University at Buffalo The State University of New York | REACHING OTHERS

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

Annotation of the *Bacillus anthracis str. Ames* Genome from Locus Tags BA_0633 to BA_0634 and BA_0709 to BA_0711

Grace Collura*, Thomas Griffin*, Kristen Kibler*, Penelope Sergi*, Adam Wojtulski* and Gregory Jubulis (* indicates students equally contributed to gene annotation)

West Seneca West Senior High School, West Seneca, New York 14224 and The Western New York Genetics in Research and Health Care Partnership

Abstract

The group of genes in the locus tag interval BA_0634 to BA_0635 and BA_0709 to BA_0711 from the bacterium Bacillus anthracis str. Ames were annotated using GENI-ACT, a compilation of multiple genomic data bases that use FASTA sequences to deduce a gene's final protein function and name. The proposed gene name was assessed using the following GENI-ACT modules: Basic Information, Sequence Based Similarity, Structure Based Evidence, Cellular Localization Data, Evidence for Horizontal Gene Transfer, and Alternative Open Reading Frame. The final annotation of the gene deduced by the aforementioned modules was consistent with the proposed gene annotation made by GENI-ACT. The genes were called correctly.

Introduction

Bacillus anthracis str. Ames is a gram positive, virulent bacterium, that was isolated from a dead 14 month old cow from Sarita, Texas in 1981. This strain of anthrax is believed to earliest known strain to be used in laboratories around the world, making it a reference for other strains. This is one of 89 genetic variants of anthrax which became infamous, as it was utilized as a bioterror weapon in September of 2001. Letters laced with anthrax spores were sent through the United States resulting in 22 infections with five fatalities.

Robert Koch experimentally proved bacteria caused illness through his work with *Bacillus anthracis*. The bacteria produces a spore that is extremely resilient, which makes it fatal after germination. The capsular polysaccharide that surrounds the bacteria itself helps to protect it from possible threats such as the host's immune system responses. People are infected by anthrax when they come in contact with infected animals or contaminated animal products. Cutaneous, gastrointestinal, and inhalation are three types of anthrax infection with inhalation providing the optimal conditions required for spore germination, making it the most virulent.

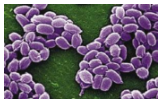

Figure I. Genomic sequencing of Bacillus anthracis can be used in microbial forensics.
Nature **454**, 813 (2008)

Genes of *Bacillus anthracis* can be studied using bioinformatics. Specific locus tags of the bacterium's DNA are analyzed to determine the coded protein's function, shape, and location in the cell. The primary component of the study is the compiling of the data found in multiple data bases and comparing them to each other in order to surmise the true intention of a particular gene. Each student was assigned a unique locus tag of a gene in the range BA_0633-BA_0634 and BA_0709-BA_0711.

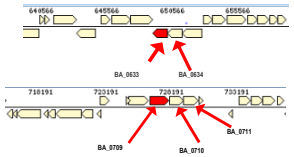

Figure II. Gene Neighborhoods of BA_0633-BA_0634 and BA_0709-BA_0711

Methods and Materials

Modules of the GENI-ACT (http://www.geni-act.org/) were used to complete *Bacillus anthracis str. AMES* 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- Structure-Based Evidence	TIGRfam, Pfam, PDB	Are there functional domains in my protein?
Module 4- Cellular Localization Data	Gram Stain, TMHMM, SignalP, PSORT, Phobius	Is my protein in the cytoplasm, secreted or embedded in the membrane?
Module 5- 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 8- Evidence for Horizontal Gene Transfer	Phylogenetic Tree,	Has my gene co-evolved with other genes in the genome?
Final Annotation	Review data from all modules	Does the student proposed name of the gene agree with that proposed by the automated computer annotation? Are any changes proposed to the pipeline annotation?

Table 1. Modules employed for gene annotations of BA_0633-BA_0634 and BA_0709-BA_0711

Results

Bacillus anthracis str. Ames; BA_0633: Based on the GENI-ACT prediction this protein is spore germination protein GerKC. The Web Logo showed that this protein is mildly conserved throughout the amino acid sequence. This is not an integral membrane protein because TMHMM predicted no transmembrane helixes. According to signal IP, a signal peptide is present at the terminal end of the protein suggesting it is secreted from the cell. This is further supported by the Phobius result stating that the protein in non-cytoplasmic.

Bacillus anthracis str. Ames; BA_0634 : The proposed function of the gene is to code for spore germination protein GerKB of the *Bacillus anthracis Ames strain*. The Web Logo created with the gene's FASTA Sequence shows that there is a high level of amino acid conservation present throughout the sequence. The data TMHMM tool indicates the predicted transmembrane helices in BA_0634. This leads us to believe the prediction of the gene's function is correct. The cytoplasmic membrane score of 10.0 as indicated by PSORT-B states the protein is most likely located within the cell membrane.

Bacillus anthracis str. Ames; BA_0709: The Web Logo displayed higher levels of conservation in multiple locations. HMM Logo showing a highly conserved region of the protein suggesting a functional role. PSORTB indicated a cytoplasmic membrane score of 10. SignalP showed that the protein is npt exported out of the cell. In conclusion, the gene context and chromosome viewer GC showed similar neighborhoods and distribution showing no evidence for horizontal gene transfer.

Bacillus anthracis; BA_0710: The data provided by TMHMM, PSORT-B and Phobius indicates the protein created by my gene is an integral component of the plasma membrane as it possesses multiple transmembrane helices and has a cytoplasmic membrane score of 10.0. Additionally, SignalP predicted the protein was not exported out of the cell. The Web Logo for this protein showed relatively high levels of conservation in multiple sections of the amino acid sequence. The top gene product of the Blast search for this protein was germination protein GerLB.

Bacillus anthracis; BA_0711: Based on the information found in PubMed, it was concluded that *Bacillus anthracis* is gram positive. TMHMM helped us determine that the protein is an integral membrane protein on account of the presence of transmembrane helices. SignalP alluded to the presence of a signal peptide, which suggests that our protein is secreted from the anthrax bacterium into the surrounding medium. The final predicted location for the protein is extracellular. This information was found with psortb, where the localization score for extracellular rested above a 7.5.

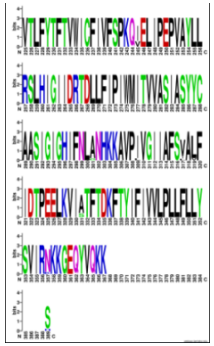

Figure III. BA_0634 Web logo. The produced Web Logo shows many large capital letters, which shows multiple contingencies and displays that genes have been preserved between different variants

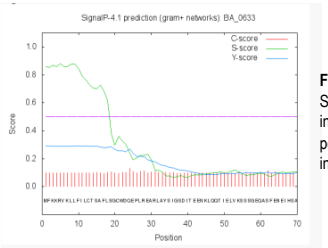

Figure IV. Signal IP result indicating signal peptide present in BA_0633.

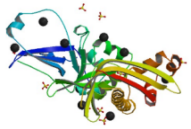

Figure V. BA_0711 Depiction showing crystal structure of the GerBC protein.

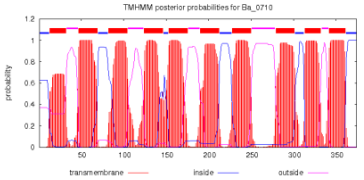

Figure VI. TMHMM results for BA_0710 predicting the presence of 10 transmembrane helices in the amino acid sequence of the protein it codes for.

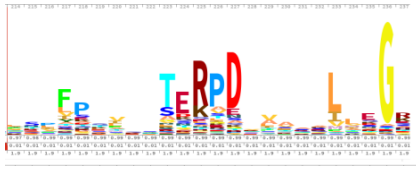

Figure VII – BA_0709 HMM Logo showing a highly conserved region of the protein suggesting a functional role.

Conclusion

The GENI-ACT proposed gene product did not differ significantly from the proposed gene annotation for each of the genes in the group. The genes appear to be correctly annotated by the computer database.

GENE LOCUS	GENI-ACT PRODUCTS	PROPOSED ANNOTATION
0633	spore germination protein GerKC	spore germination protein GerKC
0634	spore germination protein GerKB	spore germination protein GerKB
0709	spore germination protein GerLA	spore germination protein GerLA
0710	spore germination protein GerLB	spore germination protein GerLB
0711	spore germination protein GerLC	spore germination protein GerL

References

Rasko, D. A., Worsham, P. L., Abshire, T. G., Stanley, S. T., Bannan, J. D., Wilson, M. R., ... Ravel, J. (2011). *Bacillus anthracis* comparative genome analysis in support of the Amerithrax investigation. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(12), 5027–5032. http://doi.org/10.1073/pnas.1016657108

Acknowledgments

Special thank you to SUNY at Buffalo, Dr. Stephen Koury, and Dr. Rama Dey-Rao.
Supported by an NIH Science Education Partnership (SEPA) Award - R25OD10536-1 A1