

Annotation of the *Clostridium botulinum* Genome at Locus Tags

CLJ_0039 and CLJ_0061

Anthony Oedbauer, Rachael Empie, Tomi Cordz, Eirinn Mangan and Mr. Kirk Winters
Pioneer High School and Yorkshire NY and The Western New York Genetics in Research Partnership



Abstract

Two genes from the microorganism *Clostridium botulinum* Ba4 657 CP001081 (CLJ_0039 and CLJ_0061) were annotated using the collaborative genome annotation website GENI-ACT. The Genbank proposed gene product name for the 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, potential alternative open reading frames, enzymatic function, presence or absence of gene duplication and degradation, the possibility of horizontal gene transfer, and the production of an RNA product. 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 r database.

Introduction

Clostridium botulinum is a Gram-positive, rod-shaped, anaerobic, spore-forming, motile bacterium with the ability to produce the neurotoxin botulinum. This organism is classified as a Clostridia, also this bacteria is an important cause to diarrhea. The discovery of *Clostridium botulinum* and botulinum toxin was made by E. van Ermengem in 1895. About 75 years later a subtype of the toxin proved to be highly effective in the treatment of dystonia's and is now widely used.

Clostridium botulinum is a microorganism of interest for several reasons. This bacterium is a natural source of the Botulinum toxin, one of the most toxic poisons known. Although when a small amount of Botulinum toxin is injected into a muscle, it blocks nerve signals that tell your muscles to contract. The effect is that it temporarily weakens or paralyzes the facial muscles and smooths or eliminates wrinkles in the skin for a few months. This organism has been tried to be weaponized by USA and Russia but it could not be effectively weaponized.

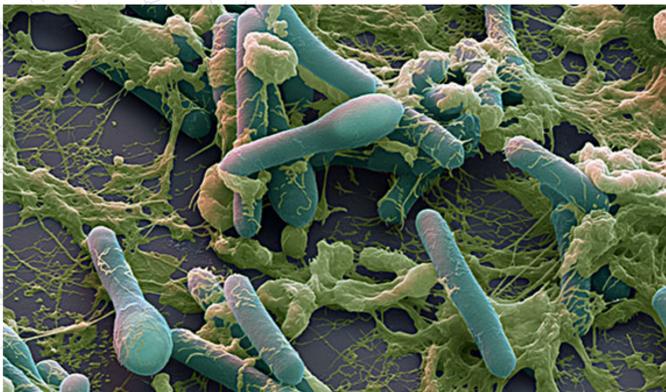


Figure 1- Scanning electron micrograph of *Clostridium botulinum*

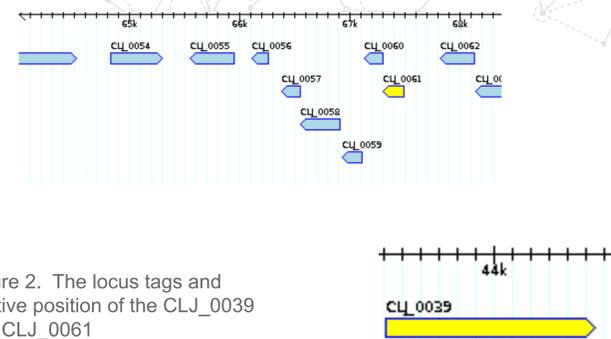


Figure 2. The locus tags and relative position of the CLJ_0039 and CLJ_0061

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- 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?
Module 5- Structure-Based Evidence	TIGRFam, Pfam, PDB	Are there functional domains in my protein?
Module 6- Enzymatic Function	KEGG, MetaCyc, E.C. Number,	In what process does my protein take part?
Module 7- Gene Duplication/ Gene Degradation	Paralog, Pseudogene	Are there other forms of my gene in the bacterium? Is my gene functional?
Module 8- Evidence for Horizontal Gene Transfer	Phylogenetic Tree,	Has my gene co-evolved with other genes in the genome?
Module 9- RNA	RFAM	Does my gene encode a functional RNA?

Results

CLJ_0039:
The initial proposed product of this gene by GENI-ACT was DNA polymerase. 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, the transmembrane topography of the amino acid sequence, and the cellular location of the amino acid sequence. As such, the proposed annotation is DNA polymerase.

CLJ_0061:
The initial proposed product of this gene by GENI-ACT was hypothetical 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, the transmembrane topography of the amino acid sequence, and the cellular location of the amino acid sequence. As such, the proposed annotation is hypothetical protein.

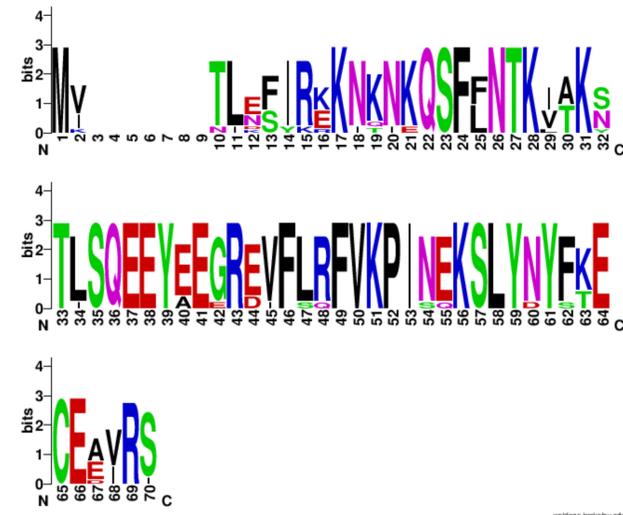


Figure 3 – A portion of the CLJ_0061 Weblogo

Query	42	IILSGSS-----DFTLKQELSN---IIISEEGNINISVITLNAIIKKMSNEEISLFIDPK	93
Sbjct	43	+IL+GS + T+ +++ + I+E G++ + IIKK+ +E+ L + + VILTGSDSSEIIEITIPKQVGDVEIETETG5VLPGRFFVDIINKLPGKEVKLSTNEQ	102
Query	94	NNDVLKIKQGRNN-SKIALAKEHTISDILKSDIISLNLNPKIFKDLIESTSYAVAQDES	152
Sbjct	103	++ N S + + + + + D ++ L+ K+ K++I T++AV+ E+ FQTLITSGHSEFNLSGLDPDQYPLPEVSRDD---AIQLSVKVLKNIIAQTNFVSTSET	159
Query	153	RPILLGQYIEENLNLIKITALDGFRLISNSAKIENKHKKK-ILIHANNLKNLSSLDIST	211
Sbjct	160	RP+L G +++N + TA D RL ++E++ ++K ++I L L+ ++ DS RPVL TGVNMLIQNELICTATDShRLAVRKLQLEDESENKNVITPGKALSELNKHMSDS	219
Query	212	EELILNKNKKAIVEVKNKNYAAIASNIFSEEFIEYKQLIGDLSFPPIKIVDRLQLLN	271
Sbjct	220	E++ + +N+ V N N+ S + + + + L + ++ IK+ ++ + EDIDIFFASNQVLFVRVGINIFI---SRLLLEGHYPDTRLFPPE--NYEIKLGINNGDFYH	273
Query	272	TVDRISVVSFVKEGTHLMKLNKDS--IEIDGNSS-VASMKGEVNI--IKNNFEEFET	325
Sbjct	274	+DR S+++ +EG + ++KL+ + +E+ S + ++K EVN ++ N + AIDRASLLA---REGGNVVKLSTGNELVELSSTSPETGTVKEEVNANDVEGGNLIK----	326
Query	326	IGFNPKYILDALKNINTKEIDIFFNSMNSPAIINKPEKAEDISVNALVLPVKT	379
Sbjct	327	I FN KY++DALK I+ E+++ F + P I+ KP+ +D SV L+LP++T ISFNSKYMMDALKAIIDNDEVEVEFFGTMKPFIL--KPK--DDDSVTQLPIRT	376

Figure 4 – CLJ_0039 First BLAST hit pairwise alignment hit for product proposal DNA polymerase.

Score	Expect	Method	Identities	Positives	Gaps
125 bits(314)	1e-36	Compositional matrix adjust.	63/63(100%)	63/63(100%)	0/63(0%)
Query	1	MVTLFIRKKNKNSQFFNTKVAKSTLSQEEYEEERDVFRLFVKPINEKSLYNYFTECEE	60		
Sbjct	1	MVTLFIRKKNKNSQFFNTKVAKSTLSQEEYEEERDVFRLFVKPINEKSLYNYFTECEE	60		
Query	61	IRS 63			
Sbjct	61	IRS 63			

Figure 5 – CLJ_0061 First BLAST hit pairwise alignment

Conclusion

The GENI-ACT proposed gene product did not differ at all from the proposed genes annotation, the genes appear to be correctly annotated by the computer database.

References

Botulinum Toxin: A Bioterrorism Weapon
BY DOUG HANSON, PHD ON APR 1, 2004-
<http://www.emsworld.com/>

Acknowledgments

Supported by an NSF Innovative Technology Experiences for Students and Teachers (ITEST) Award - 1311902