Annotation of the *Bacillus Anthracis* Genome from Locus Tags BA_0632 to BA_0635

Nicolas Gravelle*, Kiana Ray*, Hunter Talley*, Jacob Tonello* and Melanie Roth Erie 1 BOCES at Maryvale Work Readiness and The Western New York Genetics in Research and Health Care Partnership *indicates the same effort

Abstract

A group of 4 consecutive genes from the microorganism Bacillus anthracis str. Ames (BA_0632 - BA_0635) 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 str. Ames is a gram-positive, rod-shaped bacterium. This bacterium can make spores that can be inhaled, acquired through skin lesions by contact with infected animal hides, or ingestion of contaminated meats to produce gastrointestinal anthrax. *B. anthracis* makes a good biological weapon because it can be made in large quantities. Inhalation anthrax is more lethal than skin lesions and gastrointestinal anthrax. If a person has been exposed to anthrax, they should be given a course of antibiotic treatment (Gast, et al., 2011).



Figure 1. The locus tags and relative position of the genes under investigation in this research



https://decodingscience.Missouri.edu/tap/bacillus-anthracis/

Methods

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
Basic Information	DNA Coordinates and Sequence, Protein Sequence	What is the sequence of the gene and protein? Where is it located in the genome?
Sequence-Based Similarity	Blast, CDD, T- Coffee, WebLogo	How similar is the protein under investigation to other proteins in Genbank?
Structure-Based Similarity	TIGRfam, Pfam, PDB	What functional domains are present in the protein under investigation?
Cellular Localization	Gram Stain, TMHMM, SignalP, LipoP, Psortb, Phobius	Is the protein under investigation located in the cytoplasm, secreted, in the periplasm or embedded in the cell membrane or cell wall?
Final Annotation	Evaluate data from all modules	Has the gene been correctly called by the pipeline annotation?

Results

BA_0632				
Basic Information	DNA Coordinates	648309649970		
	DNA Length	1662		
	Amino Acid Sequence Length	553		
Sequence Based Similarity	COGs	COG0366 [carbohydrate transport and metabolism]		
	TCOFFEE	well conserved in middle, poorly conserved on amino and carboxy ends		
	WEBLOGO	well conserved in middle, not well conserved on amino or carboxy end		
Cellular Localization	Transmembrane Helices	no transmembrane helix found		
	Signal Peptide	no signal peptide		
	PSORTB Final Prediction	protein is cytoplasmic		
Structure Based Evidence	TIGRFAMs	TIGR02403 trehalose_treC: alpha,alpha-phosphotrehalas		
	PFAMs	PF00128 Alpha amylase, catalytic domain		
	HMM Logo Key Residues	G1, H129		
	PDB	5BRQ: Entity 1 containing Chain A, B, C, D		

BA_0632	MKDWHKSVVYQIYPKSFNSYYNKETGDIKGVTEKLDYLKELGVDYIW
CCG47173.1	MIEQPWWKKAVVYQIYPKSFNDTTGNGVGDIQGIIEKLDYLNELGVDVLW
COE53185.1	MKDWHKSVVYQIYPKSFNSYYNKETGDIKGVTEKLDYLKELGVDYIW
SHP64325.1	MT-QPWWKKAVVYQIYPKSFNDTTGNGVGDIKGIIEKLDYLATLGVDVVW
WP_012575044.1	MHPWWKTAVVYQIYPKSFKDTNGDGIGDLQGVIEKLDYLKMLGVDAIW
WP_019416513.1	MHPWWKTAVVYQIYPKSFKDTNGDGIGDLQGVIQKLDYLQTLGVDAIW
WP_035299168.1	ME-KPWWKKAVVYQIYPKSFKDTTGSGTGDLQGVMEKLDYLRELGVDVIW
WP_062754046.1	MKQQPWWKKAVVYQIYPKSFKDTNGDGIGDLQGIIEKLDYLKLLGVDVIW
WP_075037985.1	ME-QPWWKKAVVYQIYPKSFNDTTGNGVGDIQGIIEKLDYLKKLGVDVLW
WP_080861970.1	MK-QPWWKKSVVYQIYPKSFKDTTGNGIGDLQGIIEKLDYLKTLGIDVIW
WP_081161885.1	MEQQPWWKKAVVYQIYPKSFNDTNGDGIGDLPGIIEKLDYLKLLGVDVIW
WP 097354486.1	MTKQPWWKKAVVYQIYPKSFKDTNGDGIGDLQGIIEKLDYLKLLGVDVIW
-	* *:::*******: **: *::**************
_	* *:::******:: :: *::*:*: *::**********
- BA_0632	* *:.:*********************************
- BA_0632 CCG47173.1	* *:.:*********************************
	* *:.:*********************************
BA_0632 CCG47173.1 COE53185.1 SHP64325.1	* *:.:*********: **: *: :***** **:* :* LTPIYQSPQNDNGYDVSDYYSIDPSYGTMEEFEELLEEAKARNIEIMLDI LTPMYKSPQNDNGYDISDYYDIHDEYGTMQDFEQLLAETHKRGMKLIMDL LTPIYQSPQNDNGYDVSDYYSIDPSYGTMEEFEELLEEAKARNIEIMLDI LTPIYESPQRDNGYDISDYFRIYEPYGKMEDFDSLLKEAHNRGIKIIMDI
	* *:.:*********: **: *::***** **:* :* LTPIYQSPQNDNGYDVSDYYSIDPSYGTMEEFEELLEEAKARNIEIMLDI LTPMYKSPQNDNGYDISDYYDIHDEYGTMQDFEQLLAETHKRGMKLIMDL LTPIYQSPQNDNGYDVSDYYSIDPSYGTMEEFEELLEEAKARNIEIMLDI LTPIYESPQRDNGYDISDYFRIYEPYGKMEDFDSLLKEAHNRGIKIIMDI LTPIYESPQRDNGYDISDYFRIYEPYGTMEDFENLIKEAHKRNIKMVMDI
	* *:::********************************
	* *:::********************************
	* *:::*********: **: *::**** **:* :* LTPIYQSPQNDNGYDVSDYYSIDPSYGTMEEFEELLEEAKARNIEIMLDI LTPMYKSPQNDNGYDISDYYDIHDEYGTMQDFEQLLAETHKRGMKLIMDL LTPIYQSPQNDNGYDVSDYYSIDPSYGTMEEFEELLEEAKARNIEIMLDI LTPIYESPQRDNGYDISDYFRIYEPYGKMEDFDSLLKEAHNRGIKIIMDI LTPIYESPQRDNGYDISDYFRIYEPYGTMEDFENLIKEAHKRNIKMVMDI LTPIYESPQRDNGYDISDYYRIYEPYGTMEDFEKLIEEAHKRNIKIVMDI LTPIYPSPQRDNGYDISDYCGIDPAYGTMETFERLLQEAHARGIRVIMDI LTPIYKSPQRDNGYDISDYFQIHDEYGTMEDFDRLLEEVHRRGMKLIMDM
<pre>BA_0632 CCG47173.1 COE53185.1 SHP64325.1 WP_012575044.1 WP_019416513.1 WP_035299168.1 WP_062754046.1 WP_075037985.1</pre>	* *:::*********: **: *::**** **:* :* LTPIYQSPQNDNGYDVSDYYSIDPSYGTMEEFEELLEEAKARNIEIMLDI LTPMYKSPQNDNGYDISDYYDIHDEYGTMQDFEQLLAETHKRGMKLIMDL LTPIYQSPQNDNGYDVSDYYSIDPSYGTMEEFEELLEEAKARNIEIMLDI LTPIYESPQRDNGYDISDYFRIYEPYGKMEDFDSLLKEAHNRGIKIIMDI LTPIYESPQRDNGYDISDYYRIYEPYGTMEDFENLIKEAHKRNIKMVMDI LTPIYESPQRDNGYDISDYYRIYEPYGTMEDFEKLIEEAHKRNIKIVMDI LTPIYPSPQRDNGYDISDYCGIDPAYGTMETFERLLQEAHARGIRVIMDI LTPIYKSPQRDNGYDISDYFQIHDEYGTMEDFDRLLEEVHRRGMKLIMDM LTPMYKSPQNDNGYDISDYFDIHEEYGTMEDFERLLQETHDRGMKLIMDM
BA_0632 CCG47173.1 COE53185.1 SHP64325.1 WP_012575044.1 WP_019416513.1 WP_035299168.1 WP_062754046.1 WP_075037985.1 WP_080861970.1	* *:::********: **: *::**** **:* :* LTPIYQSPQNDNGYDVSDYYSIDPSYGTMEEFEELLEEAKARNIEIMLDI LTPMYKSPQNDNGYDISDYYDIHDEYGTMQDFEQLLAETHKRGMKLIMDL LTPIYQSPQNDNGYDVSDYYSIDPSYGTMEEFEELLEEAKARNIEIMLDI LTPIYESPQRDNGYDISDYFRIYEPYGKMEDFDSLLKEAHNRGIKIIMDI LTPIYESPQRDNGYDISDYFRIYEPYGTMEDFENLIKEAHKRNIKMVMDI LTPIYESPQRDNGYDISDYYRIYEPYGTMEDFEKLIEEAHKRNIKIVMDI LTPIYPSPQRDNGYDISDYYRIYEPYGTMEDFEKLIEEAHKRNIKIVMDI LTPIYPSPQRDNGYDISDYFQIHDEYGTMEDFERLLQEAHARGIRVIMDI LTPIYKSPQRDNGYDISDYFQIHDEYGTMEDFDRLLEEVHRRGMKLIMDM LTPMYKSPQNDNGYDISDYFDIHEEYGTMEDFERLLQETHDRGMKLIMDI LTPIYASPQRDNGYDISDYFSVHQDYGTMADFERLLAEAHQRGIKVIMDM
<pre>BA_0632 CCG47173.1 COE53185.1 SHP64325.1 WP_012575044.1 WP_019416513.1 WP_035299168.1 WP_062754046.1 WP_062754046.1 WP_075037985.1 WP_080861970.1 WP_081161885.1</pre>	* *::*********************************

Figure 2 – T-COFFEE of BA-0632 is shown to be well conserved amongst many different rows. Rows with a * display that the amino acid is conserved.

BA 0632 Summary:

The initial proposed product of this gene by GENI-ACT was a trehalose-6-phosphate hydrolase. This gene product is supported by BLAST for the amino acid sequence had a low evalue and few gaps. It is also supported by results found in TIGRFAM, T-COFFEE and PDB. The proposed annotation of this protein is trehalose-6-phosphate hydrolase.



BA 0634

shows 10 ¦

transmem

domains.

brane

helix

buffalo.edu



The initial proposed product of this gene by GENI-ACT was spore germination protein GerKB. This gene product proposal is supported by WebLogo, which showed a well conserved region among the species. This annotation is also supported by PFAM and PSORTb. As such, the proposed annotation is a

BA_0635				
DNA Coordinates	652194653696			
DNA Length	1503			
Amino Acid Sequence Length	31			
COGs	No COGS found			
TCOFFEE	Amino portion is poorly conserved			
WEBLOGO	Residues 400-407 are well conserved			
Transmembrane Helices	3 transmembrane helices found			
Signal Peptide	no signal peptide found			
PSORTB Final Prediction	Cytoplasmic Membrane			
TIGRFAMs	None found			
PFAMs	PF03323 spore germination protein			
HMM Logo Key Residues	D12, F135, Y174, P252			
PDB	None found			
	BA_0635 DNA Coordinates DNA Length Amino Acid Sequence Length COGs TCOFFEE WEBLOGO Transmembrane Helices Signal Peptide PSORTB Final Prediction TIGRFAMs PFAMs HMM Logo Key Residues PDB			

Figure 5 – TMHMM shows that BA_0635 has 3 transmem brane helix domains.

The initial proposed product of this gene by GENI-ACT was spore germination protein GerKA. This gene product proposal is supported by BLAST, the top hit is a spore germination protein KA with a low E-value and few gaps. TMHMM shows the protein is integrated in the cell membrane. It is also supported by the results of T-COFFEE and PFAM. The proposed annotation is spore germination

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

cus	GENI-ACT Gene Products	Proposed Annotation
	trehalose-6-phosphate hydrolase	trehalose-6-phosphate hydrolase
	spore germination protein GerKC	spore germination protein GerKC
)	spore germination protein GerKB	spore germination protein GerKB
)	spore germination protein GerKA	spore germination protein GerKA

References

Gast, A. P., Relman, D. A., Casadevall, A., Connell, N. D., Inglesby, T. V., Kafadar, K., . . . Cline, A. (2011). Review of the Scientific Approaches Used During the FBIs Investigation of the 2001 Anthrax Letters. The National Academy of Sciences. doi:10.17226/13098

Acknowledgments Supported by an NIH Science Education Partnership (SEPA) Award - R25GM129209