Annotation of the Clostridium perfringens Genome from Locus Tags CPF_1683, CPF_1670, CPF_1672,CPF_1682

Maureen Ferguson, Sabrina Lonas, Ryan Melber, Kiera Quinlivan, Mr. Richards
City Honors School PS 195 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

We studied CPF_1683, CPF_1670, CPF_1682, and CPF_1672 on the computer program GENI-ACT. We were meant to annotate the genes using various computer programs to verify the gene product. We would be able to determine if the computer program made an error. We used websites, models, and programs such as Pfam, BLAST, PSORT-B, HMM logo, and the KEGG pathway ID to analyze the gene. As a result, we detected no errors made by the computer program. The data we collected correlated with the predicted gene product.

Introduction

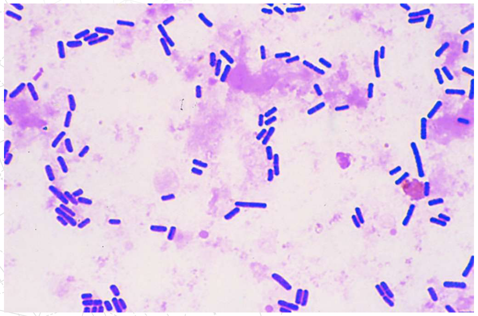
Clostridium perfringens is a bacteria that grows spores. It is found in various environments such as soil and animal intestines. It is capable of surviving with limited oxygen, and grows quickly at room temperature. This bacteria is a widespread pathogen. *Clostridium perfringens* can cause up to one million food poisoning incidents each year.

The gene CPF_1683 is a major cardiolipin synthase. It is part of the cardiolipin synthetase protein family. The gene is responsible for the metabolism of lipids and phospholipids. It is found in the organism bacillus subtilis.

The gene CPF_1672 is a sodium/alanine symporter. It is part of the sodium/alanine symporter family. It is responsible for the transport of sodium, ions, and amino acids. It is found in the organism methanococcus maripaludis.

The gene CPF_1670 is a homocysteine. It is part of the homocysteine S-methyltransferase family. This gene is responsible for amino acid transport and metabolism. It is found in the organism bacillus subtilis.

The gene CPF_1682 is a relA/spoT protein. It is part of the HMM PF04607 protein family. It is responsible for nucelotide transport and amino acid binding. It is found in the organism bacillus subtilis.



Gram stain of *Clostridium perfringens*, illustrating that it is a gram positive bacterium

Methods

Modules of the GENI-ACT (http://www.geni-act.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?

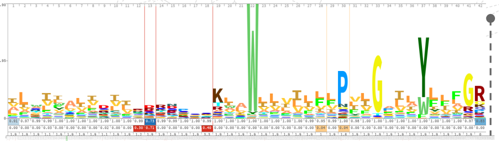
Results

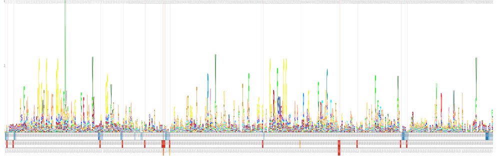
CPF_1683:
The initial proposed product of this gene by GENI-ACT was a cardiolipin synthase. This gene product proposal was supported by the top BLAST hits for the amino acid sequence, the Pfam results, the HMM logo, and the PSORT-B. As such, the proposed annotation is a cardiolipin synthase.

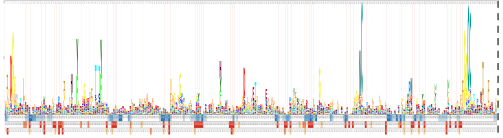
CPF_1670:
The initial proposed product of this gene by GENI-ACT was a homocysteine. This gene product proposal was supported by the top BLAST hits for the amino acid sequence, the Pfam results, the HMM logo, and the PSORT-B. As such, the proposed annotation is a homocysteine.

CPF_1672:
The initial proposed product of these genes by GENI-ACT was a sodium:alanine symporter. This gene product proposal was supported by the top BLAST hits for the amino acid sequences, the Pfam results, the HMM logo, and PSORT-B. The proposed annotation is a sodium/alanine symporter.

CPF_1682:
The initial proposed product of this gene by GENI-ACT was a relA/spot protein. This gene product proposal was supported by the top BLASTT hits for the amino acid sequence, the Pfam results, the HMM logo,and the PSORT-B. The proposed annotation is a relA/spot protein.







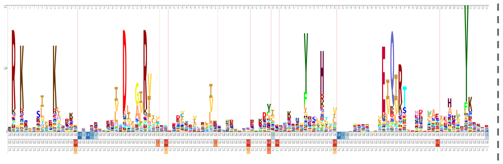


Figure 1 – CPF_1683 is compared to the cardiolipin synthase family using an HMM logo. It shares several amino acids with other members of its family. For example, the amino acids represented by the W (sea green), G (yellow), and Y (dark green).

Figure 2 – CPF_1672 is compared to the sodium/alanine synthase family using an HMM logo. There are many amino acids in the same locations of others in its family. For example, the amino acids represented by G(yellow), Y (dark green), and W (sea green).

Figure 3-CPF_1670 is compared to the homocysteine S-methyltransferase family using an HMM logo. There are many amino acids in the same locations of others in its family. For example, the amino acids represented by D(red) and Y(dark green).

Figure 4- CPF_ 1682 is compared to the HMM PF04607 protein family using an HMM logo. It shares several amino acids with other members opf family. For example, the amino acids represented by the R (maroon), Y (brown, and D (red).

PSORT-B Evidence:

CPF_1683: This gene received a 9.51 cytoplasmic membrane score on the PSORT-B. Since this gene synthesizes lipids and phospholipids, it is a rational conclusion that it would be located in the cytoplasmic membrane. The cytoplasmic membrane is composed of two phospholipid layers.

CPF_1670: This gene received a 7.50 score for cytoplasmic on the PSORT-B. Since this gene is responsible for metabolic processes, it would be logical that it is located in the cytoplasm. The function of the cytoplasm is to perform metabolic activity.

PSORT-B continued:

CPF_1672: The gene received a 10.0 cytoplasmic membrane score on the PSORT-B. Since this gene is predicted transport sodium and amino acids, it is likely that it is located in the cytoplasmic membrane. The cytoplasmic membrane is responsible for the transport of molecules into and out of the cell.

CPF_1682: This gene received a 7.5 score for cytoplasmic on the PSORT-B. This gene is responsible for nucleotide transport and the function of the cytoplasm is to transport molecules through the cell.

CPF_1683:
The initial proposed product of this gene was a major cardiolipin synthase. This was supported by the top BLAST hits for the amino acid sequence, the presence of well-curated functional domains within the amino acid sequence, the cellular location of the amino acid sequence, and the enzymatic function of the amino acid sequence.

CPF_1672:
The initial proposed product of this gene was a sodium/alanine symporter.. This was supported by the top BLAST hits for the amino acid sequence, the presence of well-curated functional domains within the amino acid sequence, the transmembrane topography of the amino acid sequence, the cellular location of the amino acid sequence, and the enzymatic function of the amino acid sequence.

CPF_1670:
The initial proposed product of this gene by GENI-ACT was a homocysteine. This gene product proposal was supported by the top BLAST hits for the amino acid sequence, the presence of well-curated functional domains within the amino acid sequence, the transmembrane topography of the amino acid sequence, the cellular location of the amino acid sequence, and the enzymatic function of the amino acid sequence.

CPF 1682:
The initial proposed product of this gene by GENI-ACT was a homocysteine. This gene product proposal was supported by the top BLAST hits for the amino acid sequence, the presence of well-curated functional domains within the amino acid sequence, the transmembrane topography of the amino acid sequence, the cellular location of the amino acid sequence, and the enzymatic function of the amino acid sequence.

Conclusions

The GENI-ACT proposed gene product did not differ significantly from the proposed gene annotation for CPF_1683, CPF_1670, CPF_1682, and CPF_1672 in the group and as such, the genes appear to be correctly annotated by the computer database.

Acknowledgments

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

www.buffalo.edu