Annotation of the *Clostridium botulinum* Genome from DNA Coordinates CB00210, CB00211, CB00214

Gabrielle Cole, Elise Gonzalez, Doniquelle Joseph, Felix Rodriguez, Aleixa Serrano, Richard Nagler and Nicole Welka Emerson Annex Buffalo, NY and The Western New York Genetics in Research and Health Care Partnership

Abstract

A group of three genes from the microorganism Clostridium botulinum (CBO0210, CBO0211, CBO0214) were annotated using the collaborative genome annotation website GENI-ACT. The Genbank proposed a gene product name for each gene. It was assessed in terms of the general genomic information, amino acid sequence-based similarity data, and structure-based evidence from the amino acid sequence. The Genbank proposed gene product name did not differ significantly for two of the three genes from the proposed gene annotation for each of the genes in the group and as such, two of the three genes appear to be correctly annotated by the database.

Introduction

Clostridium botulinum is a bacteria that grows on food and produces toxins, that when ingested, causes paralysis. It's found in soil and untreated water throughout the world. It produces spores that survive in improperly preserved or canned foods, where they produce a toxin. The toxin that is produced is a neurotoxin known as botulinum, which is rare. The toxin affects your nerves, paralyzes you, and even causes death. Taking a small bite of food containing this toxin can be deadly. A frequent question asked is, can you get botulinum from a dented can? The answer is yes. When a dented or damaged can creates a pinhole-size opening that allows the mixture of air and moisture it can spur the growth of bacteria, which contaminates the food. There are three main kinds of naturally occurring botulinum. They are foodborne, wound and infant. They occur when the spores produce the toxin in an anaerobic environment. There is currently no vaccine, but the U.S. Centers for Disease Control has a supply of antitoxin for treating cases of botulinum.



Figure 1. Clostridium botulinum

Methods

Modules of the GENI-ACT (http://www.geni-act.org/) were used to complete Clostridium botulinum genome annotation. The modules are described below (only modules 1, 2, and 5 were used for annotation):

, ,		
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

CBO0210:

The initial proposed product of this gene by GENI-ACT was a *Clostridium botulinum* malic enzyme. Malic enzymes are enzymes that catalyze the oxidative decarboxylation of malate to pyruvate and carbon dioxide concomitantly reducing NADP+ to NADPH. Malic enzymes are also used for the reductive biosynthesis of molecules such as fatty acids. This result was supported by the results of the BLAST hits from the amino acid sequence, the results from the WebLogo, along with the T-Coffee results. The transmembrane topography of the amino acid sequence was also supporting evidence to further conclude that the proposed annotation is a *Clostridium botulinum* malic enzyme.



CBO0211: The initial proposed product of this gene by GENI-ACT was a competence/damage-inducible protein cinA. This gene product proposal was supported by the top BLAST hits for the amino acid sequence. The protein cinA allows the bacteria to use any DNA strands in the surrounding area as a means of transformation. As such, the proposed annotation is a competence/damage-inducible protein cinA.

NADP-c	lependent malic enzyme [[Eubacterium] combesii]	
NCBI Refere	ence Sequence: WP_099839543.1	
Identical Prote	eins FASTA Graphics	
<u>Go to:</u> 🕑		
LOCUS DEFINITION ACCESSION	<pre>WP_099839543 390 aa linear BCT 13-NOV-2017 NADP-dependent malic enzyme [[Eubacterium] combesii]. WP_099839543</pre>	
VERSION	WP_099839543.1	1
KEYWORDS	RefSeq. [Eubacterium] combesii	
ORGANISM	[Eubacterium] combesii	
	Bacteria; Firmicutes; Clostridia; Clostridiales; Eubacteriaceae;	
COMMENT	Eubacterium.	
COMMENT	sequence which may be annotated on many different RefSeq genomes	
	from the same, or different, species.	
	COMPLETENESS: full length.	
		1

Figure 3 – Top blast hit for CBO0210 not of the same genus and species.



Figure 4 – Blast hits from Swiss pro for CBO0211



Figure 5 – Gene neighborhood for CBO0211

CBO0214:

The initial proposed product of this gene by GENI-ACT was a vancomycin resistance protein. This gene product proposal was primarily supported by the top BLAST hits for the amino acid sequences, however most species showed this sequence to be a hypothetical protein. Vancomycin is a strong antibiotic used as ¹ a last resort in the treatment of a bacterial infection. Due to this gene conferring a resistance to Vancomycin, it would make it very difficult for someone to be cured. Further research would be needed in order to definitively call this proposal correct.



Figure 7 – Top blast hits for CBO0214 showing hypothetical protein

Conclusion

The GENI-ACT proposed gene product did not differ significantly from the proposed gene annotation for two of the three genes in the group and as such, two genes appear to be correctly annotated by the computer database.

Gene

CBO0

CBO0

References

https://www.fsis.usda.gov/wps/portal/fsis/topics/food-safetyeducation/get-answers/food-safety-fact-sheets/foodborne-illness-anddisease/clostridium-botulinum/ct_index accessed on April 26, 2018



assistance.

ocus	GENI-ACT Products	Proposed Annotation
210	Malic enzyme	Malic Enzyme
211	Competence/Damage-Inducible Protein CinA	Competence/Damage-Inducible Protein CinA
214	Vancomycin Resistance Protein	Vancomycin Resistance Protein

Acknowledgments

Supported by an NIH Science Education Partnership (SEPA) Award - R250D010536 A special thanks to Dr. Stephen Koury for all his support and

www.buffalo.edu