

Annotation of the *Bacillus anthracis* strain A0248 at Locus Tags BA_0067, BA_0041, BA_0069, BA_0102 and BA_0043

Alana Curtis, Taylor Gadd, Viktoria Hamilton, Abigail Noonan, Megan Weber, Laura O'Donnell, Elizabeth Winters
Attica High School and The Western New York Genetics in Research Partnership



Abstract

A group of 5 genes from the microorganism *Bacillus anthracis* were annotated using the collaborative genome annotation website GENI-ACT. The Genbank proposed gene product name 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, and cellular localization data. 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 the computer database.

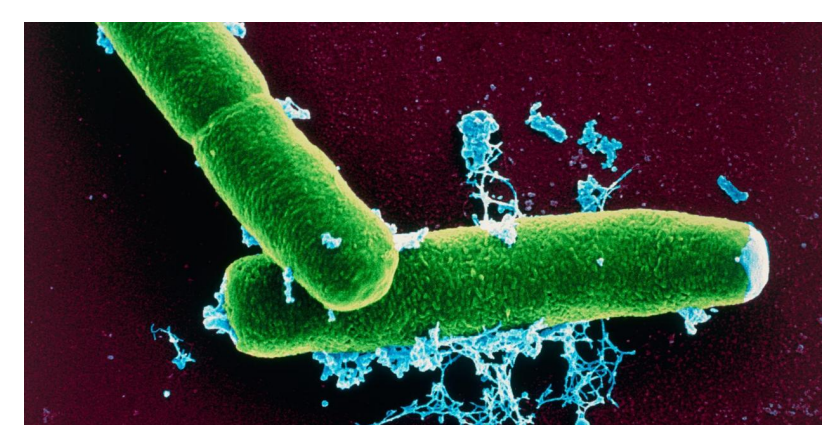
Introduction

Bacillus anthracis is a rod-shaped, gram-positive bacterium that forms endospores. It can be grown in either anaerobic or aerobic conditions in an ordinary nutrient medium. *B. anthracis* is one of the very few species of bacteria to synthesize a protein capsule (poly-D-gamma-glutamic acid) which allows it to be highly resilient, surviving extremes of temperature, low-nutrient environments, and harsh chemical treatment over decades or centuries. The bacteria is able to remain dormant for many years, however once it reaches a favorable environment it will begin to grow again.

The endospores created by the bacteria can resist heat, drying, and many disinfectants, even up to 95% ethanol. Due to the resiliency of the endospores, they are extraordinarily well-suited to use in biological warfare. It can be turned into a powdered form, and reside in water, or other salty environments.

Once the bacteria enter the body and proliferate, the host will start to show flu-like symptoms such as, sore throat, muscle aches, shortness of breath, and even coughing up blood. Most cases of anthrax can be treated with antibiotics however, if the symptoms appear and a person continues to not get treated, the bacteria will continue to rapidly replicate and it will continue to spread throughout the whole body. Since the bacterium is able to survive in many extreme environmental conditions, it is very hard to try and eradicate the bacteria completely.

Scanning electron micrograph of *Bacillus anthracis*
Credit: Dr Gary Gaugler/Science Photo Library
bbc.com



Methods and Materials

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?

Results

BAA_B0067		
Basic Information	DNA Coordinates	complement(53190..53333)
	DNA Length	144 bp
	Amino Acid Sequence Length	47 aa
Sequence Based Similarity	COGs	no COG hits
	TCOFFEE	not well conserved
	WEBLOGO	not well conserved
Cellular Localization	Transmembrane Helices	one
	Signal Peptide	none
	PSORTB Final Prediction	Cytoplasmic membrane
Structure Based Evidence	TIGRFAMs	No significant TIGRFAMs
	PFAMs	No significant PFAMs
	HMM Logo Key Residues	none
	PDB	No PDB results

Bacillus anthracis BAA_B0067:

Although many of the databases had no results, data from the BLAST search predicts BAA_B0067 to be involved with polyglutamate capsule biosynthesis (protein CapE). The results showed only bacteria from the genus *Bacillus* since it is one of the few bacteria that make this type of protective protein capsule and is therefore unique. Cellular Localization Data predicts this protein to be found on the cytoplasmic side of the membrane with one transmembrane helix.

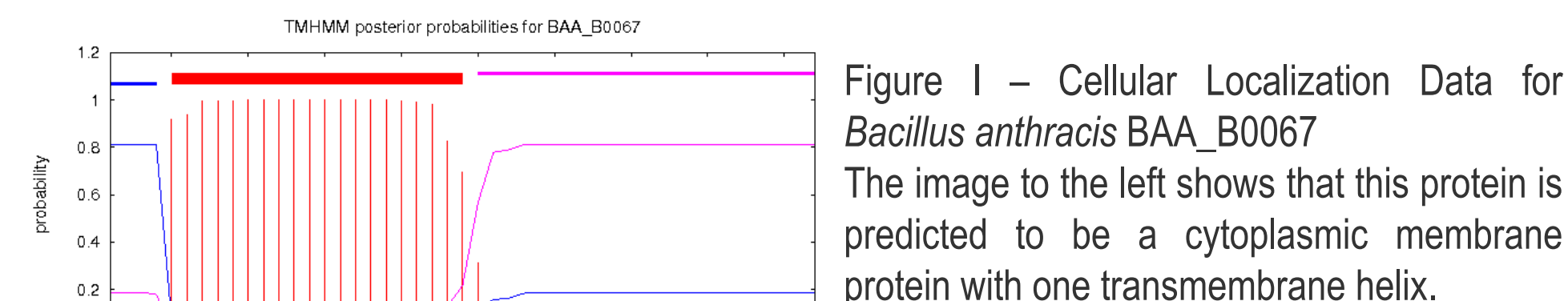
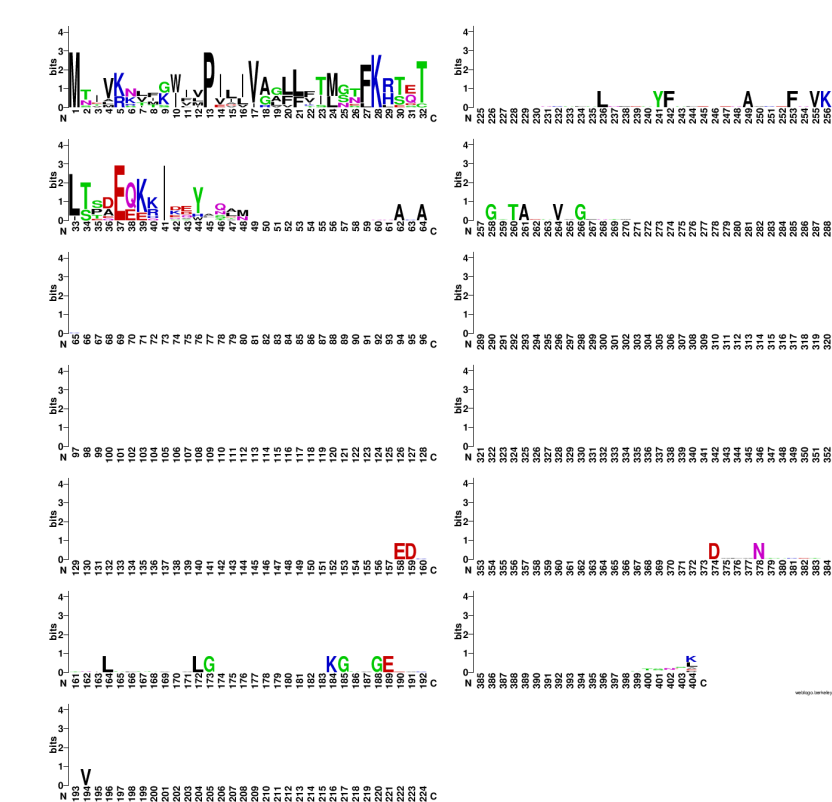


Figure II – WebLogo Data for *Bacillus anthracis* BAA_B0067
The image to the right shows the protein is not well conserved.

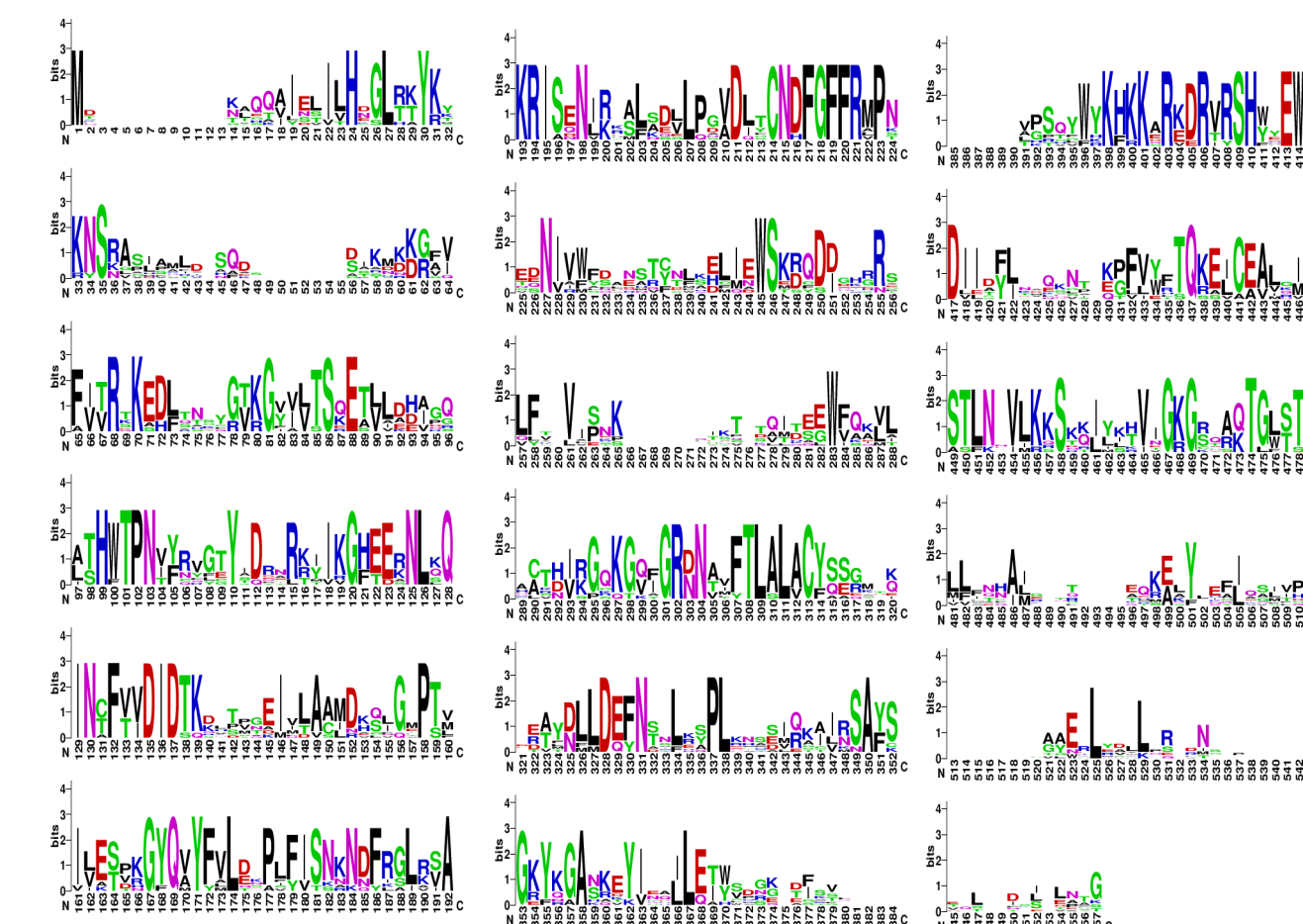


BAA_B0041		
Basic Information	DNA Coordinates	complement(33130..34668)
	DNA Length	1539 bp
	Amino Acid Sequence Length	512 aa
Sequence Based Similarity	COGs	No significant COGs
	TCOFFEE	Many orthologs and mostly well conserved sporadically well conserved throughout the sequence, but not at the beginning or end
	WEBLOGO	
Cellular Localization	Transmembrane Helices	none
	Signal Peptide	none
	PSORTB Final Prediction	Cytoplasmic
Structure Based Evidence	TIGRFAMs	No significant TIGRFAMs
	PFAMs	Primase C terminal 1 (PriCT-1)
	HMM Logo Key Residues	R13, S60
	PDB	No PDB results

Bacillus anthracis BAA_B0041 :

Blast results and PFAM data supports BAA_B0041 to be a C terminal primase involved in DNA replication. PSORTB data predicts the protein to be found within the cytoplasm since that is where DNA replication occurs within the cell. The presence of many orthologs indicate this protein is similar to other proteins in other species primarily because DNA replication is common in bacteria and can occur often.

Figure III – WebLogo Data for *Bacillus anthracis* BAA_B0041
The image to the right shows little conservation of the protein at the C and N terminals and more conservation in the middle.



BAA_B0069		
Basic Information	DNA Coordinates	complement(54931..56166)
	DNA Length	1236 bp
	Amino Acid Sequence Length	411 aa
Sequence Based Similarity	COGs	COG2843 Poly-gamma-glutamate biosynthesis protein CapA/YwtB (capsule formation), metallophosphatase superfamily [Cell wall/membrane/envelope biogenesis]
	TCOFFEE	Slightly conserved in the middle
	WEBLOGO	Some sporadic conservation throughout
Cellular Localization	Transmembrane Helices	one
	Signal Peptide	none
	PSORTB Final Prediction	Cytoplasmic membrane
Structure Based Evidence	TIGRFAMs	No significant TIGRFAMs
	PFAMs	Calcineurin-like phosphoesterase superfamily
	HMM Logo Key Residues	G6, G112, Y248, G251
	PDB	No significant results found

Bacillus anthracis BAA_B0069 :

COG and PFAM data supports BAA_B0069 to be a poly-gamma-glutamate protein involved with the outer capsule formation. This allows *B. anthracis* to reside in more extreme environments such as one with a high salt concentration. Cellular Localization Data predicts this to be a membrane protein with one transmembrane helix. There were few orthologs indicating that this is a protein unique to *B. anthracis*.

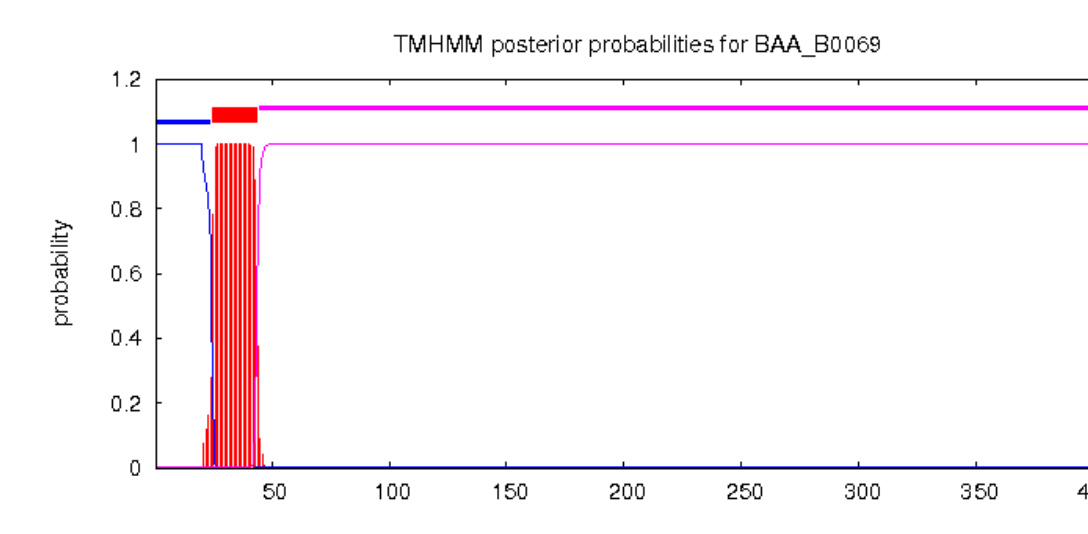


Figure IV – Cellular Localization Data for *Bacillus anthracis* BAA_B0069
The image to the left shows that this protein is predicted to be a cytoplasmic membrane protein with one transmembrane helix.

BAA_B0102		
Basic Information	DNA Coordinates	82880..84505
	DNA Length	1626 bp
	Amino Acid Sequence Length	541 aa
Sequence Based Similarity	COGs	No significant COGs
	TCOFFEE	Less conserved in the beginning then well conserved throughout the rest
	WEBLOGO	Well conserved in the middle
Cellular Localization	Transmembrane Helices	none
	Signal Peptide	none
	PSORTB Final Prediction	Unknown
Structure Based Evidence	TIGRFAMs	No TIGRFAMs
	PFAMs	SIR2-like domain
	HMM Logo Key Residues	T5, T6, N7, Y8, D9, G52, Y67, L91, F92, G94, S96, D99
	PDB	No PDB

Bacillus anthracis BAA_B0102 :

Blast results and PFAM data supports BAA_B0102 to be SIR2-like protein that assists with the binding of FAD and NAD in various cell processes, such as cell respiration. Although the results from PSORTB database shown an unknown final prediction, this could mean that the protein is found in multiple locations throughout the cell. There were many orthologs which means that it is a common protein found in a variety of other bacterial species.

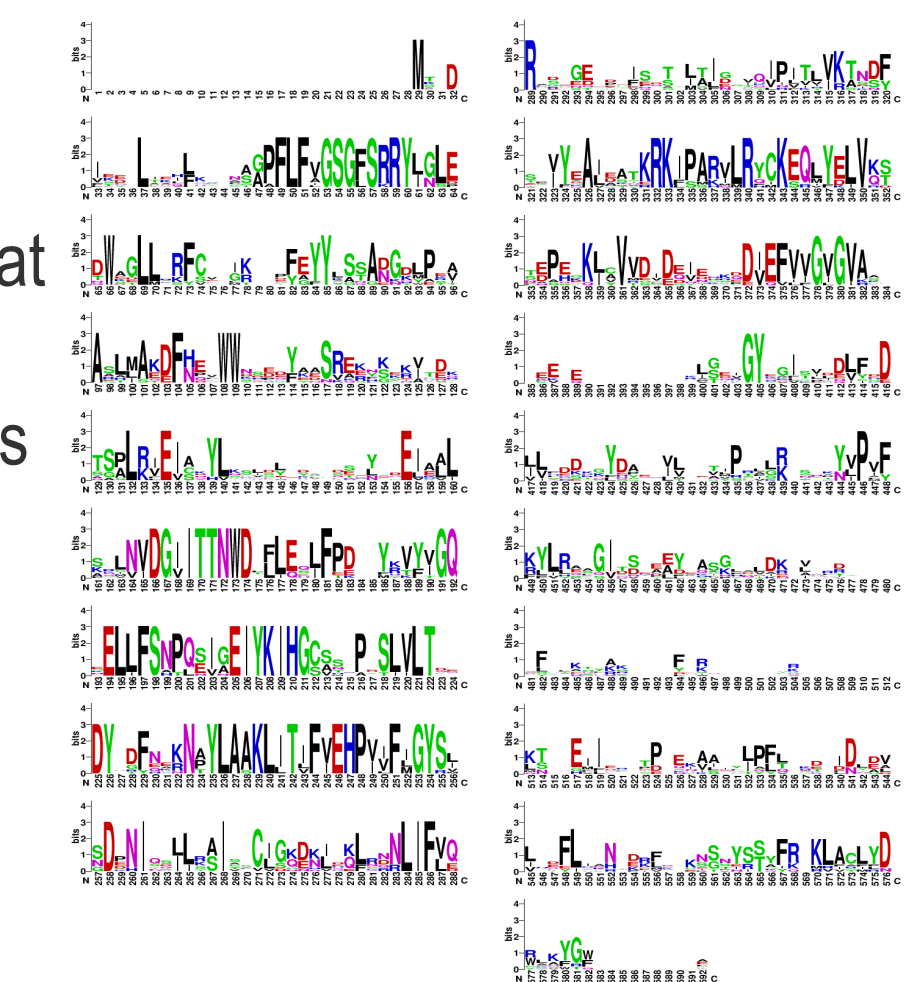


Figure V – WebLogo Data for *Bacillus anthracis* BAA_B0102
The WebLogo image above shows the protein is well conserved throughout the middle.

BAA_B0043		
Basic Information	DNA Coordinates	35574..36440
	DNA Length	867 bp
	Amino Acid Sequence Length	288 aa
Sequence Based Similarity	COGs	COG1192: Cellulose biosynthesis protein BcsQ [Cell motility]
	TCOFFEE	Mostly well conserved
	WEBLOGO	Well conserved throughout the middle
Cellular Localization	Transmembrane Helices	none
	Signal Peptide	none
	PSORTB Final Prediction	Cytoplasmic
Structure Based Evidence	TIGRFAMs	No significant TIGRFAM
	PFAMs	AAA domain
	HMM Logo Key Residues	G11, G12, D39, P127
	PDB	Crystal structure of Delta protein of pSM19035 from <i>Streptococcus pyogenes</i>

Bacillus anthracis BAA_B0043 :

COG, PDB and PFAM data supports BAA_B0043 to be involved in sporulation initiation. The dormant spores preserve the genetic material when conditions are inhospitable. Cellular Localization Data predicts this protein to be in the cytoplasm where sporulation occurs.

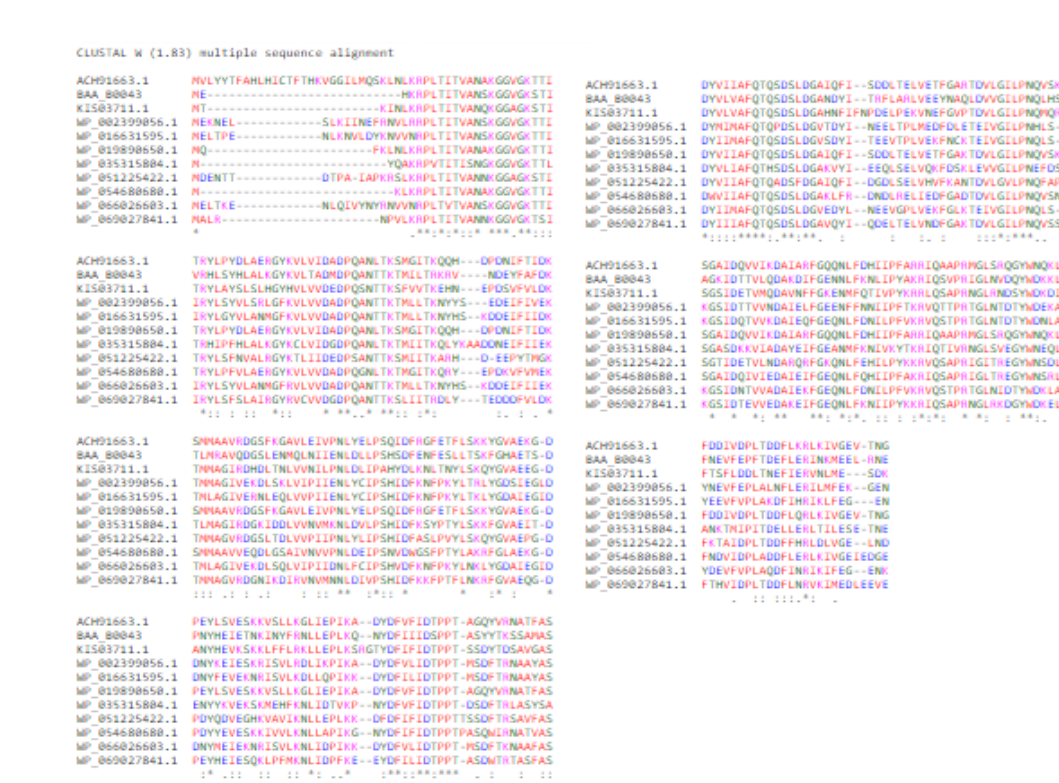


Figure VI – T-Coffee Data for *Bacillus anthracis* BAA_B0043. The image above shows conservation throughout most of the protein.

Conclusion

The Geni-Act proposed gene product 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 the computer database.

Gene Locus	Proposed Annotation
0067	Polyglutamate capsule biosynthesis (protein CapE)
0041	Primase C terminal 1 (PriCT-1)
0069	Poly-gamma-glutamate biosynthesis protein CapA/YwtB
0102	SIR2-like protein
0043	Cellulose biosynthesis protein BcsQ

References

Bacillus anthracis. (2017, April 26). Retrieved May 20, 2017, from https://en.wikipedia.org/wiki/Bacillus_anthraxis

Acknowledgments

Supported by NSF ITEST Strategies Award Number 1311902