



Annotation of the *Clostridium botulinum* Genome for Locus Tags CLH_0149 to CLH_0151

Alanna Swanson, Chloe Heinrich, Beth Brombos, Sarah England, Sanna Vollbrecht and Mrs. Rich
Frontier Central High School, 4432 Bayview Road Hamburg, NY 14075 and The Western New Genetics in Research and Health Care Partnership

Abstract

A group of consecutive 3 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 in the nr database.

Introduction

Clostridium botulinum is a gram positive, rod shaped spore forming bacteria. It is divided into 4 phenotypic groups and it is classified into seven stereotypes based on the antigenicity of the toxin it produces. *Clostridium botulinum* can be found in 3 main forms of botulinum, wound, infant, and foodborne botulism. The most common form of botulinum is infant botulism where the spores grow within the baby's intestines. Wound botulism can be contracted by getting contaminated soil in a wound or an injection with a contaminated needle. Foodborne botulism is contracted by consuming improperly canned food that the spores have grown in.

A team of seven students from Frontier Central High School were randomly chosen from fourteen volunteers to participate in the gene annotation research. The students selected *Clostridium botulinum* as their bacteria to research and were assigned genes CLH_0147 to CLH_0153. Four other students remained as part of the control group and researched background information of the bacteria as well as gene functions. (1,2)



This picture shows what a scanned microscopy image of *Clostridium botulinum* bacteria looks like.

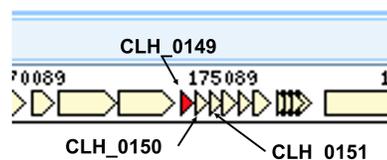


Figure 1. The locus tags and relative position of the genes under investigation in this research

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 |
|-----------------------------------|--|--|
| Basic Information | DNA Coordinates and Sequence, Protein Sequence | What is the sequence of the gene and protein? Where is it located in the genome? |
| Sequence-Based Similarity | Blast, CDD, T-Coffee, WebLogo | How similar is the protein under investigation to other proteins in Genbank? |
| Structure-Based Similarity | TIGRFam, Pfam, PDB | What functional domains are present in the protein under investigation? |
| Cellular Localization | Gram Stain, TMHMM, SignalP, LipoP, Psortb, Phobius | Is the protein under investigation located in the cytoplasm, secreted, in the periplasm or embedded in the cell membrane or cell wall? |

Results

CLH_0149

Alanna Swanson

This gene turned out to be a DNA binding protein which is exactly what the computer originally thought it would be in the basic information. All of this was supported by the research in the nr database, protein BLAST, and other databases like Pfam, WebLogo, T-Coffee, CDD, PDB and TIGRFAM. DNA-binding proteins are proteins that bind to single- or double-stranded DNA, generally in the major groove if the binding is sequence-specific – as with transcription factors that regulate expression of genes, and nucleases that cleave DNA between nucleotides. DNA-binding proteins are proteins that attach to DNA. Transcription factors bind to regulatory sequences on DNA and turn transcription of genes on or off. ... It binds to DNA using three domains, the palm, finger and thumb. Histone proteins are structural proteins that bind to DNA using positively charged amino acids. (3)

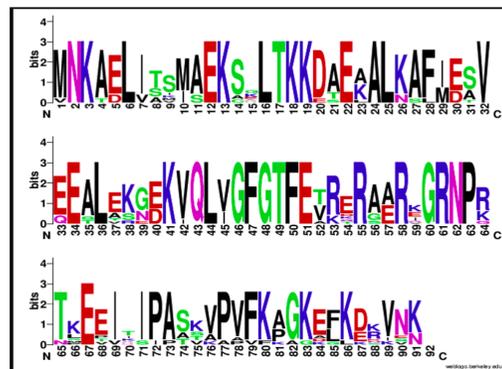


Figure 3: WebLogo CLH_0149

This WebLogo shows all the similarity between the binding protein and other genes. It shows how CLH_0149 has a lot of genes that are practically the same and others that are very different. This WebLogo is very conserved and well put together.

CLH_0150

Chloe Heinrich

The computer assumes that this gene is an S4 domain protein, which is furthermore proven by the BLAST hits for the amino acid sequence as shown in Figure 5. In addition to this, the Pfam database also says that the Pfam name is S4 domain as well as the clan name. The S4 domain is a group of enzymes that are involved in the translation of RNA.

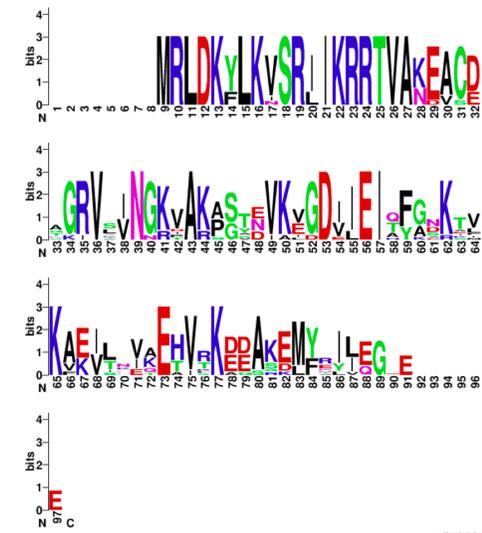


Figure 4: This WebLogo is showing the similarities in this protein's sequence compared to other proteins. There was a lot of matches in the beginning of the sequence but following that the matches get pretty scattered.

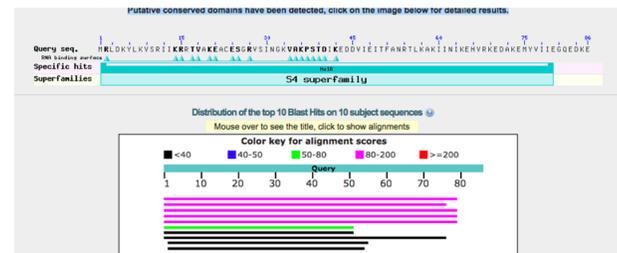


Figure 5: Above is the BLAST results. The pink lines show a strong hit the green line shows a medium hit and the black line shows a weak hit.

CLH_0151

Beth Brombos

The computer proposed that the product of this gene was a sporulation protein YabP. This was supported by the top BLAST hits for the amino acid sequence. This was also shown in curated databases such as TIGRFam, Pfam, Web Logo, T-Coffee and PDB. The function of sporulation protein YabP is not fully known. It is likely that YabP does have a function in sporulation or germination, one that is either unappreciated or partially redundant with that of another protein.

Figure 6: CDD Hits for TIGRFam and Pfam databases for gene CLH_0151

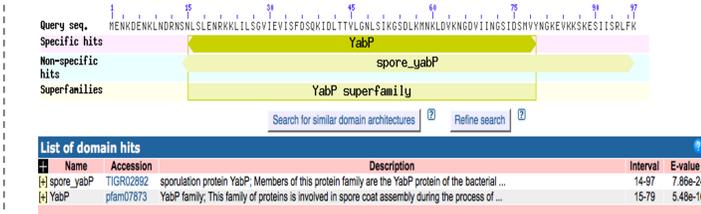
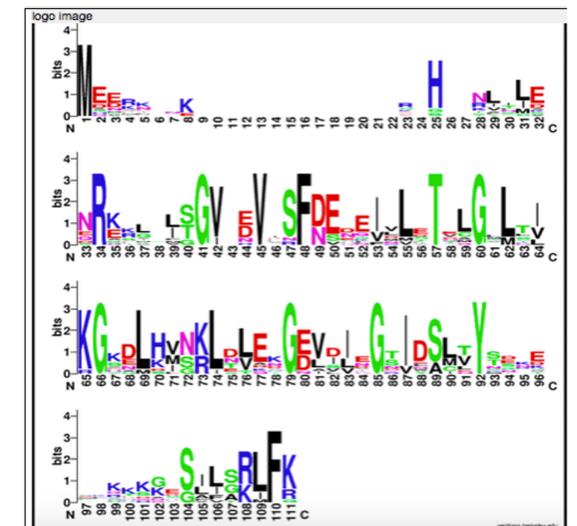


Figure 7: WebLogo for gene CLH_0151 compared to genes CLH_0149 and CLH_0150 has little conservation. Hits are scattered throughout the amino acid sequence. Numbers 34 to 92 have the most similarity to other proteins while the beginning and end of the sequence have few similarities.



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.

| Locus Tag | Pipeline Annotation Product Name | Proposed Annotation | Changes Proposed? |
|-----------|----------------------------------|--------------------------|-------------------|
| CLH_0149 | DNA-binding protein HU | DNA-binding protein HU | No |
| CLH_0150 | S4 domain protein | S4 domain protein | No |
| CLH_0151 | Sporulation protein YabP | Sporulation protein YabP | No |

References

- (1) Hoecker, J. I. (2018, 5, 15). Infant botulism: Can it be prevented? retrieved from <http://mayoclinic.org>
- (2) Botulism (2019, May 19). Retrieved from <http://www.foodsafety.gov>
- (3) <https://www.nature.com/subjects/dna-binding-proteins>

Acknowledgments

Supported by an NIH Science Education Partnership (SEPA) Award - R25GM129209; Dr. Koury, Dr. Dey-Rao