Annotation of the *Bacillus anthracis* Genome from Locus Tags Bant_0053 to Bant_0055

Dylan DelConte, Tai Hartman, Zachary Hinkley, Corey Krug, Ashley Perry, Jessica Stock, Emily Wyszomirski, and Richard Salton*
Gates-Chili High School, Rochester, NY and the Western New York Genetics in Research and Health Care Partnership
*All students contributed equally

GATES CHILI SCHOOLS
SEPA SCIENCE EDUCATION PARTNERSHIP AWARD
Supported by the National Institutes of Health
University at Buffalo

Abstract

Three consecutive genes from the microorganism *Bacillus Anthracis* (Bant_0053 - Bant_0055*)* were annotated using the collaborative genome annotation website GENI-ACT. The gene product name for each gene was proposed by Genbank (NIH genetic sequence database). The objective of GENI-ACT was to assess these proposals. They were assessed in six modules, each looking to find/reconfirming information about the function of Bant_0053 - Bant_0055. In module one basic information about our genes, such as the sequence and location in genome was found. In module two programs such as BLAST, CDD, T-Coffee and Weblogo were used to determine if our sequences are similar to other proteins in Genbank. In Module three we determined weather our proteins are cytoplasmic, secreted or embedded in the membrane. In module four potential alternative open reading frames were assessed. In module five TIGRfam, Pfam and PDB were used to determine if there are functional domains in our proteins. Lastly in module six, using KEGG and MetaCyc the enzymatic function of our proteins was determined. Based on the information found from working through the modules, the proposed gene product name by Genbank for each protein seems to be correct. The gene product for BA_0053 is Stage V sporulation protein T, for BA_0054 Stage V sporulation protein B and for BA_0055 MazG family protein.

Introduction

Bacillus anthracis is an endospore forming, rod-shaped bacteria that has a gram-positive stain. Because this is gram-positive it shows it has a thick peptidoglycan layer, making it harder to eradicate. Most environments are hospitable because it can thrive in aerobic and anaerobic atmospheres. Among livestock it is a very common disease but also infects humans. *Bacillus anthracis* is hard to kill due to the fact it produces endospores. Aloys Pollender, a German physician, discovered the microbe and it was one of the first discovered to produce a disease such as Anthrax.

Bacillus anthracis was discovered in 1850, it was the first bacterium to be shown to cause a disease. There are three ways that anthrax can infect you. The main reason why it is important to research Bacillus anthracis is because the normal infection causes a small painless ulcer but if left untreated it can make its way into the blood stream and will cause death in 20% of infections. If inhaled the infection will resemble the common cold. This is important because it could possibly be used as a deadly bio-terrorism weapon that will be hard to detect and is fatal if left untreated due to the extremely resilient nature of the spores.

Geni-act is used to help identify functions of various loci on a certain genome. It utilizes different programs and databases to assemble an overview of what each specific locus does. By knowing what every locus does, we're closer to understanding the entirety of the virus and how to combat it.

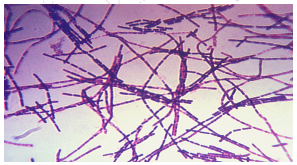


Figure 1. Bacillus anthracis gram positive test from: https://upload.wikimedia.org/wikipedia/commons/a/a1/Bacillus_anthracis_Gram.jpg

Methods

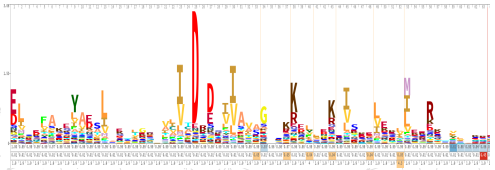
Modules of the GENI-ACT (http://www.geni-act.org/) were used to complete *Bacillus anthracis* genome annotation . The modules are described below:

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?

Results

Bant_0053:

Based on the databases in Geni-Act, locus BA_0053 is exclusively involved in the transcription of endospores. More specifically, it produced stage V sporulation protein T. This is supported with evidence from pairwise alignments with TIGRFAM and Pfam, which compared BA_0053 with a protein that deals with sporulation and germination processes in *Bacillus subtilis*. Also, this annotation is further proven by conserved domains found with WebLogo produced which received a near-perfect match with related species *Bacillus subtilis* in T Coffee. On top of this, TMHMM graphs documenting transmembrane topology assured the proper location for this kind of locus. Thus, the proposed annotation for this gene is stage V sporulation protein T.



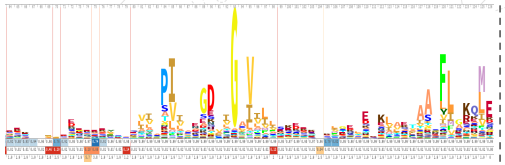


Figure 2 – The pairwise alignment of our amino acid sequence and that of Stage V sporulation protein T C-terminal, transcription factor produced by Pfam, showing a large amount of conserved domains between the two.

Bant_0054:

The top BLAST hit for BA_0054 was stage V sporulation protein B. In WebLogo, our sequence was overall very conserved with sections 135-152 being the most clearly conserved. The top TIFGRAM hit was also stage V sporulation protein B and the top Pfam hit was polysaccharide biosynthesis protein. The COD name was RfbX and the COG number is COG2244. RfbX is a Membrane protein involved in the export of O-antigen and teichoic acid. Our protein helps form the spore coat in stage five of the sporulation process.

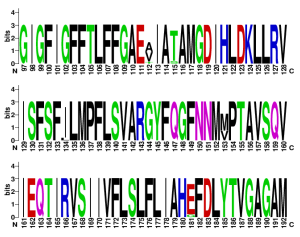


Figure 3 – BA_0054 was very conserved from sections 135-152 taken from web logo

Bant_0055:

The initial proposed product of BA_0055 by GENI-ACT was tetrapyrrole methylase family protein/MazGfamily protein. The top BLAST hit for BA_0055 was MazG Family Protein. The COG name for the protein was MazG. The number was COG1694. The top TIGRFAM hit was MazG: MazG nucleotide pyrophosphohydrolase and the top Pfam hit was Tetrapyrrole (Corrin/Porphyrin) Methylases. The second Pfam hit was MazG nucleotide pyrophosphohydrolase domain. The proposed DNA coordinates predicted by IGG are correct based on Alternative Open Reading Frame results. The E.C. number of our protein is 3.6.1.9 based on a search for MazG on Metacyc. Lastly based on the results for MazG in Metacyc and 3.6.1.9 in ExPASy the function for MazG and subsequently our protein is the hydrolysis of ATP and other nucleoside 5'triphosphates.

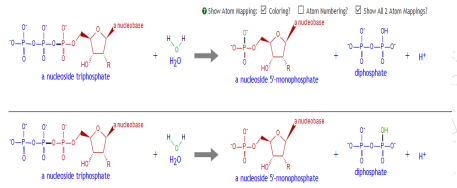


Figure 4 – This reaction shows the hydrolysis of a nucleoside triphosphate into a nucleoside 5'-monophosphate, a diphosphate and hydrogen. The enzyme MazG does this .

Conclusion

Through the utilization of the Genbank, GENI-ACT is able to determine the likely gene product of a pre-determined gene locus. The GENI-ACT proposed gene product proved very similar to the proposed gene annotation for the gene locus' below. This further validates that the genes have been correctly annotated by the computer database.

Gene Locus	GeniAct Gene Products	Proposed Annotation
Bant_0053	Stage V sporulation protein T	Stage V sporulation protein T
Bant_0054	Stage V sporulation protein B	Stage V sporulation protein B
Bant_0055	Tetrapyrrole methylase family protein/MazGfamily protein	MazGfamily protein

References

Boydston, J., 2006. The ExsY Protein Is Required for Complete Formation of the Exosporium of Bacillus anthracis. J Bacteriol 188(21): 7440–7448.

Slamti, L; Perchat, S; Gominet, M., 2004. Distinct mutations in PlcR explain why some strains of the Bacillus cereus group are nonhemolytic. Journal of Bacteriology 186 (11): 3531–8.

Acknowledgments

Supported by an NIH Science Education Partnership (SEPA) Award - R25ODO10536
Significant Contributor: Dr. Stephen Koury
www.buffalo.edu