Annotation of the *Clostridium botulinum* Genome Strain 657/Type Ba4 at Locus Tags CLJ_0014 & CLJ_0018

Joseph S. Bowers, Olivia L. Watroba & Margaret S. Jolles
Silver Creek Central School District & The Western New York Genetics in Research and Health Care Partnership

SEPA Science Education Partnership Award
Supported by the National Institutes of Health
NIH National Institutes of Health Turning Discovery Into Health
University at Buffalo

Abstract

Two students from Silver Creek Middle and High Schools participated in the Western New York Genetics in Research and Health Care Partnership Gene Annotation Research Study. Annotation is the process of assigning function or biological significance to a gene. Using the Geni-Act website, students were able to annotate their gene based off of Basic Information, Sequence Based Similarity Data, Structure Based Evidence, Cellular Localization Data, Alternative Open Reading Frame, Duplication and Degradation and Horizontal Gene Transfer. The students were able to obtain data including protein structure, function and location. They were also able to hypothesize if horizontal gene transfer occurred.

Introduction

Clostridium botulinum is a gram positive, anaerobic, rod shaped bacterium associated with improperly produced canned goods. It produces a toxin that is lethal. This is most commonly obtained when a canned food item is consumed that was not stored at proper temperature, was not prepared properly at time of canning or when a can has begun to swell and the contents are consumed. *Clostridium botulinum* was first discovered and isolated by Emile van Ermengem in 1896, and it was later deemed to survive by forming spores, remaining in a dormant (sleep like) state until environmental conditions are perfect for growth.

Clostridium botulinum consists of seven subtypes. Each subtype produces a different botulinum toxin; with the exception of subtypes three and four, all are human pathogens. Types one and two, commonly found in soil are the primary cause of botulism outbreaks in the United States. Type five commonly found in fish is also a contributor to the cases of botulism in the United States.

Our team analyzed *Clostridium botulinum* Strain 657/ Type Ba4. Our Locus tags were CLJ_0018 being one of the largest genes and CLJ_0014 being one of the smallest genes in the group. Our goal was to analyze and find the functions of our two genes.

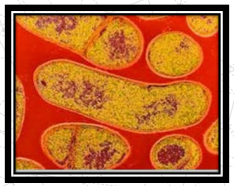

Figure 1.
Clostridium botulinum
Gram Positive
Anaerobic Bacterium

http://parasites.ftz.czu.cz/food/_data/141.jpg

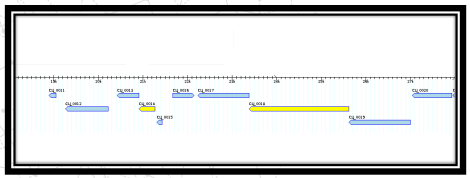

Figure 2: The locus tags (in yellow) and relative position of the genes under investigation in this research

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- 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?

Results

CLJ_0014: Of the two proteins described in this research, *CLJ_0014* is quite short with a sequence of 122 amino acids. *CLJ_0014* was thought to be a transcriptional regulator/factor, AraC family, and this was supported by annotation. The first hit from curated Swiss Protein Data Base resulted in description of an HTH (Helix turn Helix) type transcriptional regulator. A transcription regulator/factor is a protein that binds DNA in a sequence specific manner to regulate transcription. *CLJ_0014* is therefore not an enzyme, and Module #6 was not necessary in this annotation. The results of both COG and Pfam searches also supports the protein being one that binds DNA and the WebLogo confirms that the protein is highly conserved when compared to other similar proteins. The computer called the start and stop codons correctly, due to the position of the Shine-Delgarno sequence, and the stop is at the proper position for the length of the protein. Protein Data Base produced 3 acceptable hits showing that sections of the protein *CLJ_0014* may fold in a similar way. *CLJ_0014* is a transmembrane protein with a single helix. This is supported by both the TMHMM score and Phobius. Horizontal gene transfer is not predicted because it shares a Phylum, *Firmicutes,* with the most closely related organisms seen on cladogram in Figure 6. *CLJ_0014* has two paralogs, but is not a pseudogene.

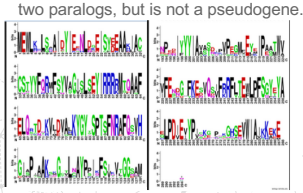

Figure 3: WebLogo for *CLJ_0014* illustrating high conservation across the amino acid sequence

Prediction of CLJ_0014

```
ID   CLJ_0014
FT   TOPO_DOM     1     29       NON CYTOPLASMIC.
FT   TRANSMEM    30     48
FT   TOPO_DOM    49    122       CYTOPLASMIC.
//
```

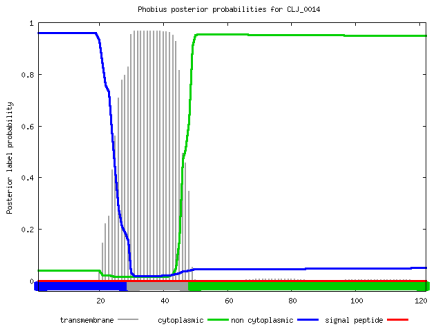

Figure 4:
Phobius results for *CLJ_0014* predicting a single transmembrane helix

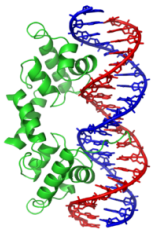

Figure 5: Helix Turn Helix domain (green) bound with DNA

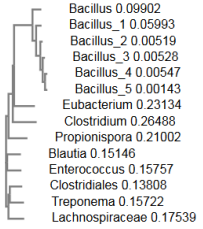

Figure 6: Cladogram; *Clostridium* and 13 other species

CLJ_0018: In contrast with the first protein discussed, *CLJ_0018* is very long having a sequence of 746 amino acids. This protein was hypothesized in Geni-Act to be a *Chaperone Clp B.* In general chaperone proteins assist other biomolecules to fold into the correct 3 dimensional shape. The first and second BLAST hits from the curated Swiss Protein Data Base resulted in the identification of this protein as an *ATP-dependent Clp protease ATP-binding subunit ClpC*, with alignment length and e-values being identical. A single COG hit showed similar results but instead of a *ClpC*, it identified a *ClpA*. *Clp* proteins belong to a group known as *Collagen-Like Proteins* and have been identified in a broad range of bacteria and have similar functions. T-Coffee and WebLogo comparisons of 12 other proteins with *CLJ_0018* resulted in mixed results of high and low conservation across the 13 proteins.

The location of *CLJ_0018* is most likely cytoplasmic, having no transmembrane helices predicted (TMHMM), no signal peptide predicted (Signal P) and a high cytoplasmic score of 9.97 from PSort B.

TIGRFAM gave one result, once again a *chaperone ClpB; ATP-dependent chaperone protein ClpB.* Pfam data base, which focuses on protein domains and families showed 2 results, *AAA_2* or *ATPases Associated with diverse cellular Activities* & *ClpB_D2-small.* Protein Data Base also confirms similarity with *CLP Complexes, AAA type protein and chaperone activity.*

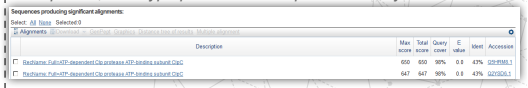

Figure 7: Top BLAST hits from Swiss Protein for *CLJ_0018*

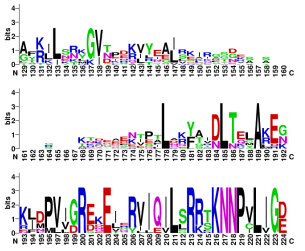

Figure 8: WebLogo for CLJ_0018 showing mixed conservation of sequences

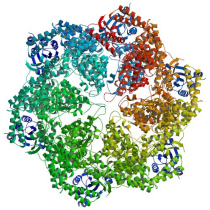

Figure 9: PDB result; 3D protein 3J3S - Structural dynamics of the MecA-ClpC complex revealed by cryo-EM

Conclusion

The GENI-ACT proposed gene product did not differ significantly from the proposed gene annotation for each of the genes. The genes appear to be correctly annotated by the computer database.

Gene Locus	Geni- Act Gene Product	Proposed Annotation
CLJ_0014	Transcriptional activator, AraC family	Transcriptional regulator containing an HTH domain
CLJ_0018	Chaperone ClpB	ATP dependent Clp Protease ATP binding subunit; Chaperone protein

References

- http://www.sciencedirect.com/science/article/pii/S0923250814002046
- http://bioweb.uwlax.edu/bio203/s2008/strandwi_phil/schedule.htm
- http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0117414
- http://www.rcsb.org/pdb/explore/explore.do?structureId=3J3S

Acknowledgments

- Supported by an NIH Science Education Partnership (SEPA) Award - R25OD10536-1 A1
- Stephen Koury Ph.D. – Research Associate Professor – Dept. of Biotechnical & Clinical Laboratory Sciences
- Jennifer R. Cooke, *MPS, CSA*, Program Coordinator WNY Rural Area Health Education Center (R-AHEC)