# Annotation of *Clostridium botulinum* at Locus Tags CBO0001, CBO0604, CBO3572, CBO1389, and CBO1774

Sarah Alfiero, Bianca Foeller, Makenna Jarnot, Cassidy Pajda, Allison Russell, and Laura O'Donnell

Attica High School- Attica, NY and The Western NY Genetics in Research and Health Care Partnership

Abstract

A group of 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 rod-shaped bacterium, mostly found as spores, that produces toxins that causes food poisoning. They tend to thrive and grow in places with little oxygen, and can lay dormant for a very long time due to their hard, protective coating and their protective membrane layers. These bacteria cause a condition known as botulinum, or more commonly, food poisoning. The neurotoxins this bacterium produces are among one of the most toxic in the world, and even a microscopic dosage can cause death or severe sickness. Neurotoxins are poisonous chemicals that destroy, paralyze, and has harmful effects on the nerve tissues. It can be most commonly found in homecanned foods, unrefrigerated salsa, baked potatoes in aluminum foil, honey, garlic in oil, and poorly-prepared fish. They like more basic environments, so they will not be commonly found on acidic foods, such as tomatoes and most fruits. The effects of this poisoning can be disruption of the nerve functions, nausea, blurred vision, trouble breathing, paralysis or death. The main cause of food poisoning is improper home-canning of foods. If the process is done incorrectly, *Clostridium botulinum* and other bacteria present during the canning process will be destroyed by the boiling water and food, but the spores it produced will not be destroyed.

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



## 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, SignalP, PSORT, Phobius	Is my protein in the cytoplasm, secreted or embedded in the membrane?

Results

•						
Gene Locus: CBO0001				Gene Locus: CBO3572		
Basic Information	DNA Coordinates	11347			DNA Coordinates	complement(38158843817110)
	DNA Length	1347 bp	•	<b>Basic Information</b>	DNA Length	1227 bp
	Amino Acid Sequence Length	448 aa			Amino Acid Sequence Length	408 aa
		COG0593: DnaA - Chromosomal replication				COG2814:AraJ-Predicted arabinose efflux permease,
	COGS	initiation ATPase DnaA			0000	MFS family [Carbohydrate transport and metabolism]
Saguanca-basad		[Replication, recombination and repair]		Sequence-based	TCoffoo	Well conserved at the middle but not the beginning or
Sequence-based Similarity	T-Coffee	Well conserved throughout		Similarity		the end
Sinnanty		Very well conserved through amino acid 121			Wob-Logo	Fairly well conserved in the middle but not the
	Web-Logo	through amino acid 433. Not as well conserved in			Web-Logo	beginning or the end
		the beginning and the end.		TICDEAMo	2_A_01_02: multidrug resistance protein,	
	TIGRFAMs	DnaA: chromosomal replication initiator protein		Structuro-based		2A0115: MFS transporter, aromatic acid:H+ s
	PFAMs	Bac_DnaA: chromosomal replication initiator		Structure-based Evidence	PFAMs	MFS_1: transporter, major facilitator family
Structure-based	HMM Logo Key Residues	F7, F10, G13, P36, G44, G46, H49, Y69, F75,		LVIGENCE	HMM Logo Key Residues	G53,G60,R62,G96,G133W,150,F150,W210,Y233
Evidence		D99, D104, D105, D138, P141, G158, R203			PDB	no results
Lvidelice	Crystal structure of domain III from the		Collular	Transmembrane Helices	12	
	PDBThermotoga maritima replicatioDnaA	Thermotoga maritima replication initiation protein	i		Signal Peptide	none
		DnaA		LUCANZALION	PSORTb Final Prediction	Cytoplasmic membrane
	Transmembrane Helices	None	; (	Clostridium hotulinum CR03572		
Cellular Localization	Signal Peptide	None				
	<b>PSORTb Final Prediction</b>	Cytoplasmic		COG, HGRFAM, and PFAM data support CBO35/2 to be a carbohydrate		
			1 1		· · · · · · · · · · · · · · · · · · ·	

Clostridium botulinum CBO0001:

COG, TIGRFAM, PFAM, and PDB data all support CBO0001 to be a replication initiation protein that begins the process of DNA replication during binary fission. Cellular localization data predicts that the protein resides in the cytoplasm where replication would occur in the bacteria. Due to the numerous orthologs, it can be inferred that many species of bacteria contain a similar protein with a similar function.

	1-1-258 522-0-0-528
	4- 3- 3122- 1- 0 80 80 80 80 80 80 80 80 80 80 80 80 8
	4- 3- \$122- 1- N 8266
	4- 3- 1- 0- N SE SSE
TEVI GNGNREAHAASLAVAEAPAKAVAPLEIY	4- 3- 1- 0- 5986 N
ġ ĊĠŴĠĹĠĸŦĦĹĬŀĦĂĬĠĦŸŸĿĿĦŊ₽ġĸĸŴĸŴţĸŚŚ	4- 3- 1- 0-115 814 814 814 814 814 814
<b>EKETNE</b> E INS B <mark>OKK AND RANDOW I ID</mark>	4- 3- 212 - 1- 200 84 80 84 80 84 81 81 81 81 81 81 81 81 81 81 81 81 81
\$1 • DIGFAGKERIQEEFFHIENTLHE <sub>est</sub> kolviS	

SURPOKE DTLEDRIRSREENGLID OPPOL RIA LEKKALZERER PREVEZZIARZES. KRTKSVÆPROLANVLSRELTDSSLPK IG CCRDHTTV HAHEK S.L. DE LODA 

Figure 1 – WebLogo data for *Clostridium botulinum* CBO0001 showing that the protein is very well conserved from amino acid 121 through 433.

Gene Locus: CBO0604			
	<b>DNA Coordinates</b>	702029702421	
<b>Basic Information</b>	DNA Length	393 bp	
	Amino Acid Sequence Length	130 aa	
	COCS	XRE- DNA-binding transcriptional regulator, XRE-	
Sequence-based	6005	family HTH domain [Transcription];	
Similarity	T-Coffee	Well conserved throughout	
	Web-Logo	Well Conserved throughout	
Structure-based Evidence	TIGRFAMs	couple_hipB: transcriptional regulator, y4m	
	PFAMs	HTH_3: DNA-binding helix-turn-helix protein	
	HMM Logo Key Residues	L1, I1, R5, A16, G20, E30, G32, P36, L44, L48	
	PDB	30P9: Entity 1 containing Chain A	
Cellular Localization	Transmembrane Helices	1	
	Signal Peptide	None	
	<b>PSORTb Final Prediction</b>	Cytoplasmic	

Clostridium botulinum CBO0604:

TIGRFAM, PFAM, and PDB data CBO0604 to be a transcriptional support regulator during gene expression. Although some of the cellular localization data is conflicting, the PSORTb final prediction is that this protein resides in the cytoplasm which coincides with the protein's proposed function as a transcriptional regulator. Due to the presence of many orthologs, it can be inferred that CBO0604 is a common protein in many different species of bacteria.

Figure 2 – WebLogo data for *Clostridium botulinum* CBO0604 showing that the protein is very well conserved throughout.

Clostridium botulinum CBO1389: BLAST results and PFAM data supports CBO1389 to be a tetratricopeptide repeat motif protein. A structural motif is a protein that is a supersecondary structure to a variety of proteins or nucleic acids. Although some of the cellular localization data is conflicting, the PSORTb final prediction is that this protein resides in the cytoplasm. Due to a significant amount of orthologs present, the inference can be made that this protein is common in different bacteria species in need of tetratricopeptide repeat motif proteins.





transporter, possibly of the sugar arabinose. Membrane transport proteins such as this usually contain many transmembrane helices. Cellular localization data supports this prediction with the presence of 12 transmembrane helices. Due to the presence of many orthologs, it can be inferred that this is a common **REAL** protein in many different species of bacteria.

NTVSRKVA	SLAIBLAPAGR AQALSLIAT UTALAMYLCH	AS LY S.A. LL.YCLLLIPA S. HLAY
w <b>LR</b> vyTLA, AAF (FNT)	PIGRIVGOYEG RYJE-AIGUGALIJILGL.K	
A. FHULTAOVG MUTIYAW		LAPPATDY AMALESSIEN GIGAGAL YONOXS
		HUSMS CAN AYPA AA L. W. FRew
TVLVISBIGYAFAHALEWSIT	FS.N- ATVILLIGABILOS IK ONO.	



Figure 3 – WebLogo data and TMHMM data for Clostridium botulinum CBO0001 The Weblogo image on the left shows that the protein was fairly well conserved throughout the middle. The cellular localization data on the right from the TMHMM database shows the presence of 12 transmembrane

Gene Locus: CBO1389		
DNA Coordinates	15129121513805	
DNA Length	894 bp	
Amino Acid Sequence Length	297 aa	
COGS	No significant COG hit	
T Coffoo	Well conserved at beginning and not well conserved	
I-COllee	at the middle or end	
Web-Logo	Well conserved through 33-191 and not well	
	conserved from 192 on	
TIGRFAMs	No TIGRFAMs given	
PFAMs	TPR_2: tetratricopeptide repeat	
HMM Logo Key Residues	G8, A20	
PDB	No significant hit	
Transmembrane Helices	1	
Signal Peptide	none	
<b>PSORTb Final Prediction</b>	Cytoplasmic	
	Gene Locus:DNA CoordinatesDNA LengthAmino Acid Sequence LengthCOGST-CoffeeWeb-LogoTIGRFAMsPFAMsHMM Logo Key ResiduesPDBTransmembrane HelicesSignal PeptidePSORTb Final Prediction	

KKK - VLLL - P		
GYY.Y.K.K.YN LLY AN MERKYDRA	4 3 4 3 4 5 4 5 5 5 5 5 5 5 5 5 5 5 5 5	
AL X P AL X P		
NE GENEKRILAL EKY D K.Y.		
AOK ECKKY NERKAKANNSY		A A A A A A A A A A A A A A
WebLogo data for Clostridium	botulinum CBO1389 showing	that the protein is very well conser

Basic I

#### Cellular Loc

Clostridium botulinum CBO1774: COG, TIGRFAM, PFAM, and PDB data support CBO1774 to be an ATP-binding cassette transporter protein (ABC transporter) belonging to a transport system superfamily. It most likely hydrolyzes ATP into ADP and utilizes the released energy to function and do its job. It can be found in the cytoplasm and data suggests that the protein is a type of import protein that helps to transport various materials into cells.

A B YLOYDA BAR BAR L P. LOL	ENDER CONSTURY OF A CONSTURY.	L GL P. G	Figure 5 – <i>Clostridium</i> showing that conserved in not well conse or end.
-----------------------------	---	-----------	---

## 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 Loc CBO000

CBO060 CBO357

CBO138 CBO177

## References

FSIS. (n.d.). Retrieved from https://www.fsis.usda.gov/wps/portal/fsis/topics/food-safety-education/getanswers/food-safety-fact-sheets/foodborne-illness-and-disease/clostridiumbotulinum/ct index

## Acknowledgments

	Gene Locus: CBO1774	
Basic Information	<b>DNA Coordinates</b>	Complement (18829631884543)
	DNA Length	1581 bp
	Amino Acid Sequence Length	526 aa
	0000	Uup - ATPase components of ABC
Converse based	COGS	transporters with duplicated ATPase
Sequence-based	TCoffee	
Similarity	I-Cottee	well conserved in the middle and the end.
	Web-Logo	Well conserved in the middle, but not at
		the beginning or end.
	TIGREAMs	ABC_ABC_ChvD: ATP-binding cassette
		protein
Structure-based Evidence	PFAMs	ABC_tran: ABC transporter, ATP-binding
		protein
	HMM Logo Key Residues	G18, G21, G23, K24, S110, G112, D133
		EttA binds to ribosome exit site and
	PDB	regulates translation by restricting
		ribosome and tRNA dynamics
ellular Localization	Transmembrane Helices	None
	Signal Peptide	None
	PSORTb Final Prediction	Cytoplasmic

WebLogo data for botulinum CBO1774 the protein is well the middle, but then erved at the beginning

us	Geni-Act Gene Products	Proposed Annotation
1	DnaA: chromosomal replication initiator protein	DnaA: chromosomal replication initiator protein
4	DNA-binding transcriptional regulator	DNA-binding transcriptional regulator
2	Arabinose efflux permease	Arabinose efflux permease
9	Tetratricopeptide repeat motif protein	Tetratricopeptide repeat motif protein
<b>'</b> 4	ABC transporter, ATP-binding protein	ABC transporter, ATP-binding protein

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

### www.buffalo.edu