Annotation of the *Photorhabdus luminescens* Genome at Locus Tag PLU_RS1965

Leah Stucke and Lawrence Grisanti
East Aurora High School and the Western New York Genetics in Research and Health Care Partnership

SEPA SCIENCE EDUCATION PARTNERSHIP AWARD
Supported by the National Institutes of Health

Abstract

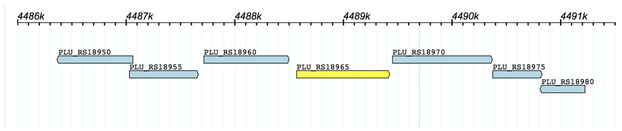
A gene from the microorganism *Photorhabdus luminescens* (PLU_RS18965) was annotated using the collaborative genome annotation website GENI-ACT. The GenBank proposed gene product name for the gene was assessed in terms of the general genomic information, amino acid sequence-based similarity data, structure-based evidence from the amino acid sequence, cellular localization data, and potential alternative open reading frames. Through the use of various programs, including BLAST, CDD, T-Coffee, WebLogo, TMHMM, and other programs, the gene was annotated and found to match the proposed product names. The GenBank proposed a gene product name for the gene and the results did not differ significantly from the proposed gene annotation for the genes in the group and as such, the gene appears to have been correctly annotated by the r database. According to gene annotation websites and software gene PLU_RS18965 was correctly named as a chaperone protein.

Introduction

The genes that was studied was from the genome of *Photorhabdus luminescens*. In order to properly investigate the properties of this gene, a variety of software tools were employed. These included BLAST, CDD, T-Coffee, WebLogo, TMHMM, among others. By using these platforms, a higher educational purpose has been served. Knowledge regarding the similarity of genomes as well as a general understanding of bioinformatics has resulted in response to this educational stimulus.

Gene PLU_R18965 (location shown in Figure 1) was computer pipeline annotated previously using some of the software listed above and those services proposed gene product names for the protein. However, to investigate the relative characteristics of these genes, research on organisms that showed similarity in sequences was conducted to fully understand the traits of the genes in *Photorhabdus luminescens*. Research involving the enzymatic function of the gene, the duplication and degradation of the gene, the horizontal gene transfer and the RNA family of the gene has not been conducted.

To further understand the genes of *Photorhabdus luminescens*, certain aspects were investigated. These include sequence based and structure based similarity, as well as alternative reading and cellular localization. The results and findings from the research done will be expounded upon in the following content.



Methods

Modules of the GENI-ACT (http://www.geni-act.org) were used to complete *Photorhabdus luminescens* 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, T-Coffee	In what process or structure is the protein under investigation involved?

Results

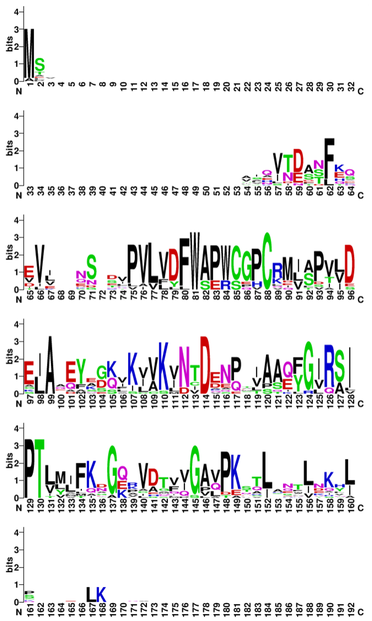
PLU_RS18965 Annotation Findings

Through the annotation of gene PLU_RS18965 and the extensive studies comparing its proposed gene product name and the results from various software it was found that the gene functions as a chaperone protein. The initial proposed product name for this gene was a chaperone protein which assists in the covalent folding or unfolding of macromolecule structures. The sequences of the protein were analyzed in comparison to sequences of similar molecules through BLAST. The top BLAST hit was a chaperedoxin located in Escherichia coli K-12. A chaperedoxin has a combination of a chaperone protein and redox-protective functions. A sequence alignment was conducted through the T-Coffee software and was used to create a WebLogo (Figure 2).

The areas of protein PLU_RS18965 alternate between a high number of matching amino acids or having no amino acids match to areas with little conservation of amino acid, examples of this can be seen in the T-Coffee multiple sequence alignment and the WebLogo and the concentration of amino acids in certain regions exemplified in the analyses of these software calculations.

Figure 1. (Left) The image to the left shows the location of gene PLU_RS18965 (highlighted in yellow) in its gene neighborhood in the organism. 1

Figure 2. The image on the right is a portion of the WebLogo made from the analysis of the genes sequences and their similarity to other sequences. WebLogo The WebLogo has sections that have very few letters and sections that have many letters, the sections that contain many letters have a good balance of types of amino acids. They often have two amino acids per row, one in a small quantity and the other with a larger quantity, but there are also often many smaller quantities of proteins in a row, showing negligible conservation in the regions of protein PLU_RS18965



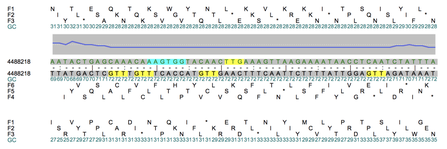
After determining how the PLU_RS1965 sequence aligned to similar genes, it was analyzed for the structural similarity to other genes. Gene PLU_RS18965 was entered into CDD, TIGRFAM, Pfam, and PDB software for further analysis. The data analysis from these programs showed that the gene being studied was a chaperone protein that also has redox functions.

Figure 3. Gene PLU_RS18965 results from CDD analysis that show the gene is a chaperone protein with thioredoxin-like domains.

The protein PLU_RS18965, which BLAST results identify as a chaperone protein, is most likely found in the cytoplasm since TMHMM and Phobius did not predict it to have transmembrane helixes and PSORTb determined it had the highest probability of being a cytoplasmic protein.

For the protein PLU_RS18965 the collaborative GenBank website GENI-ACT had the start codon located correctly, an open reading frame is found in the beginning of the protein sequence in the results from IGM/M software. The diagramed results show the start codon regions in yellow and the Shine-Dalgarno regions in teal, showing the areas of the open reading frames that correctly matched the expected regions from the previous gene annotation (Figure 4).

Figure 4. The first portion of the IGM/M results of alternate open reading frames.



The gene from the microorganism *Photorhabdus luminescens* identified by locus tag PLU_RS18965 has been analyzed through various software including BLAST, WebLogo, and CDD to examine the sequences and structure of the protein and to confirm the gene's proposed product name. We hypothesize this gene was annotated correctly as a chaperone protein. However the gene also contains thioredoxin and redox like capabilities. YbbN has been proposed to act as a chaperone or co-chaperone that aids in heat stress response and DNA synthesis in *E. coli* (Lin and Wilson, 2011).

Conclusion

The GENI-ACT proposed gene products did not differ significantly from the proposed gene annotations for each of the genes and as such, the genes appear to be correctly annotated by the computer database.

PLU_RS18965
Proposed Gene Product Name: Co-Chaperone YbbN
Pipeline Annotated Gene Product Name: Co-Chaperone YbbN
Changes made: None

Notes: PLU_RS18965 is a chaperone protein that has some redox qualities

References

Lin and Wilson. 2011. J. Biol. Chem. jbc.M111.238741. doi:10.1074/jbc.M111.238741

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