

Annotation of the *Clostridium Botulinum* Genome for Locus Tags CLH147,CLH148, CLH152, CLH153

Sebastian Baco, Jaclyn Conley, Nathaniel Grosskopf, Sam McLaughlin, Emma Bartoe, Madison Santangelo, and Mrs. Rich Frontier Central High School and The Western New Genetics in Research and Health Care Partnership



Abstract

A group of four 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, 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 four phenotypic groups and is classified into seven serotypes based on the antigenicity of the botulinum toxin produced. The bacteria grows on food and produces toxins. It is prevalent in soil and marine sediments, most commonly as spores. The spores are not harmful until they begin to grow out into active bacteria and produce neurotoxins. The neurotoxins in botulism prevent neurotransmitters from functioning properly. As it progresses, paralysis will occur starting with the eyes and face and then moves to the throat, chest and extremities. When paralysis reaches the chest, death from the inability to breathe will occur unless the person is ventilated. Symptoms appear 12 to 72 hours after consuming the contaminated food. Illness lasts from 1-10 days with treatment, however full recovery can take weeks to months and some never recover. There is a 35- 65% chance of death for patients that haven't been treated immediately. Some of the symptoms include nausea, fatigue, dizziness, slurred speech, body aches, drooping eyelids, difficulty swallowing and lack of a fever. Some treatments are induced vomiting by doctors and they also give medicine to induce bowel movements to try and clear out the digestive system. Injected anti-toxins reduce the risk of complications. It keeps the toxin from attacking your nerves however, it cannot reverse the damage already done.

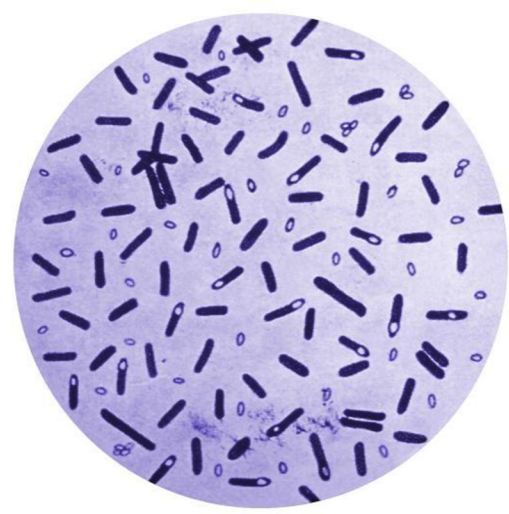


Figure 1: Microscopic view of gentian violet-stained Clostridium botulinum and their spores. Source: Public Health Image Library, Center for Disease Control (1979)

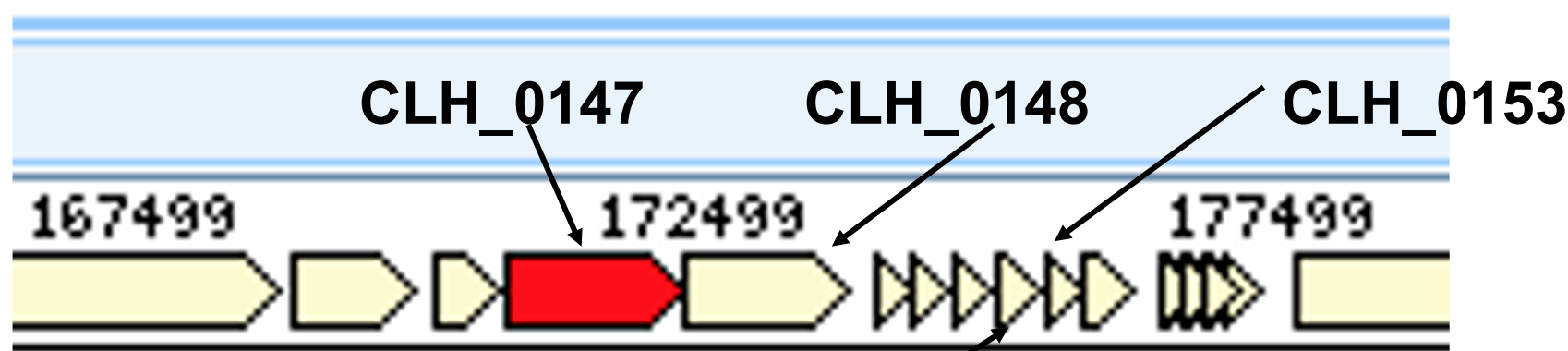


Figure 2. The locus tags and relative position of the genes under investigation in this research

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_0152 (Sam McLaughlin):

The computer pipeline proposed product of this gene was a spore cortex biosynthesis protein. This was supported by the top BLAST hits for the amino acid sequence, as well as TIGRFAM and Pfam hits. The Pfam description predicted that the gene would span the membrane several times, as well as suggested that it's only found in genomes of species that perform sporulation. In this context, sporulation is the formation of nearly dormant forms of bacteria. Spores can preserve the genetic material of the bacteria when conditions are inhospitable and lethal for the normal (vegetative) form of the bacteria.

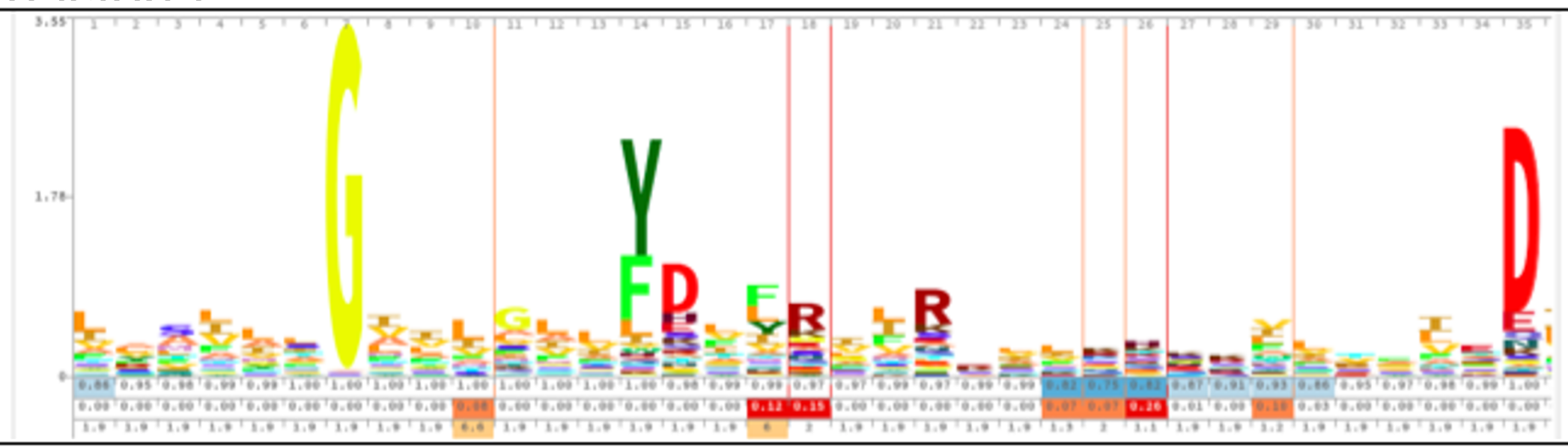


Figure 3 - CLH_0152
Segment of WebLogo of the sequence alignment of significant orthologs from the BLAST search. <https://weblogo.berkeley.edu/>

CLH_0147 (Jaclyn Conley):

The computer had originally proposed this gene as a stage V sporulation protein B. Sporulation is the process of the particular cell/gene protecting its genetic material in unfavorable conditions. This was supported by the results from the top BLAST hits for the amino acid sequence and the TIGRFAM results. Additional evidence from TMHMM, LipoP, and Phobius showed results that this protein is found in the outer cell membrane. Module 4 added onto this information by showing that there is a multitude of Transmembrane helices throughout this particular gene (Figure 4). Overwhelming evidence forced the conclusion that this protein is most likely where the computer predicted, and that it also performs the functions previously mentioned.

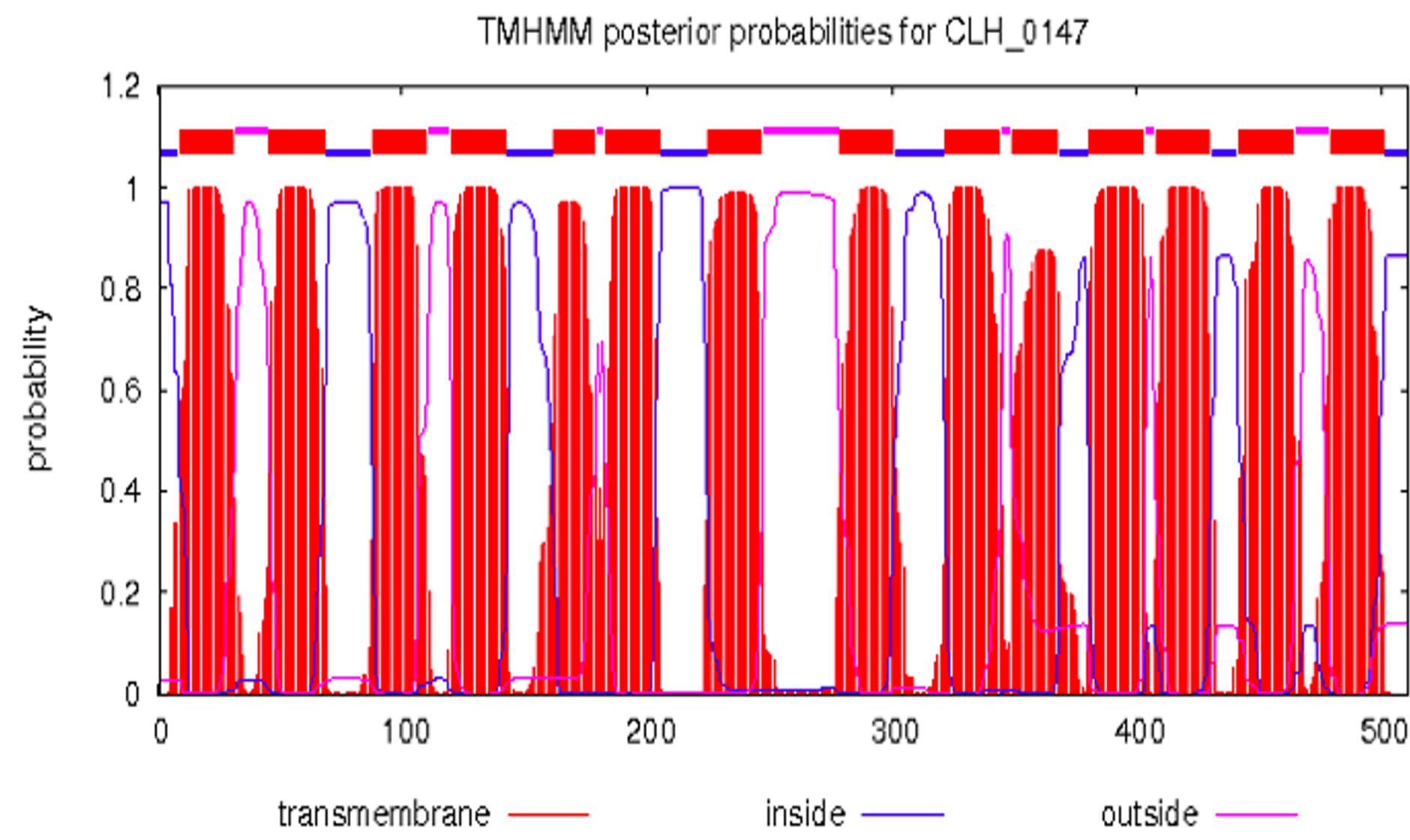


Figure 4 – CLH_0147 image from the results of TMHMM. The graph clearly shows the protein entering and exiting the cell membrane.

CLH_0148 (Sebastian Baco):

The computer pipeline proposed product of this gene was a MazG family protein. This was supported by the top BLAST, TIGRFAM, and Pfam hits for the amino acid sequences. Results from InterPro suggest that the MazG family is a “regulator of oxidative stress and nutritional stress responses” [2]. The location of this protein is not clear, as TMHMM and Phobius predicted it as non cytoplasmic, while LipoP and PSORT-B predicted it as cytoplasmic.

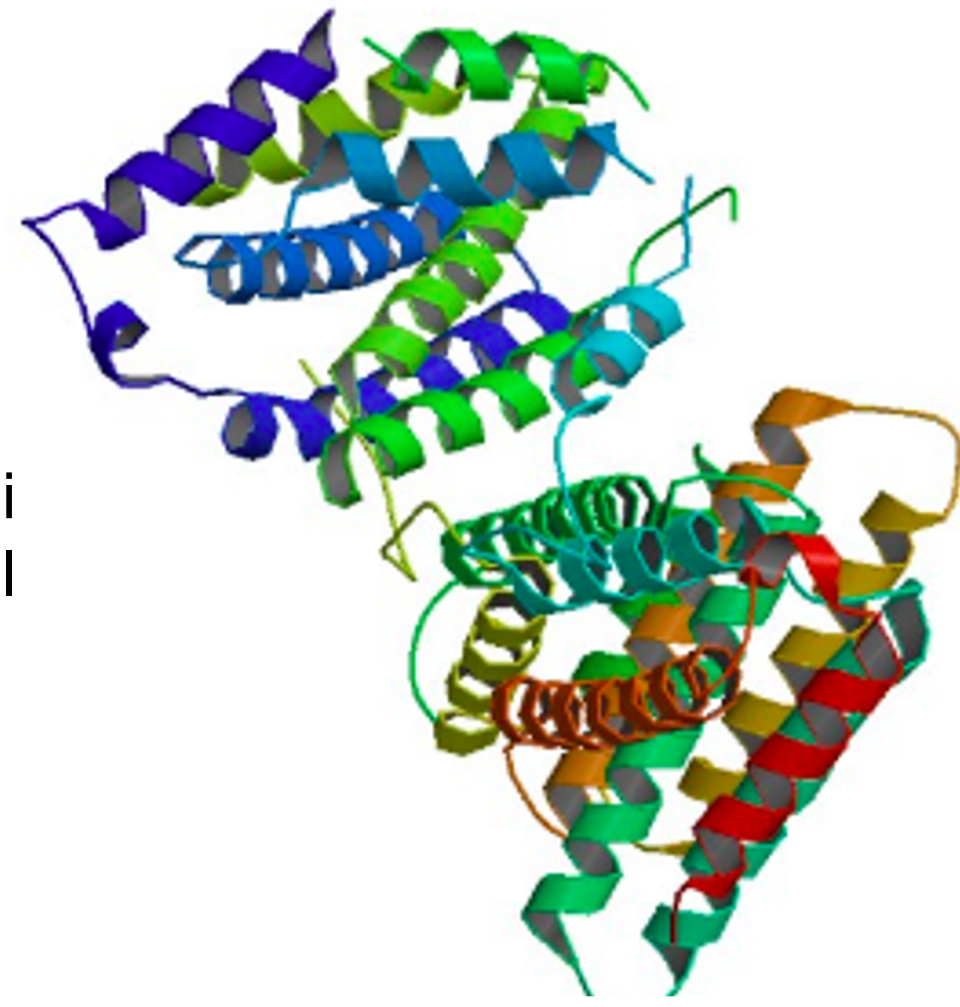


Figure 5 - CLH_0148
Crystal Structure of Escherichia coli MazG, the Regulator of Nutritional Stress Response
<http://www.rcsb.org/structure/3CRA>

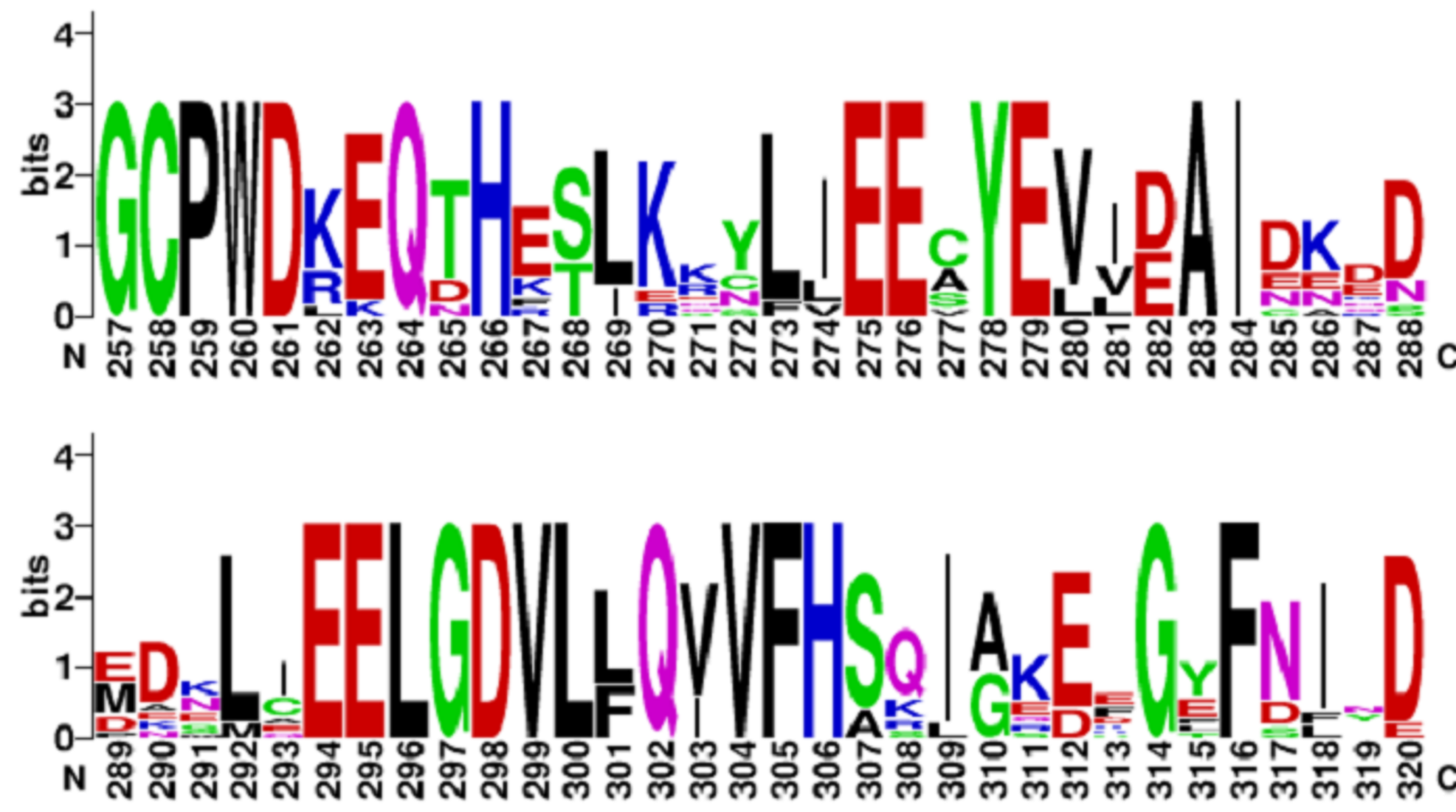


Figure 6 - CLH_0148
Segment of WebLogo of the sequence alignment of significant orthologs from the BLAST search
<https://weblogo.berkeley.edu/>

CLH_0153 (Nathaniel Grosskopf):

The computer pipeline proposed product of this gene was a cell division protein. Cell division is the process of a singular cell dividing into two of the same cells(daughter cells) [3]. This proposal was supported by the top BLAST hits for the amino acid sequence as well as the Pfam hit. CLH_0153 wasn't present in multiple databases, such as TIGRMFAM, PDB, and only one Pfam hit, which are curated databases. This means there is not a lot of researched information on this protein.

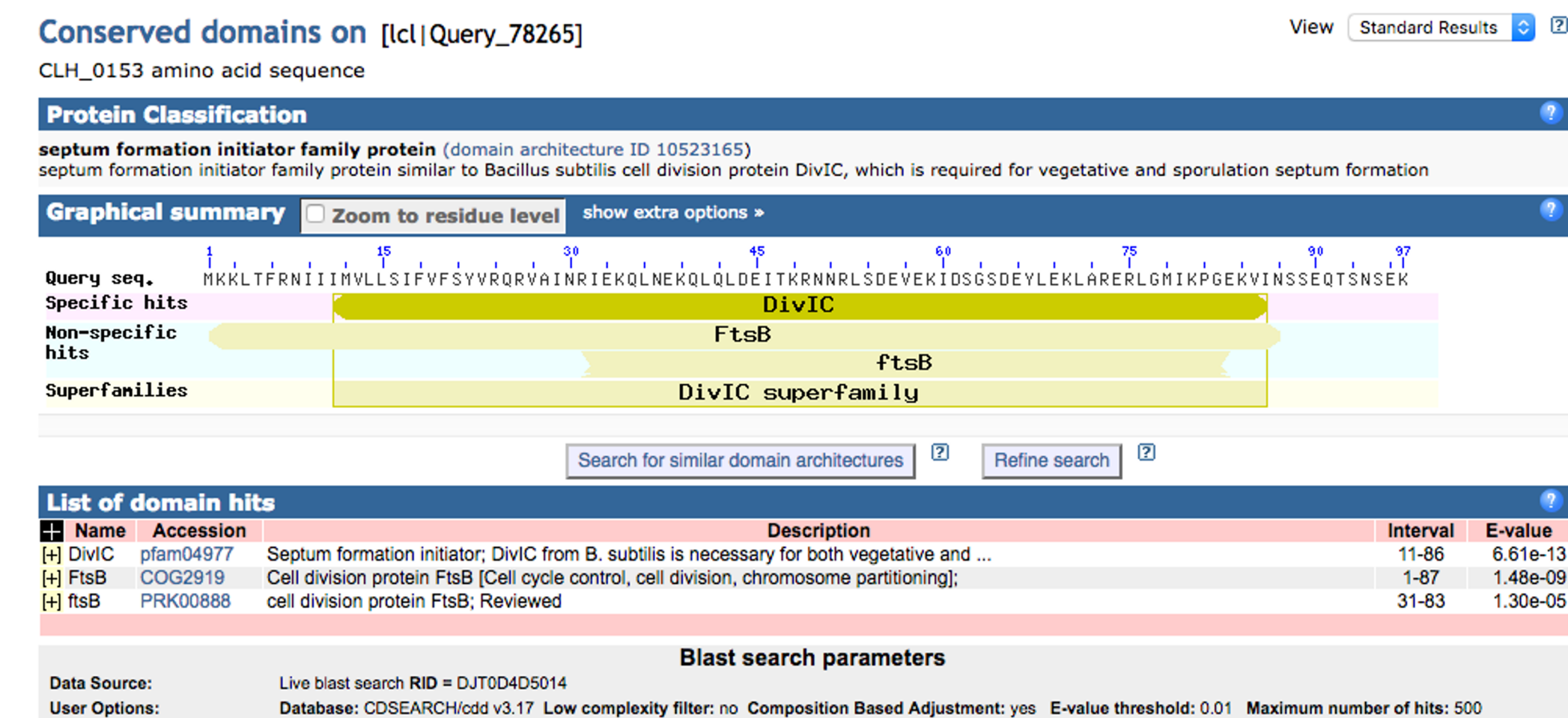


Figure 7 - CLH_0153 image of the Pfam and TIGRFAM BLAST hits. The image shows the lack of TIGRFAM which supports the lack of research on this protein.

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_0152	spore cortex biosynthesis protein YabQ	spore cortex biosynthesis protein YabQ	No
CLH_0147	Stage V sporulation protein B	Stage V sporulation protein B	No
CLH_0148	MazG family protein	MazG family protein	No
CLH_0153	Cell division protein	Cell division protein	No

References

- [1]https://www.fsis.usda.gov/wps/portal/ffsis/topics/food-safety-education/get-answers/food-safety-fact-sheets/foodborne-illness-and-disease/clostridium-botulinum/ct_index
- [2] <http://www.ebi.ac.uk/interpro/entry/IIPR011551>
- [3] <https://biologydictionary.net/cell-division/>
- [4]<https://www.encyclopedia.com/science-and-technology/biology-and-genetics/cell-biology/sporulation>

Acknowledgments

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