Annotation of the *Clostridium perfringens* Genome at Locus Tags CPF_1662, CPF_1667,and CPF_1673

Sumaiya C., Nadya E., Bushra N., and Todd Richards
City Honors School, Buffalo, NY and the Western New York Genetics in Research and Health Care Partnership

SEPA SCIENCE EDUCATION PARTNERSHIP AWARD
Supported by the National Institutes of Health

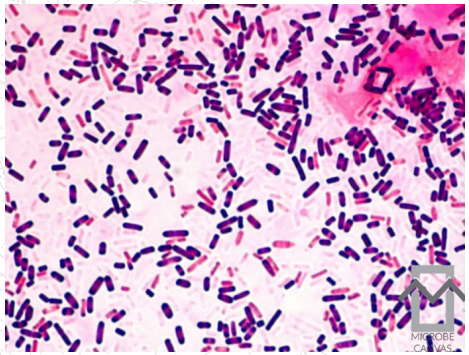
University at Buffalo

Abstract

Three genes from the organism *Clostridium perfringens* (CPF_1662, CPF_1667,CPF_1673) was interpreted using using Geni-act, which is a collaborative annotated genome website. The gene was evaluated based on essential genomic information, similarities to any amino acid sequences in the Genebank, cellular localization data, alternative reading frames, enzyme function, and gene duplication or denigration. The product name for the gene was also predicted by the Genebank. There were differences discovered between the Genebank's annotation and the results after testing the experiment.

Introduction

Clostridium perfringens is an anaerobic, gram positive, pathogenic bacterium. This bacteria is present everywhere in nature and it can be found in decaying vegetation, marine sediments, insects and even soil. This bacteria is commonly carried by food and is the third most common causes of food poisoning in the US. Fortunately, it is not too severe but sometimes It is mistaken for the 24 hour flu. This bacterium can also cause infections. These infections show evidence of tissue necrosis, or cell death. Bacteremia , or the presence of bacteria in the blood is also caused by clostridium perfringens. This bacteria also causes Gas gangrene which is a bacterial infection where gas is produced in the tissues.



Photomicrograph of gram stained *Clostridiumm perfringens* (1)

Modules of the GENI-ACT (http://www.geniact.org/) were used to complete *Clostridium perfringens* 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?
Module 9- RNA	RFAM	Does my gene encode a functional RNA?

Results

CPF_1662:

The initial proposed product was voltage- gated chloride channel family protein. This gene product was supported by the Blast hits for amino acid sequence. Voltage gated channels display a lot of important physiological and cellular roles that include the regulation of pH, volume homeostasis, organic solute differentiation and so on. Voltage gated chloride channels are important for setting cell resting membrane potential and maintaining proper cell volume. However, after doing the experiment. the top hit gene product name that came up after using BLAST was H(+)/Cl(-) exchange transporter ClcA. The organism that came up was *Salmonella enterica* subsp. This organism is a subspecies of *Salmonella enterica*, the rod shaped, flagellated, aerobic, Gram- negative bacterium. This organism also contains serovars which is a distinct variation within a species of bacteria that can infect a range of vertebrate hosts. The pathogenic serovars of *S. enterica* species are in this sub species which includes the ones accountable for typhoid.

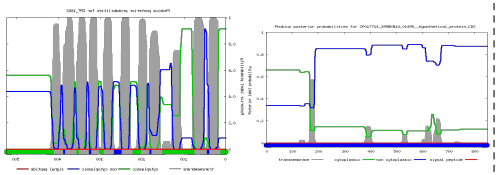
CPF_1667:

The gene of this product that was originally proposed was the aminotransferase, class V. However, while conducting this experiment the gene product was concluded to be Cysteine desulfurase IscS. Evidence that supported this was the top hits of the amino acid sequence based BLAST and the multiple sequence alignment. Both the aminotransferase and the Cysteine desulfurase are two very different enzymes. The aminotransferase is used to catalyze a reaction between an amino acid and an a-keto acid and is important when synthesizing proteins. The Cysteine desulfurase transfers sulfur containing groups and participates in thiamine metabolism.

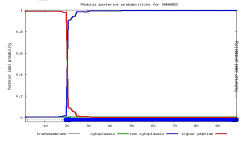
CPF_1673:

Originally the proposed gene product was the HD domain protein, but after conducting the experiment it was inferred to be the Multifunctional CCA protein. This was supported by the top hits of the amino acid based BLAST. The HD domain protein is a conserved protein domain. It is found in a superfamily of enzymes with a predicted or known phosphohydrolase activity. These enzymes appear to be involved in nucleic acid metabolism, signal transduction and possibly other functions in bacteria, archaea and eukaryotes. On the other hand the Multifunctional CCA protein is involved in RNA binding. It Catalyzes the addition and repair of the essential 3'- terminal CCA sequence in tRNAs without using a nucleic acid template.

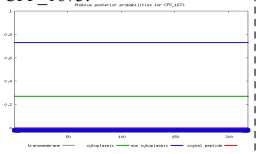
Phobius Probability Graph:This graphs combines both methods used in the SignalP and TMHMM.CPF_1662



CPF_1667:



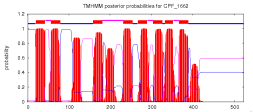
CPF_1673:



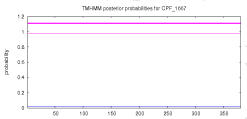
Transmembrane Topography map:

These graphs shows the probability of the amino acids on the x-axis being located on the transmembrane. The pink lines are the predicted amino acids outside the membrane while the blue is the predicted inside the membrane.

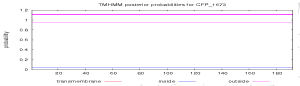
CPF_1662:



CPF_1667:



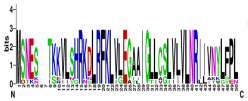
CPF_1673:



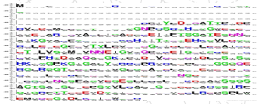
Weblogo:

This logo shows the conservation from multiple sequence alignments.

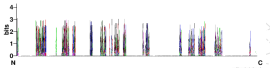
CPF_1662



CPF_1667:



CPF_1673:



Conclusions

CPF_1662 CPF_1673 CPF_1667

The proposed gene products differed significantly from the gene products identified in the results of our experiments. After analyzing our results it could be seen that the proposed organism in CPF_1667, which was the aminotransferase class V, differed from the gene product which was concluded to be Cysteine desulfurase IscS. In CPF_1662 the proposed product was voltage- gated chloride channel family protein, although after analyzing the BLAST results the product was identified as H(+)/Cl(-) exchange transporter ClcA. Lastly in CPF_1673 the proposed organism was the HD domain protein, yet after analyzing our results it came out to be the Multifunctional CCA protein.

References

1. http://microbe-canvas.com/Bacteria/anaerobic-gram-positive-rods/spores-positive/lecithinase-positive-2/indole-negative/clostridium-perfringens.html

Acknowledgments

Supported by an NIH Science Education Partnership (SEPA) Award - R25ODO10536

www.buffalo.edu