Annotation of the *Borrelia burgdorferi* Genome at Locus Tags BB_0141 and BB_0142

Caroline Ding, Ryann Donlon, Jessica McMahon, Catherine Elliott and Kate Carretto
Alfred-Almond Central School District, 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

Two consecutive genes from the microorganism *Borrelia burgdorferi* (BB_0141 – BB_0142) were annotated using the collaborative genome annotation website GENI-ACT. The first purpose of this research was to learn about the structure and functions of the coded proteins. The second purpose was to confirm if the gene is correctly annotated in the database. The GenBank proposed gene product name for BB_0141 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. BB_0142 was assessed in terms of the general genomic information, amino acid sequence-based similarity data, and cellular localization data. Programs used include protein BLAST, WebLogo, TMHMM, SignalP, etc. BB_0141 and BB_0142 appear to perform similar functions, as both were found to be involved in outward membrane transport. BB_0142 was found specifically to be involved in nodulation, acriflavin resistance, heavy metal efflux, and multidrug resistance proteins. The GenBank proposed gene product names 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 well-preserved and correctly annotated by in the r database.

Introduction

Borrelia Burgdorferi is a free-living pathogenic inhabitant of mud and water, typically thriving in anaerobic environments forming a gram negative spiral or wavelike body shape and flagella (Burgdorfer W, Barbour AG, Hayes SF, et al., 1982) It most commonly infects humans and other mammals. *Borrelia Burgdorferi* changes its gene expression to adapt to the different environments within the organisms (Yang X, Goldberg MS, Popova, et al., 2000).

Borrelia Burgdorferi was identified in 1981 by Willy Burgdorfer and Alan Barbour from studying a group of arthritis cases in children that appeared in the early 1970s. This genome was researched and found to be a bacterium transmitted by deer ticks and mice (Burgdorfer W, Barbour AG, Hayes SF, et al., 1982). This bacterium is associated with multiple illnesses including Lyme disease, and with the annotation of the genome many unanswered questions can be solved. For example, the function of the genome and why infection of different organisms presents varied symptoms might be understood.

The purpose of this research was to annotate and analyze two genes found in *Borrelia Burgdorferi*. The annotators were split into two different groups, and each group worked on a different section of the gene. Websites such as PSPORT-B T-Coffee, BLAST, and Pfam were used to help determine the function of our genes and to see if they are annotated correctly.

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?

Results

BB_0141:
The computer pipeline proposed product of this gene was a transporter periplasmic adaptor unit. This was supported by the top BLAST hits for the amino acid sequence, the presence of well-curated functional domains within the amino acid sequence, and the cellular location of the amino acid sequence. The amino acid sequence seems well-studied and correctly conserved according to multiple BLAST results had E-values of 0.0 and WebLogo. Structure-based similarity data revealed the protein's function. Results from TIGRFAM confirmed that the protein is a membrane fusion protein involved in outward transport. In addition, PDB results and top hit of PFAM confirmed that the protein is involved in heavy metal efflux. The second hit of PFAM revealed that biotin and lipoamide acyltransferases play an important role in the protein's function. The structure of the protein was understood through cellular localization. The protein is most likely in the cytoplasm according to PSORTB predictions. Results from SignalP and LipoP indicated that the protein has a lipoprotein signal peptide cleaved by SPII. LipoP predicted that there are less than 4 cleavage sites, and the most likely position for a lipoprotein signal peptide (according to Signal IP) is between position 24 and 25 LVA-CV. There are no transmembrane helices in the protein according to TMHMM and Phobius, which shows that the protein is not an integral membrane protein.

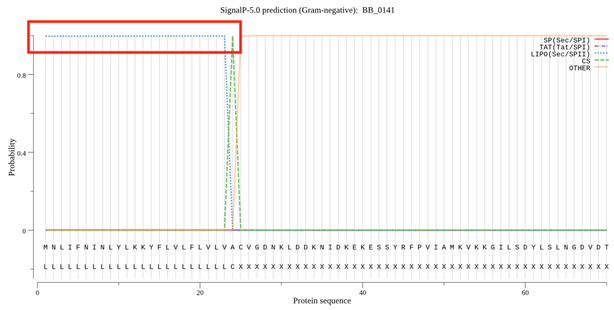

Figure 1 – BB_0141 SignalP results predict that the protein is a lipoprotein signal peptide with likelihood of 0.9976.

BB_0142:
The computer pipeline proposed product of this gene was an outer membrane efflux 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, and the cellular location of the amino acid sequence. The function was also supported by several other BLAST hits with 0.0 E values. The data found supports the proposed annotation of this gene as an outer membrane efflux protein and finds it to be accurate.
The initial proposed product of this gene was an outer membrane efflux protein. This was supported by the top BLAST hits for the amino acid sequence, the presence of well-curated functional domains within the amino acid sequence, and the cellular location of the amino acid sequence, This gene product proposal is also supported by the identification of BB_0141 and as a formate as a periplasmic adaptor unit in the cytoplasm. As such, the proposed annotation is a outer membrane efflux protein.

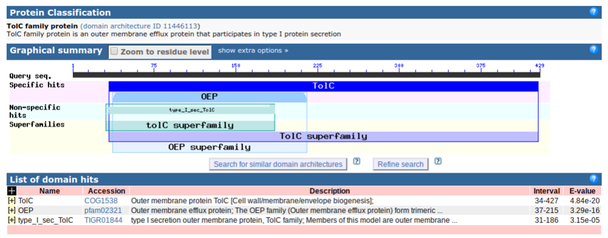

Figure 2 – BB_0142: Outer membrane efflux proteins that participate in type I protein secretion that are similar or identical to BB_0142

```
# BB_0142 Length: 428
# BB_0142 Number of predicted TMHs:  1
# BB_0142 Exp number of AAs in TMHs: 17.60326
# BB_0142 Exp number, first 60 AAs:  17.00696
# BB_0142 Total prob of N-in:        0.91838
# BB_0142 POSSIBLE N-term signal sequence
BB_0142 TMHMM2.0        inside     1    12
BB_0142 TMHMM2.0        TMhelix    13   30
BB_0142 TMHMM2.0        outside    31   428
```

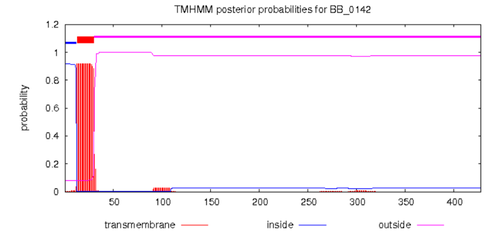

Figure 3 – BB_0142 The transmembrane topology graph indicates that gene BB_0142 contains 1 transmembrane helix and can be found on the outer membrane of the cell.

Conclusion

BB_0141 appears to be a signal peptide not residing within the membrane but possibly within the cytoplasm. 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_0141	Transporter Periplasmic Adaptor Unit in The Cytoplasm	Transporter Periplasmic Adaptor Unit in The Cytoplasm	No
BB_0142	Outer Membrane Efflux Protein	Outer Membrane Efflux Protein	No

References

Burgdorfer W, Barbour AG, Hayes SF, et al. Lyme disease - a tick-borne spirochetosis? Science. 1982.216:1317-1319.

Yang X, Goldberg MS, Popova, et al. 2000 Interdependence of environment factors influencing reciprocal patterns of gene expression in virulent Borrelia burgdorferi, Molecular Microbiology, 37(6), 1470-1479.

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

Acknowledgments
Dr. Rama Dey-Rao and Dr. Stephen Koury
Supported by an NIH Science Education Partnership (SEPA) Award - R25GM129209