Annotation of the *Brucella abortus* Genome from DNA Coordinates 2100703 to 2102352 (or Locus Tags BAA13334 to BAA103467)

Cora Blackwell, Olivia Kempf, Angela Ely, Kirk Winters
Pioneer High School, 12125 County Line Rd Yorkshire, NY 14173

PIONEER HIGH SCHOOL

University at Buffalo

Abstract

A group of 2 genes from the bacteria Brucella abortus (BAA13334_I03467 and BAA13334_I00104) 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, cellular localization data, and potential alternative open reading frames.

Introduction

Brucella abortus is a gram negative coccobacillus bacteria (short rod) that can only live and replicate in the body of its host. It is a heterotrophic, facultative bacteria that can perform aerobic and anaerobic respiration. Also, it is very hard to grow as it is a fastidious bacteria that requires nutrients from the host cell.

This bacteria is of extreme interest because it causes Brucellosis, a disease that results in the termination of a cows pregnancy, and if the disease spreads, a whole herd may be affected. The United States Government has spent millions of dollars to try to end diseases like this. Essentially, in a pregnant female cow, *Brucella abortus* localizes in the uterus during gestation and causes ulceration of the endometrium, resulting in the termination of the cow's pregnancy. In male cows, *Burcella abortus* causes the testes to swell and if left untreated, can lead to further complications. Thankfully, there is a vaccine for Brucellosis, and in humans in can be treated with antibiotics.

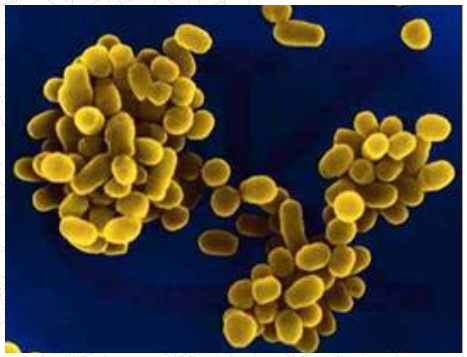


Figure 1. A microscopic view of the bacteria *Brucella abortus*

Methods

Modules of the GENI-ACT (http://www.geni-act.org/) were used to complete *Brucella abortus* genome annotation . These modules are described below, however, we were not able to complete all of them.

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?
Module 7- Gene Duplication/ Gene Degradation	Paralog, Pseudogene	Are there other forms of my gene in the bacterium? Is my gene functional?
Module 8- Evidence for Horizontal Gene Transfer	Phylogenetic Tree,	Has my gene co-evolved with other genes in the genome?
Module 9- RNA	RFAM	Does my gene encode a functional RNA?

Results

BAA13334_I03467

This protein is very likely cytoplasmic. This is supported by the database PSORT-B, who gave a predicted cytoplasmic score of 9.97. This protein is also very unique, because it is not commonly found in other organisms.

BAA13334_I00104

The initial proposed product of this gene by GENI-ACT was a histidine ammonia-lyase. This gene product proposal was supported by the top BLAST hits for the amino acid sequence, the presence of well-curated functional domains within the amino acid sequence, the cellular location of the amino acid

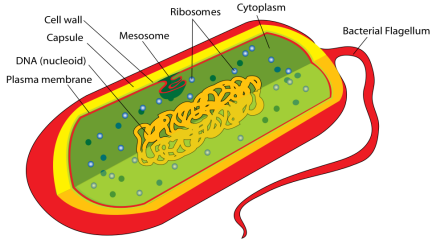


Figure 2. A diagram of a typical bacterial cell. Both genes would be located in the green, the cytoplasm.

Figure 3. A screenshot of the search results of A13334 on the Blast database NR, after excluding the organism Brucella (taxid 234). Note all the red, it shows that almost all of the bases were found in other organisms.

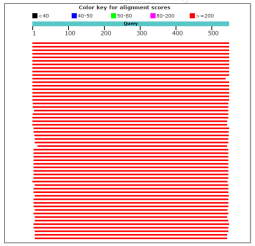


Figure 4. The website Blast's logo.

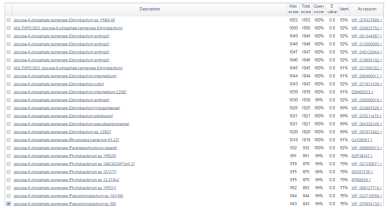


Figure 5. The search results of the same Blast search in figure 3. Note how the E-value for all the results is 0.0, and how the description for most all the results contains gluecose-6-phosphate isomerase

Figure 6. This is a WebLogo sequence for BAA13334_I00104 that indicates how well conserved the gene is. The amount of large letters shown here indicate good gene conservation.

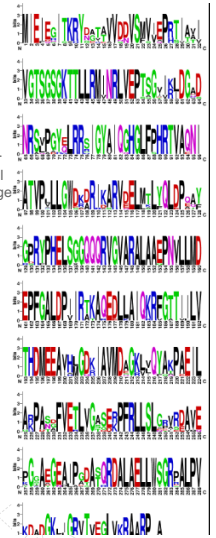


Figure 7. The lab notebook image of the amino acid sequence for A13334 , used for most research throughout the Geni-act process. This was what was inputted to most websites for research.

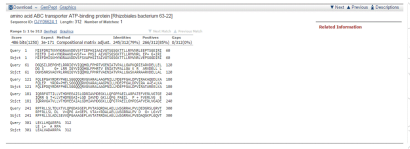
Protein Sequence

go to the Gene Page

Amino acid sequence

```
>BAA13334_I03467 amino acid sequence
MARDATKLEATVAKLKKHWAESAPRDMRAAFSTDPGRFGRYSLCLDDLLFDWSKCRVNDETMAL
LKELAVAADVEGRRAAMFAGEMINNTEDRAVLHVALRDTSSKEVLVDGHNVLPDVKHVLDRMAA
FADGIRSGALKGATGRKITDIVNIGIGGSDLGPVMATLALAPYHDEPRAHFVSNIDGAHIADTL
SPLDPASTLIIVASKTFTTIETMTNAQTARKWVADTLGEAAVGAHFAAVSTALDKVAAFGIPED
RVFGFWDWVGGRYSVWSAIGLPVMIAVGPDNFRKFLAGAHAMDVHFRDAPLEKNLPVMLGLIGY
WHRAICGYGSRAIIPYDQRLSRLPAYLQQLDMESNGKSVTLDGKPVSGPTGPVVWGEPGTNGQH
AFFQLLHQGTDTIPLEFIVAAKGHEPTLDHQHEMLMANCLAQSEALMKGRTLDEARAQLQAKNL
PASQVERIAPHRVFSGNRPSLTLIHDMLDPYALGRLIALYEHRVFVEAQIFGINAFDQWGVELG
KELATELLPVVSGKEGASGRDASTQGLVAHLHARRKA
```

Figure 8. The top hit and query sequence of the amino acid blast for BAA1334_I00104.



Conclusion

The computer database might not be entirely accurate in the annotation of these two genes investigated using the first 5 modules of GENI-ACT. Though the names and functions appear to be correct for all, further study of the amino acid blast sequences for both genes will need to be undertaken.

References

www.geni-act.org

http://geni-act.blogspot.com/p/welcome-to-geni-act-teaching-resources.html

Acknowledgments

Supported by an NSF Innovative Technology Experiences for Students and Teachers (ITEST) Award – 1311902
Thank you Mr. Winters!

www.buffalo.edu