Annotation of the *Bacillus anthracis* Genome at Locus Tags BA_0709 and BA_0710

Kelvin Cruz, AJ Rice, Emily Stoker and Rick Salton*

Gates Chili High School and The Western New York Genetics in Research and Health Care Partnership

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

Abstract

A group of two genes from the microorganism *Bacillus anthracis* (BA_0709-BA_0710) were annotated using GENI-ACT (http://geni-act.org/). The Genbank proposed gene product 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, potential alternative open reading frame and enzymatic function. 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 by in the database. The sequence data found was run through numerous sites in order to determine the name and the function of the proteins under investigation. The work done was over a span of a few weeks. The work done found that the Locus tags 0709 and 7010 were found to be both spore germination proteins GerLA and GerLB.

Introduction

Bacillus anthracis is a gram-positive, endospore forming, and rod-shaped bacterium. This organism can create mediums in either aerobic or even anaerobic conditions. *Bacillus anthracis* it was discovered and credited to a German physician named Aloys Pollender, *B. anthracis* was known bacterium to be demonstrated to cause disease . *B. anthracis* toxins are shown to be in tea, polyphenols as an example. The ability to be resistant to heat, disinfectants, and drying have shown that it can be used as biological weapons in either powdered or aerosol forms . It is known to belong to the *Bacillus cereus* group along with five other strains besides itself. Endospore staining is used to visualize bacterial endospores. These spores are made so the bacteria can survive in hostile conditions. The spores are resistant to heat, radiation, dissection, and chemicals. Gram staining is used to differentiate two large groups of bacteria based on their different cell wall constituents.

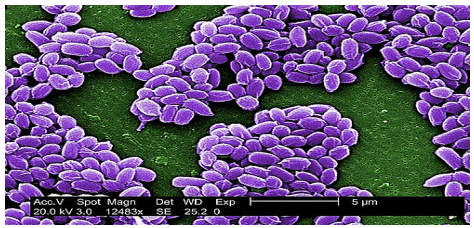


Figure 1. A picture of *Bacillus anthracis* under a electron microscope

Methods

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?

The tools we used in the GENI-ACT modules helped us identify the proteins under investigation in *Bacillus anthracis*. For module one, we found the sequences for the proteins and gene and their location in the genome. In module two, the proteins wasn't really similar to other sequences in the Genbank. In module three, the proteins localized to the membrane of the bacterium. In module four, The amino acid sequences was correctly called by the computer. In module five, there were functional domains for the proteins. Lastly, in module six, the proteins were not found not to be enzymes.

Results

BA_0710

The initial proposed product of this gene by GENI-ACT was a spore germination protein GerL.B. The Length of the DNA sequence is at 1125 nucleotides, encoding 374 amino acids. During the BLAST the protein encoded by Bant_0710 had a score of 711, with a 94% identity hit rate. Although, this gene product proposal was shown by the BLAST hits for the amino acid sequence, the presence of well-curated protein functional domains within the amino acid sequence, the transmembrane topography of the amino acid sequence, and the cellular location of the amino acid sequence. As such, the proposed annotation is a spore germination protein GerL.B.

BA_0709

The initial proposed product of this gene by GENI-ACT was a spore germination protein GerLA. *Bacillus anthracis* was determined to be a gram positive organism through a PubMed search. Also, it was noticed in the topography that the lines were sparatic yet were near each other. The score turned out to be 564.1 while the E-value was 1.1e-168. This gene product proposal was shown by the BLAST hits for the amino acid sequence, the presence of the functional domains within the amino acid sequence, the cellular location of the amino acid sequence, and also the enzymatic function of the

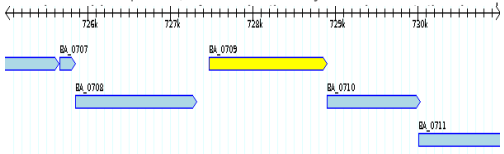


Figure 2. The gene neighborhood under investigation. The genes of oth BA_0709 and BA_0710 are found on the top strand of DNA.

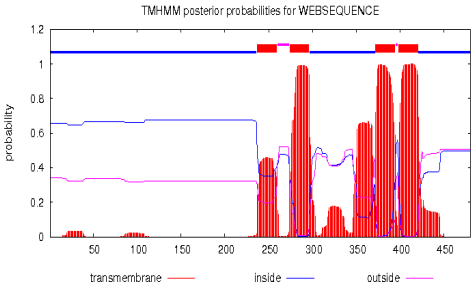


Figure 3. TMHMM results for BA_0709 predicting 4 transmembrane helixes.

```
Prediction of ba_0709

ID   ba_0709
FT   TOPO_DOM     1     273       CYTOPLASMIC.
FT   TRANSMEM   274     296
FT   TOPO_DOM   297     315       NON CYTOPLASMIC.
FT   TRANSMEM   316     334
FT   TOPO_DOM   335     345       CYTOPLASMIC.
FT   TRANSMEM   346     364
FT   TOPO_DOM   365     369       NON CYTOPLASMIC.
FT   TRANSMEM   370     389
FT   TOPO_DOM   390     395       CYTOPLASMIC.
FT   TRANSMEM   396     424
FT   TOPO_DOM   425     478       NON CYTOPLASMIC.
//
```

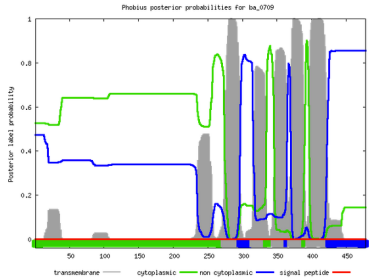


Figure 4. Phobius results for BA_0709 predicting 5 transmembrane helixes rather than the 4 predicted by TMHMM.

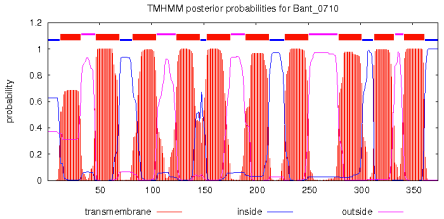
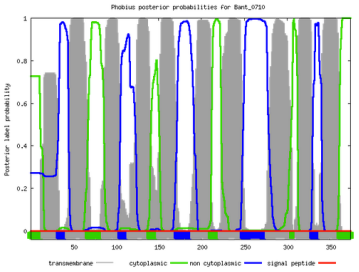


Figure 5. TMHMM results for BA_0710 predicting 10 transmembrane helixes.

```
ID   Bant_0710
FT   TOPO_DOM     1     15       CYTOPLASMIC.
FT   TRANSMEM    16     33
FT   TOPO_DOM    34     44       NON CYTOPLASMIC.
FT   TRANSMEM    45     67
FT   TOPO_DOM    68     86       CYTOPLASMIC.
FT   TRANSMEM    87    105
FT   TOPO_DOM   106    116       NON CYTOPLASMIC.
FT   TRANSMEM   117    139
FT   TOPO_DOM   140    150       CYTOPLASMIC.
FT   TRANSMEM   151    171
FT   TOPO_DOM   172    190       NON CYTOPLASMIC.
FT   TRANSMEM   191    211
FT   TOPO_DOM   212    222       CYTOPLASMIC.
FT   TRANSMEM   223    248
FT   TOPO_DOM   249    277       NON CYTOPLASMIC.
FT   TRANSMEM   278    306
FT   TOPO_DOM   307    312       CYTOPLASMIC.
FT   TRANSMEM   313    329
FT   TOPO_DOM   330    340       NON CYTOPLASMIC.
FT   TRANSMEM   341    361
FT   TOPO_DOM   362    374       CYTOPLASMIC.
//
```

Figure 5. Phobius results for BA_0710 predicting 10 transmembrane helixes, the same number as predicted by TMHMM.



Conclusions

Over all, the GENI-ACT proposed gene product did not differ significantly from the gene annotation found for each of the genes in the groups and as such, the genes were correctly annotated by the computer database and results were recorded in our GENI-ACT lab notebook. The data proved that the genes were spore germination genes in the *Bacillus anthracis* bacterium despite some errors in the module. So most of the results came out positive during the project for the bacterium. Also, the data shows that it is inside the cell membrane as well as partially sticking out of the cell. One difference noted was that there was some disagreement between Phobius and TMHMM in predicting the number of transmembrane helixes in BA_709. Wet lab experiments will be required to determine which tool has the correct prediction..

Reference

1.http://geni-act.org/lab_notebook/student/isolate_genome_gene/5e89e3a733d54928/eb9c0c359e1b4f3a/f5ae2af5b48642ac/

Acknowledgements

This work was supported by an NIH Science Education Partnership Award (SEPA)-R25ODO10536

*Names are listed in alphabetical order and not listed in contributing order

www.buffalo.edu