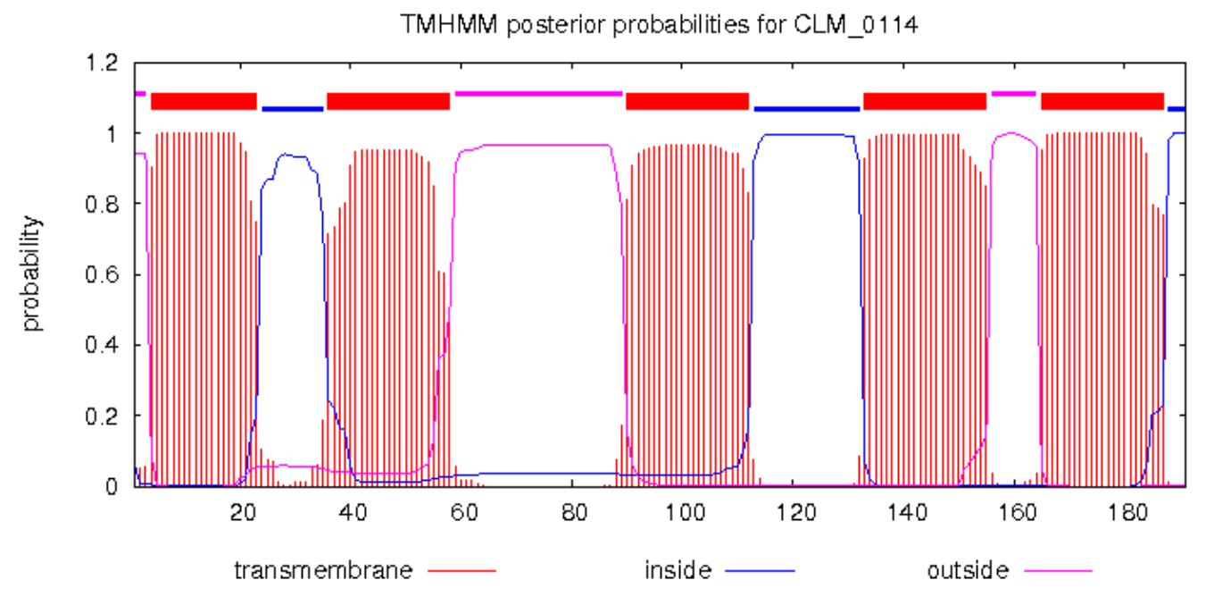


Figure I – TMHMM Data for *Clostridium botulinum* CLM_0114.
The image to the right shows the presence of five transmembrane helices, most likely present on the cytoplasmic side of the membrane in the bacteria.

CLM_0572		
Basic Information	**DNA Coordinates**	618349..618522
	DNA Length	174 bp
	Amino Acid Sequence Length	57 aa
Sequence Based Similarity	**COGs**	No significant hits
	TCOFFEE	Only one ortholog but highly conserved
	WEBLOGO	Too few orthologs
Cellular Localization	**Transmembrane Helices**	none
	Signal Peptide	Yes- between positions 44 and 45
	PSORTB Final Prediction	Extracellular
Structure Based Evidence	**TIGRFAMs**	No significant hits
	PFAMs	No significant hits
	HMM Logo Key Residues	No significant hits
	PDB	No significant hits

Clostridium botulinum CLM_0572

The BLAST database found one ortholog for CLM_0572 predicting it may be a bacteriocin protoxin related to the streptolysin S family bacteria. Research into this family indicates that this protein may be a toxin specific to this strain of *Clostridium botulinum*. Cellular localization data suggests that this is a protein containing a signal peptide and is therefore released from the cell.

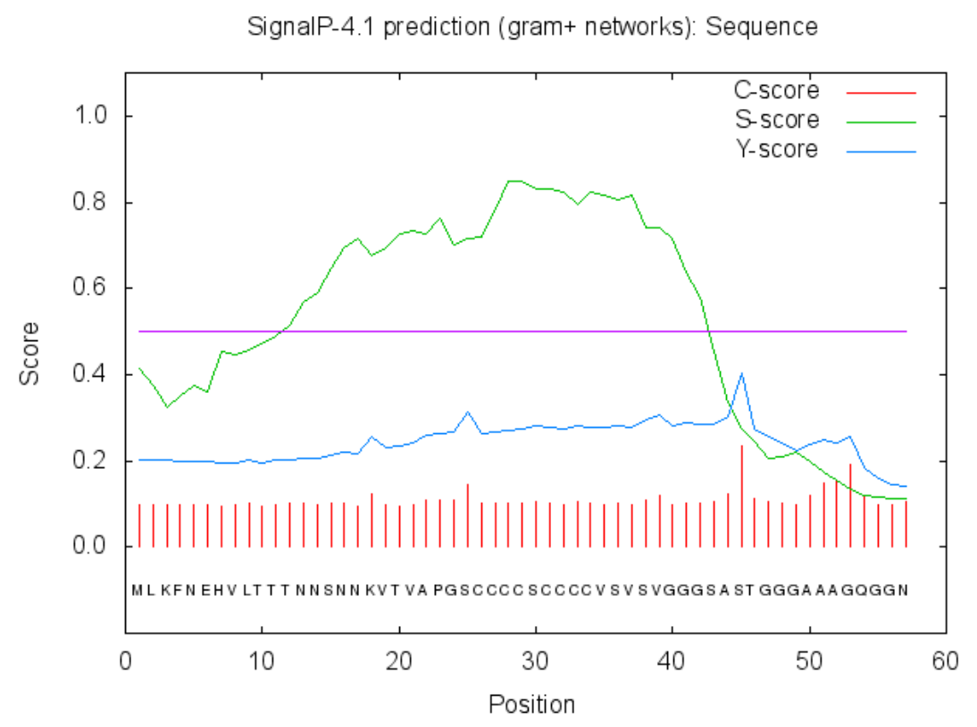


Figure II – SignalP Data for *Clostridium botulinum* CLM_0572
The image to the right shows the presence of a signal peptide indicating that this protein is released from the cell.

CLM_0623		
Basic Information	**DNA Coordinates**	674438...676948
	DNA Length	2511 bp
	Amino Acid Sequence Length	836 aa
Sequence Based Similarity	**COGs**	COG0577 SalY ABC-type antimicrobial peptide transport system, permease component [Defense mechanisms]
	TCOFFEE	conserved at beginning and end
	WEBLOGO	conserved at beginning and end
Cellular Localization	**Transmembrane Helices**	8
	Signal Peptide	none
	PSORTB Final Prediction	Cytoplasmic membrane
Structure Based Evidence	**TIGRFAMs**	No significant TIGRFAMs
	PFAMs	FtsX-like permease family
	HMM Logo Key Residues	G11
	PDB	no significant hits

Clostridium botulinum CLM_0623:

COG data supports CLM_0623 to be part of an ABC peptide transport system found in the membrane. It is suggested to be used to transport materials into and out of the cell and could more specifically be part of a defense mechanism, keeping certain toxins out of the bacterial cell. BLAST results show many orthologs with other bacteria species containing a similar protein.

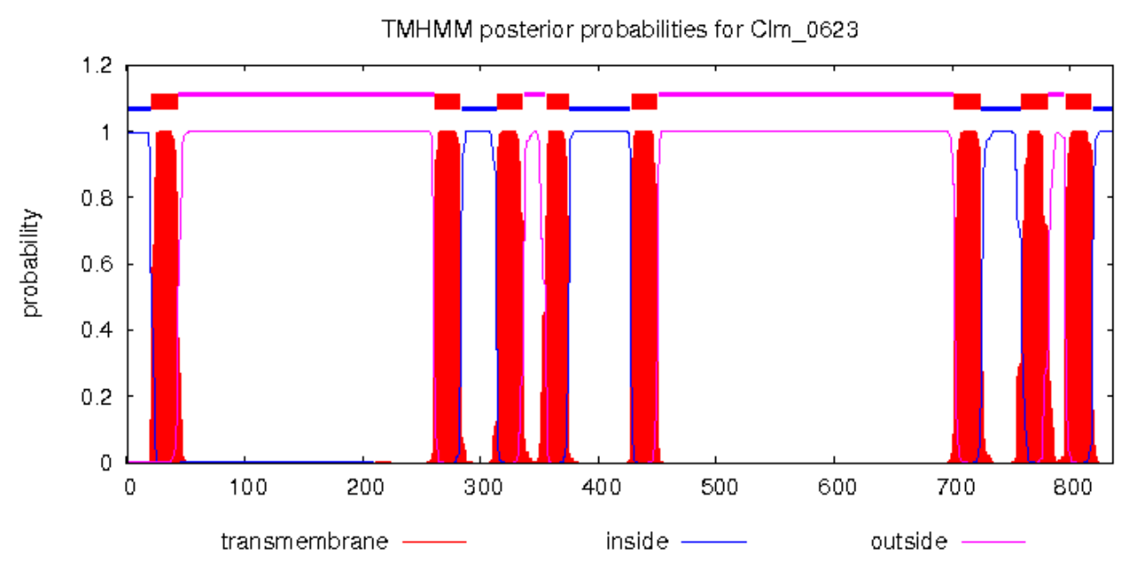


Figure III – TMHMM Data for *Clostridium botulinum* CLM_0623.
The image to the right shows the presence of eight transmembrane helices, most likely present on the cytoplasmic side of the membrane in the bacteria.

CLM_0002		
Basic Information	**DNA Coordinates**	1757..2860
	DNA Length	1104 bp
	Amino Acid Sequence Length	367 aa
Sequence Based Similarity	**COGs**	COG0592: DNA polymerase III sliding clamp (beta) subunit, PCNA homolog [Replication, recombination and repair]
	TCOFFEE	Not well conserved at the N-terminus but then sporadically conserved towards the middle and end
	WEBLOGO	Not well conserved at the N-terminus but then sporadically conserved towards the middle and end
Cellular Localization	**Transmembrane Helices**	none
	Signal Peptide	none
	PSORTB Final Prediction	Cytoplasmic
Structure Based Evidence	**TIGRFAMs**	dnan: DNA polymerase III, beta subunit
	PFAMs	DNA polymerase III beta subunit, N-termi
	HMM Logo Key Residues	R24, P28, D48, V88, F107
	PDB	Crystal structure of DNA polymerase sliding clamp from *Bacillus subtilis*

Clostridium botulinum CLM_0002:

COG, TIGRFAM, PFAM and PDB data support CLM_0002 to take part in DNA replication as a DNA polymerase III beta subunit. PSORTb final prediction suggests that it would be found in the cytoplasm of a cell because that is where DNA replication takes place. BLAST data show many orthologs for this protein since DNA replication is a common cell process in bacteria.

Figure IV – T-Coffee Data for *Clostridium botulinum* CLM_0002
The image above shows no amino acid similarities at the N-terminus of the protein among orthologs. Further down, it has more conservation but is sporadic.

Conclusion

The Geni-Act proposed gene product did not differ significantly from the proposed gene annotation for each of the genes and as such, the genes appear to be correctly annotated by the computer database.

Gene Locus	Proposed Annotation
0114	Spore maturation protein
0572	Bacteriocin protoxin
0623	ABC-type antimicrobial peptide transport system
0002	DNA polymerase III sliding clamp (beta) subunit

References

BBB - Clostridium botulinum. (2017). *Fda.gov*. Retrieved 8 March 2017, from https://www.fda.gov/food/foodborneillnesscontaminants/causesofillnessbadbugbook/ucm070000.htm

Botulism. (2017). *Healthline*. Retrieved 8 March 2017, from http://www.healthline.com/health/botulism#Overview1

Clostridium Botulinum (Botulism) food poisoning. (2017). *Foodborneillness.com*. Retrieved 8 March 2017, from http://www.foodborneillness.com/botulism_food_poisoning/

Acknowledgments

Supported by NSF ITEST Strategies Award Number 1311902