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

Gates Chili Central School

NIH National Institutes of Health — SEPA SCIENCE EDUCATION PARTNERSHIP AWARD — SUPPORTED BY THE NATIONAL INSTITUTES OF HEALTH

# Annotation of the *Bacillus anthracis* Genome from DNA Coordinates 58190 to 62321


Gabriel Brett, Emma Daley, Grace Daly, Hanna DiNorma and Richard Salton

Gates Chili Central School and the Western New York Genetics in Research and Health Care Partnership


## Abstract


A group of 4 genes from *Bacillus anthracis* (BA_0053, BA_0054, BA_0055 and BA_0056 were subjected to manual annotation using the Genomics Education National Initiative- Genome Annotations Tool kit (GENI-ACT). GENI-ACT was used to perform a series of database search's that gave general genomic information, such as amino acid sequence-based similarity data, potential alternative open reading frames and others. We used test results to determine if the gene product name was identical to that proposed for each gene by automated computer pipeline annotation. BA_0053 was found to be a Stage V sporulation protein, BA_0054 a Polysaccharide biosynthesis protein, BA_0055 a nudeotide triphosphate pryophosphohydrotase and BA_0056 an S4 domain containing protein likely involved in binding to RNA


## Introduction

*Bacillus anthracis* is part of the genus of aerobic, immobile, gram-positive and to encapsulated spores. The spores of the bacteria have are highly resistance to environmental factors and can still be infectious after a long period of time. They can survive on a simple nutrient media. *Bacillus anthracis* is the antigen that causes anthrax. Anthrax is extremely harmful to humans and animals.

It is important to study genes especially ones belonging to infectious bacteria such as *Bacillus anthracis* so we have knowledge of how the bacteria functions and to find new antibodies that work more efficiently. In 1981 the Ames Strain was discovered. The knowledge of strain greatly bettered how vaccinations were made and also saved the lives of the many victims of the 2001 Anthrax attacks.

In the GENI-ACT program we were given locations of genes in the Ames Strain. Little in know about the locus of this strain. By matching up the unknown protein sin the genes with already known proteins, we can effectively infer what job the proteins perform.

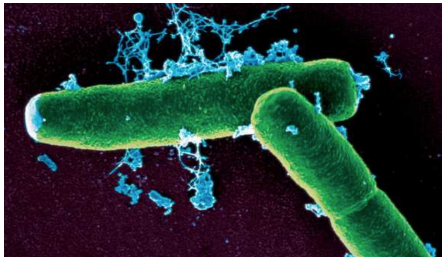

Figure 1. A scanning electron micrograph of *Bacillus anthracis* Taken from: https://microbewiki.kenyon.edu/index.php/File:Bacillus_anthracis_1.jpg

## Methods and Materials

The following modules were used in the GENI-ACT experiment that tested our four separate protein strands of *Bacillus anthracis*. Each strand was isolated and inserted into each sequencing data base to determine different information about each of them. Some data bases worked better than others and some found different results for different strands. There was a small average range but since the sizes of our strands differed, our information gained did as well, hence our GENI-ACT notebooks.

| 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

Bacillus anthracis 0053

The initial proposed product of this gene by GENI-ACT was a Stage V Sporulation Protein. This gene product proposal was supported by the top Swist_Prot BLAST hit for the amino acid sequence. The top COG hit was AbrB, a Bifunctional DNA-binding transcriptional regulator of stationary/sporulation/toxin gene expression and antitoxin component of the YhaV-PrlF to xin-antitoxin module. This protein is responsible for the positive and negative transcriptional regulator of sigma G-dependent genes. It may provide a mechanism of feedback control that is important for for spore development.

Score Expect Method Identities Positives Gaps Fra
355 bits(910) 3e-122() Compositional matrix adjust. 178/178(100%) 178/178(100%) 0/178(0%)
Features:

```
Query  1    MKATGIVRRIDDLGRVVIPKEIRRTLRIREGDPLEIFVDRDGEVILKKYSPISELGDFAK  60
            MKATGIVRRIDDLGRVVIPKEIRRTLRIREGDPLEIFVDRDGEVILKKYSPISELGDFAK
Sbjct  1    MKATGIVRRIDDLGRVVIPKEIRRTLRIREGDPLEIFVDRDGEVILKKYSPISELGDFAK  60

Query  61   EYAEALYDSLGHNVLVCDRDSIIAVSGVSKKEYLNKSVGDLIEKTMEERKSVIMTDESDV  120
            EYAEALYDSLGHNVLVCDRDSIIAVSGVSKKEYLNKSVGDLIEKTMEERKSVIMTDESDV
Sbjct  61   EYAEALYDSLGHNVLVCDRDSIIAVSGVSKKEYLNKSVGDLIEKTMEERKSVIMTDESDV  120

Query  121  SIIDGVTEKVHSYTVGPIVANGDPIGAVIIFSKEAIISEIEHKAVNTAASFLAKQMEQ  178
            SIIDGVTEKVHSYTVGPIVANGDPIGAVIIFSKEAIISEIEHKAVNTAASFLAKQMEQ
Sbjct  121  SIIDGVTEKVHSYTVGPIVANGDPIGAVIIFSKEAIISEIEHKAVNTAASFLAKQMEQ  178
```

Figure 2. The pairwise alignment of BA_0053 with its top hits. There were a large number of identical proteins from a number of *Bacillus* species that exactly matched the amino acid sequence of BA_0053.

Bacillus anthracis 0054

The top Swis_Prot BLAST hit for BA_0054 was uncharacterized membrane protein YabM. The top TIGRFAM hit was stage V sporulation protein B and the top PFAM hit was polysaccharide biosynthesis protein, which has a capD-like domain. This domain knocks out mutants that tin crease bacterial colonization. This capsule is a protective structure surrounding some species of bacteria, such as *Bacillus anthrasis*. These proteins are integral membrane proteins, and BA_0054 was found to have a total of 14 transmembrane helixes by TMHMM (Figure 4).

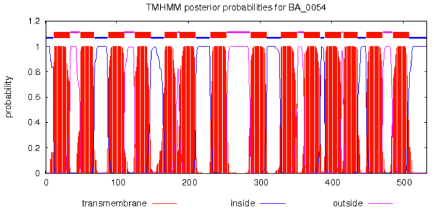


Figure 4. TMHMM results for BA_0054

Bacillus Anthracis 0055

The initial proposed product of this gene by GENI-ACT was a nudeoside triphosphate pyrophosphohgdrolase. This gene product proposal was supported by the top blast hits for the amino acid sequence (Figure 5). The presence of well-curated functional domains within the amino acid sequence. Therefore, the proposed annotation is a nudeoside triphosphate pyrophosphohydrolase.


Figure 5- Top Swis_Prot BLAST hit for BA_0055: This image shows how the query supported the idea that the gene codes for nucleoside triphosphate pyrophosphohydrolase

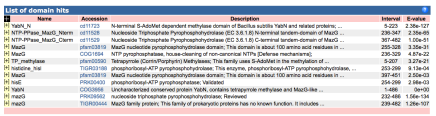


Figure 6. The Conserved Domain Database results for BA_0055 showing numerous domains consitent with BA_0056 being a nudeoside triphosphate pyrophosphohydrolase

Bacillus Anthracis 0056

The top COG hit was Ribosomal 50S subunit-recyding heat shock protein, contains S4 domain and the top PFAM hit was S4 domain containing protein (Figure 7), even though the top nr BLAST hit was multispecies hypothetical protein (Figure 8). The S4 domain is a small domain consisting of 60-65 amino acid residues that was detected in the bacterial ribosomal protein S4, probably mediates binding to RNA..

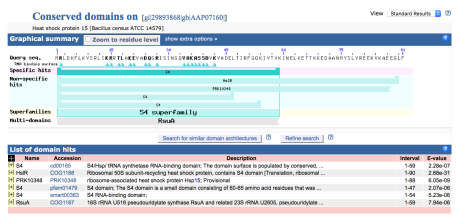


Figure 7. Conserved Domain Database results for BA_0056

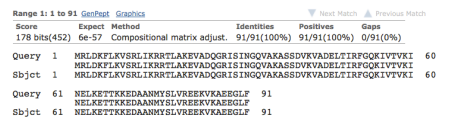


Figure 8. Top nr BLAST hit for BA_0056

## Conclusion serve

The concluding analysis of the gathered information of the geni-act enzyme and proteins vary widely. For locus tag BA_0054, the related protein linked to be Polysaccharide biosynthesis protein, which has a CapD domain. This protein attacks leukocytes and reacts with the formation of polysaccharides that make up the capsule. For locus tag BA_0055, the related enzyme is nudeotide triphosphate pryophosphohydrotase, which catalyzes chemical reactions. For locus tag BA_0053, the related protein is sporulation protein T, which its molecular and biological function is to bind DNA and regulate the transcription of DNA templates. Finally, for locus tag BA_0056, the protein only indicated that protein had an S4 domain and was likely involved in binding to RNA..

## References


UniProt, 2002-2016, UniProtKB-P3755 (SP5T_BACSU), *UniProt Consortium*, http://www.uniprot.org/uniprot/P3754 *(May 24th, 2016)*

*Bacillus Anthrax*. (2015, September 01). Retrieved May 24, 2016 From http://www.cdc.gov/anthrax


## Acknowledgments


Supported by an NIH Science Education Partnership (SEPA) Award - R25OD010536