Annotation of the *Borrelia burgdorferi B31* Genome From Locus Tags BB_0144 to BB_0146

Abby Kelly, Sage Empson, Ella Napolitano, Shannon Heinig, Andrea McMahon and Kate Carretto
Alfred Almond Central School, Almond, NY and The Western New York Genetics in Research and Health Care Partnership

SEPA SCIENCE EDUCATION PARTNERSHIP AWARD
Supported by the National Institutes of Health
ALFRED-ALMOND CENTRAL SCHOOL

Abstract

Students at Alfred-Almond Central School used the collaborative genome annotation website GENI-ACT to annotate three consecutive genes from the microorganism Borrelia burgdorferi. We used several online resources to determine aspects and functions of the genes such as general genomic information, amino acid sequence-based similarity data, structure-based evidence from the amino acid sequence, cellular localization data, potential alternative open reading frames, and enzymatic function. Our final determination was that the genes are most likely transporter permeases that reside within the cell membrane. Different information was gathered about the three different genes, so certain data is not available for some that is for others.

Introduction

Borrelia burgdorferi is the microorganism that causes Lyme Disease. It is a spirochete that is carried on ticks, and can be transmitted to both humans and animals. Lyme Disease causes flu like symptoms and a bullseye pattern rash. It causes exhaustion, and in extreme cases can be fatal. In most cases, however, it is treatable with a few weeks of antibiotics (Campbell Reece, 2008).

This bacterium is of interest because of its prevalence in rural areas, and its propensity for causing harm to humans and animals alike. Although a vaccine against this disease exists for dogs, one has not yet been developed for humans. Previous research has shown that the bacterium requires a host due to its limited metabolic capacity. (Fraser, 1997)

It was determined that the genes described in this study are located in the membrane of this organism. They aid it in transport, and have multiple transmembrane regions.

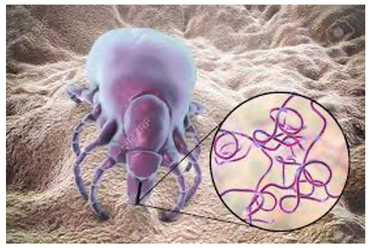

Figure 1: An image of a tick and the bacteria *Borrelia burgdorferi* B31 it carries which causes Lyme disease.

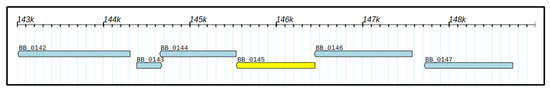

Figure 2: The locus tags and relative position of the genes under investigation in this research

Methods

Modules of the GENI-ACT (http://www.geni-act.org/) were used to complete *Borrelia burgdorferi* 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?
Alternative Open Reading Frame	IMG/M, Blast, AOR	What evidence is there that supports the proposed start codon?
Enzymatic Function	KEGG, MetaCyc, E.C. Number	In what process or structure is the protein under investigation involved?

Results

BB_0144:

The genome databases used for analysis confirmed the predicted product of a glycine betaine ABC transporter substrate-binding protein. This gene product proposal was supported by the top BLAST hits for the amino acid sequence, the presence of well-curated protein functional domains within the amino acid sequence, the transmembrane topography of the amino acid sequence, and the cellular location of the amino acid sequence.

Figure 3: Multiple Sequence Alignment WebLogo from http://weblogo.berkeley.edu/. Poor conservation is shown at positions 2-4, 5, 16 and 226 because of smaller and more numerous letters or no letters in the diagram. The prevalence of green, black and red letters supports that our protein is polar, hydrophobic and acidic.

BB_0145:

The genome databases used for analysis confirmed the predicted gene product. It is a transporter permease. This result was shown most clearly in the analysis using TMHMM, where it was shown that the protein has six transmembrane regions. The y-axis is the probability of a certain phenomenon. In this graph, there are six clear transmembrane regions. They are all at a probability of about 1, which is the same as a 100% certainty.
There is one possible location of a transmembrane region shown in the graph that does not have a high probability. This was disregarded, as the low probability made it unlikely to be a true transmembrane region. The regions of the protein on the interior and exterior were not found with the same high probabilities as the transmembrane regions, but it is likely that there is one region of the protein that is located outside the cell, in between the third and fourth transmembrane region.

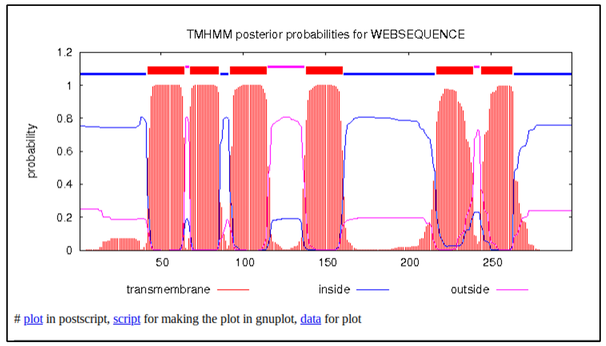

Figure 4: The graph shows the six transmembrane regions that exist in this protein.

Figure 5: Multiple Sequence Alignment WebLogo from http://weblogo.berkeley.edu/. Poor conservation is shown at positions 8, 35 and 55 because of smaller and more numerous letters in the diagram. The prevalence of green and black letters supports that our protein is polar and hydrophobic, both of which would be expected of a transmembrane protein.

BB_0146:

The databases used for analysis revealed the product was a choline ABC transporter, ATP-binding protein. This protein was supported by the top BLAST hits for the amino acid sequence, presence of functional domains within the amino acid sequence, and the location of the amino acid sequence.

Figure 6: Multiple Sequence Alignment WebLogo from http://weblogo.berkeley.edu/. Poor conservation is shown at positions 2,4 5 because of smaller and more numerous letters in the diagram. The prevalence of green and black letters supports that the protein is polar and hydrophobic.

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.

Locus Tag	Pipeline Annotation Product Name	Proposed Annotation	Changes Proposed?
BB_0144	Glycine/Betaine ABC Transporter Substrate Binding	Glycine/Betaine ABC Transporter Substrate Binding	No
BB_0145	Glycine/Betaine ABC Transporter Permease	Glycine/Betaine ABC Transporter Permease	No
BB_0146	Choline ABC Transporter / ATP Binding Protein	Glycine/Betaine ABC transporter ATP-binding protein	Yes

References

- Campbell, N. A., & Reece, J. B. (2008). Biology 8th Edition. Pearson/Cummings.
- Fraser, C. M., Casjens, S., Huang, W. M., Sutton, G. G., Clayton, R., Lathigra, R. Venter, J. C. (1997). Genomic sequence of a Lyme disease spirochaete, *Borrelia burgdorferi*.
  - Retrieved from https://www.nature.com/articles/37551

University at Buffalo The State University of New York
Jacobs School of Medicine and Biomedical Sciences
buffalo.edu

Acknowledgments

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