# Annotation of the Clostridium botulinum Genome at Locus Tags CLJ\_B3418-ClJ\_B3421, CLJ\_B1221-CLJ\_B1222 and CLJ\_B0527

Jessica Bowling, Analise Healy, Sarah Knoll, Hannah McCarthy, Jessica Pray, Trevor Stoddard, Chrystler Stortz and Lynn Barrett West Seneca West Senior High School 3330 Seneca Street, West Seneca NY 14224 and The Western New York Genetics in Research And Health Care Partnership

# Abstract

A group of 7 genes from the microorganism Clostridium Botulinum genome Locus Tags (CLJ B3418-CLJ B3421, CLJ B1221 - CLJ B1222 and CLJ B0527) 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, cellular localization data, and potential alternative open reading frames. 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 in the database.

# Introduction

Clostridium botulinum is a gram positive, anaerobic bacteria that produces the botulinum toxin in which causes flaccid paralytic disease. It is one of the most powerful toxins known to man. Toxin type A,B,E, and F cause disease in human beings. This toxin is a neurotoxin that can be very destructive to the nervous tissue. Clostridium botulinum is a bacterial illness. This illness is caused from bacteria growling on food. Clostridium botulinum causes paralysis. Botulinum neurotoxins prevent neurotransmitters from functioning properly, which causes paralysis to occur from top to bottom. Botulinum is extremely rare, however very dangerous. It is most common in infants (Smith TJ, Hill KK, Folev BT, et al. 2007)

The main purpose of the study was a way to annotate the assigned genome Locus Tags ((CLJ B3418-CLJ B3421, CLJ\_B1221 - CLJ\_B1222 and CLJ\_B0527). Through annotation, cell location and establishment of whether the genes were hypothetical or functioning, were determined. Another purpose of this project was to add further knowledge regarding Clostridium botulinum to the science community

Figure 1: A scanning electron micrograph of Clostridium botulinum showing the characteristic rod shaped growth pattern of the individual bacterial cells (Smith TJ Hill KK, Foley BT, et al, 2007



Figure 2. 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 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 - Structure-Based Evidence	TIGRfam, Pfam, PDB	Are there functional domains in my protein?
Module 4 - Cellular Localization Data	Gram Stain, TMHMM, SignallP, PSORTb, Phobius	Is my protein in the cytoplasm, secreted or embedded in the membrane?
Module 5 - 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?
Module 8 - Evidence for Horizontal Gene Transfer	Phylogenetic Tree	Has my gene co-evolved with other genes in the genome?
Final Annotation	Review data from all modules	Does the student proposed name of the gene agree with that proposed by the automated computer annotation? Are any changes proposed to the pipeline annotation?

# Results

#### CLJ B3418:

The initial proposed product of this gene by GENI-ACT was an oligopeptide/dipeptide ABC transporter, ATP-binding protein. This gene product proposal was supported by the top BLAST hits for the amino acid sequence, the presence of well-curated protein functional domains within the amino acid sequence, and the PFAM results match an ABC Transporter. As such, the proposed annotation is an ABC Transporter.

#### CLJ B3419:

Protein BLAST resulted in a top hit indicating a oligopeptide/dipeptide ABC transporter. ATP-binding protein. This agreed with the GENI-ACT proposed gene product. No transmembrane helices were predicted by TMHMM, suggesting that CLJ B3419 does not reside within the cell membrane. There is no signal peptide predicted, thus meaning there is no predicted cleavage site. The score was 0.131 which was below 0.450. As such, the proposed annotation is an ABC Transporter.

#### CLJ B3420

The initial proposed product of these genes by GENI-ACT was an oligopeptide/dipeptide ABC transporter, permease protein. This gene product proposal was supported by the top BLAST hits for the amino acid sequences and the presence of well-curated functional domains within the amino acid sequences.

# CLJ B3421

The predicted gene product by GENIACT was anoligopeptide/dipeptide ABC transporter, permease protein. BLAST results from SwisProt correlated with predicted gene. The Phobius prediction graph indicates a transmembrane, protein. Transmembrane proteins are integral proteins, which means the proteins can penetrate into or through the bilayer. Transmembrane proteins have three domains, the bilayer, the extracellular, and the intracellular domain. These proteins play important roles in the functioning of cells.



Figure 3- CLJ\_B3418: BLAST Top Hit showing a Putative Oligopeptide transport ATP-Binding protein YkfD.



prediction graph indicates a transmembrane protein.

### Peptide predicted CLJ B1221

Protein BLAST resulted in a top hit indicating a probable amino-acid ABC transporter permease protein PatM.TMHMM predicted 4 transmembrane helixes, suggesting that CLJ B1221 may reside within the cell membrane. COG puts proteins into particular function groups. Clostridium botulinum was found to be categorized into ABC-type methionine transport system, permease component [Amino acid transport and metabolism].



Figure 6 -- CLJ B1221 TMHHM posterior probabilities showing 4 transmembrane helices CLJ B122:

The predicted gene product was a polar amino acid uptake ABC transporter, PAATfamily, amino acid-binding protein. Its coordinates are 1255086..1255910. The WEBLOGO was not well conserved at the N-terminus more gaps exist. C-terminus is highly conserved. less gaps exist. Top BLAST hit for the amino acid sequence identified Vibrio harveyi. The gene product name is Putative amino-acid ABC transporter-binding protein PatH. Clostridium botulinum is a gram-positive bacteria with 1 transmembrane helice according to the TMHMM results.

# CLJ B0527

The predicted gene product by GENI-ACT for Clostridium botulinum was putative high-affinity branched-chain amino acid ABC transporter, amino acid-binding protein. It was identified by the match to protein family HMM PF01094. According to BLAST the top hit of the nr database Excluded Clostridium taxid 1485 MULTISPECIES: it was identified as a branched chain amino acid ABC transporter substrate-binding protein . A Pfam results indicates a Periplasmic binding protein. As such, the proposed annotation is an ABC Transporter.





# 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.

Figure 8 -CL | B0527 3-D structure

Gene Locus	Geni-Act Gene Products	Proposed Annotation
<u>CU_</u> B3418	oligopeptide/dipeptide ABC transporter.ATP-binding protein	Putative oligopeptide transport ATP-binding protein YkfD
CU_83419	oligopeptide/dipeptide ABC transporter.ATP-binding protein	oligopeptide/dipeptide ABC transporter,ATP-binding protein
CLJ_B3420	oligopeptide/dipeptide ABC transporter.permease protein	ABC-type dipeptide/oligopeptide/nicke I transport system, permease component
<u>CU_</u> 85421	oligopeptide/dipeptide ABC transporter. permeaseprotein	ABC-type dipeptide/oligopeptide/nicke I transport system, permease component
CLJ_B1221	amino acid ABC transporter, permease protein, His/Glu/Gln/Arg/opin e famil	ABC-type methionine transport system, permease component
<u>CU_</u> 81222	polar amino acid uptake ABC transporter, <u>PAATfamily</u> , amino acid-binding protein	amino acid ABC transporter substrate-binding protein, PAAT family
CLJ_B0527	putative high-affinity branched-chain amino acidABC transporter, amino acid-binding protein	putative branched-chain amino acid ABC transporter from Rhodospirillum rubrum

# References

Smith TJ, Hill KK, Foley BT, et al. Analysis of the Neurotoxin Complex Genes in Clostridium botulinum A1-A4 and B1 Strains: BoNT/A3, /Ba4 and /B1 Clusters Are Located within Plasmids. Salzberg S, ed. PLoS ONE. 2007;2(12):e1271 doi:10.1371/journal.pone.0001271

# Acknowledgments

Supported by an NIH Science Education Partnership (SEPA) Award - R250D010536

Dr. Stephen Koury and Dr. Rama Dev-Rao

www.buffalo.edu

