

Annotation of the *Photorhabdus luminescens* subsp. *laumondii* TTO1 Genome

At Locus Tags PLU_RS18945 and PLU_RS18970

Taylor McClure, Kaela Guest and Mr. Lawrence Grisanti
East Aurora High School | East Aurora, NY and The Western New York Genetics in Research and Health Care Partnership

Abstract

A group of 2 genes from the bacterium *Photorhabdus luminescens* subsp. *laumondii* TTO1 (PLU_RS18945 and PLU_RS18970) 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. Also of interest were enzymatic function, presence or absence of gene duplication and degradation, the possibility of horizontal gene transfer, and the production of an RNA product. 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 database.

Introduction

Photorhabdus luminescens is a gram negative bacterial insect pathogen from the *Enterobacteriaceae* family. It symbiotically resides in the digestive tract of soil-based nematodes. When the nematode is ingested by insect larvae, *P. luminescens* is released by regurgitation, thereby releasing toxins and killing the insect. Additionally, *P. luminescens* is a symbiote to soil nematodes belonging to the species *Heterorhabditis bacteriophora*. Through its complex life cycle of being a symbiote and a pathogen by adapting host environmental conditions, different processes are analyzed at different stages within the pathogen’s life (Munch et al.). By sequencing different genes within *P. luminescens*, different relationships with proteins such as MACPF proteins and gcvB RNA genes can be made (Munch et al.). The complete genome sequence is 5,688,987 base pairs long and contains 4,839 predicted protein-coding genes (Dunchaud et al.). This encodes a large number of adhesins, toxins, hemolysins, proteases and lipases, and contains a wide array of antibiotic synthesizing genes. The phylogeny of this microorganism is also a source of interest, as it is a member of the family *Enterobacteriaceae*, which has yet to be thoroughly studied utilizing bioinformatics (Dunchaud et al.)

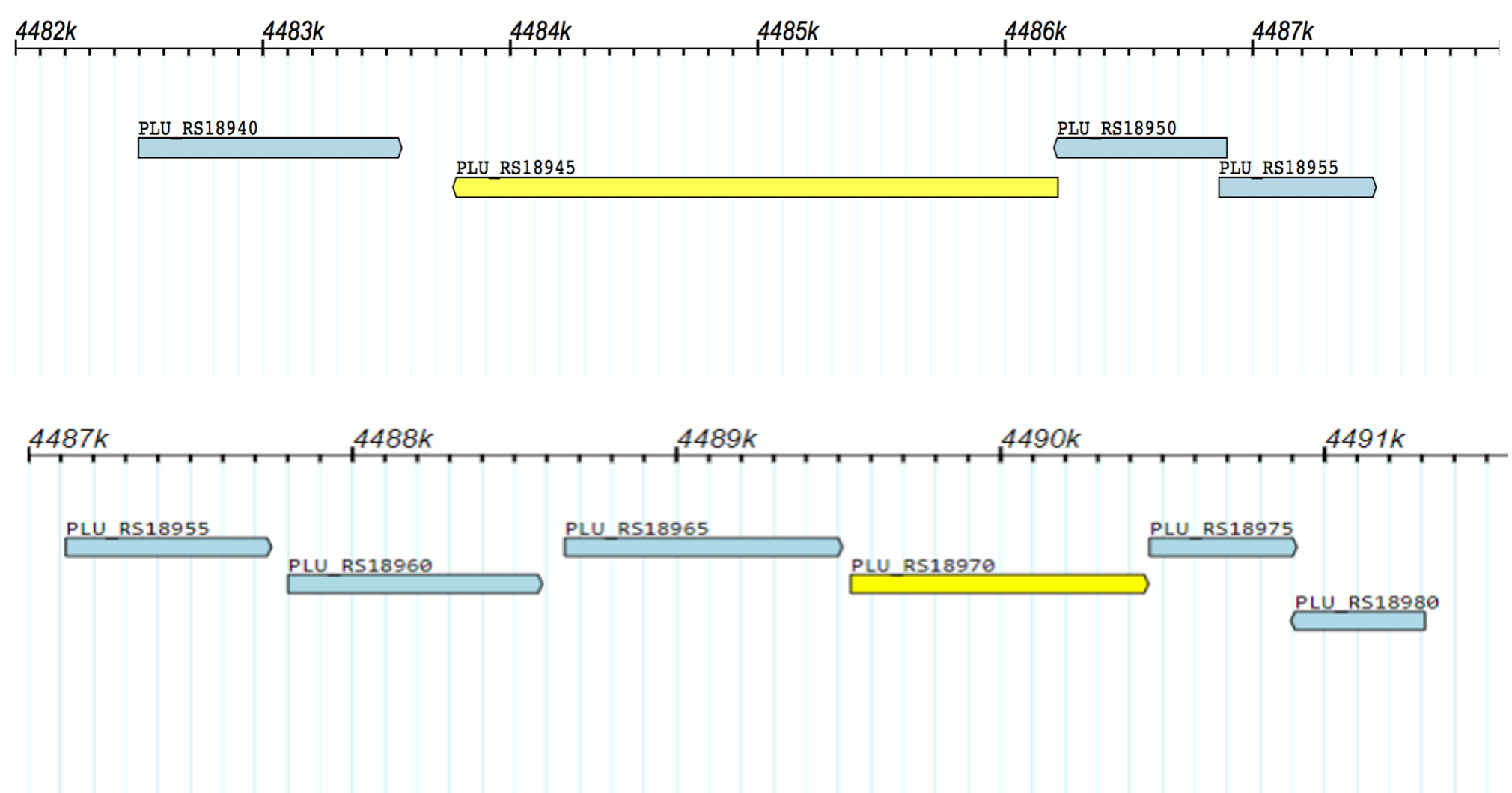


Figure 1. The locus tags and relative position of the genes under investigation in this research. (top) PLU_RS18945 (bottom) PLU_RS18970

Methods

Modules of the GENI-ACT (<http://www.geni-act.org/>) were used to complete *Photorhabdus luminescens* subsp. *laumondii* TTO1 genome annotation. The modules are described below; due to time constraints, the annotation was only completed through module 5.

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?
Enzymatic Function	KEGG, MetaCyc, E.C. Number	In what process or structure is the protein under investigation involved?
Duplication and Degradation	Paralog, Pseudogene	Are there other forms of the protein under investigation in the same genome? Is it functional?
Horizontal Gene Transfer	Phylogenetic Tree, Gene Neighborhood, GC Content	Has the protein under investigation co-evolved with the rest of the genome or has it been obtained in a different way?
RNA family	Rfam	Does the gene under investigation encode a functional RNA?
Final Annotation	Evaluate data from all modules	Has the gene been correctly called by the pipeline annotation?

Results

PLU_RS18945:
The computer pipeline proposed product of this gene was a ABC transporter permease. This gene product proposal was supported by the top BLAST hits for the amino acid sequence and the top CD hits for the amino acid sequence. Due to unattributed errors in modules, the prediction is supported with the most conclusive data attainable.



Figure 2. CDD search results for PLU_RS18945, showing inconclusive results from the NCBI database.

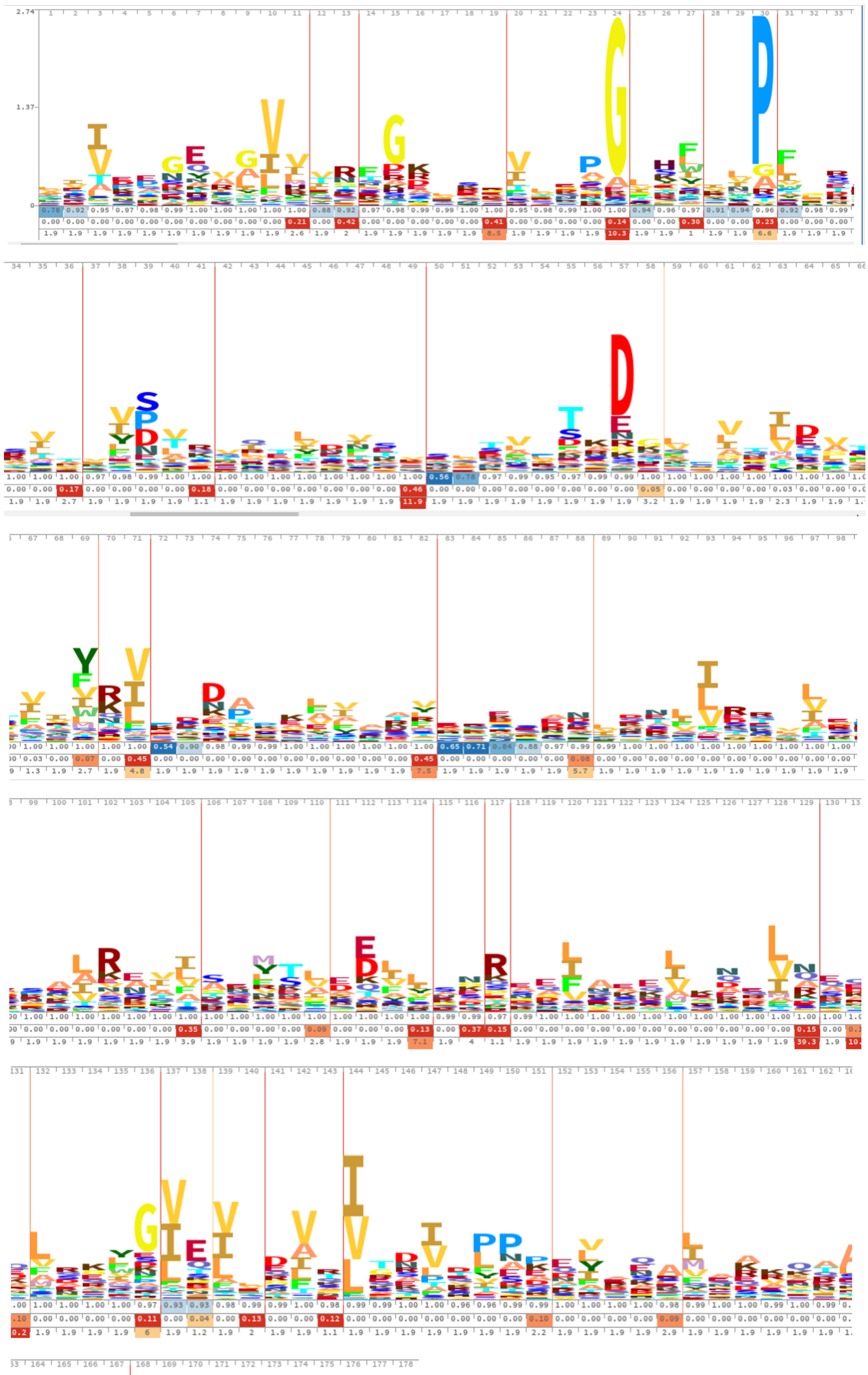


Figure 3. HMM logo of top Pfam hit for PLU_RS18970 (Band 7).

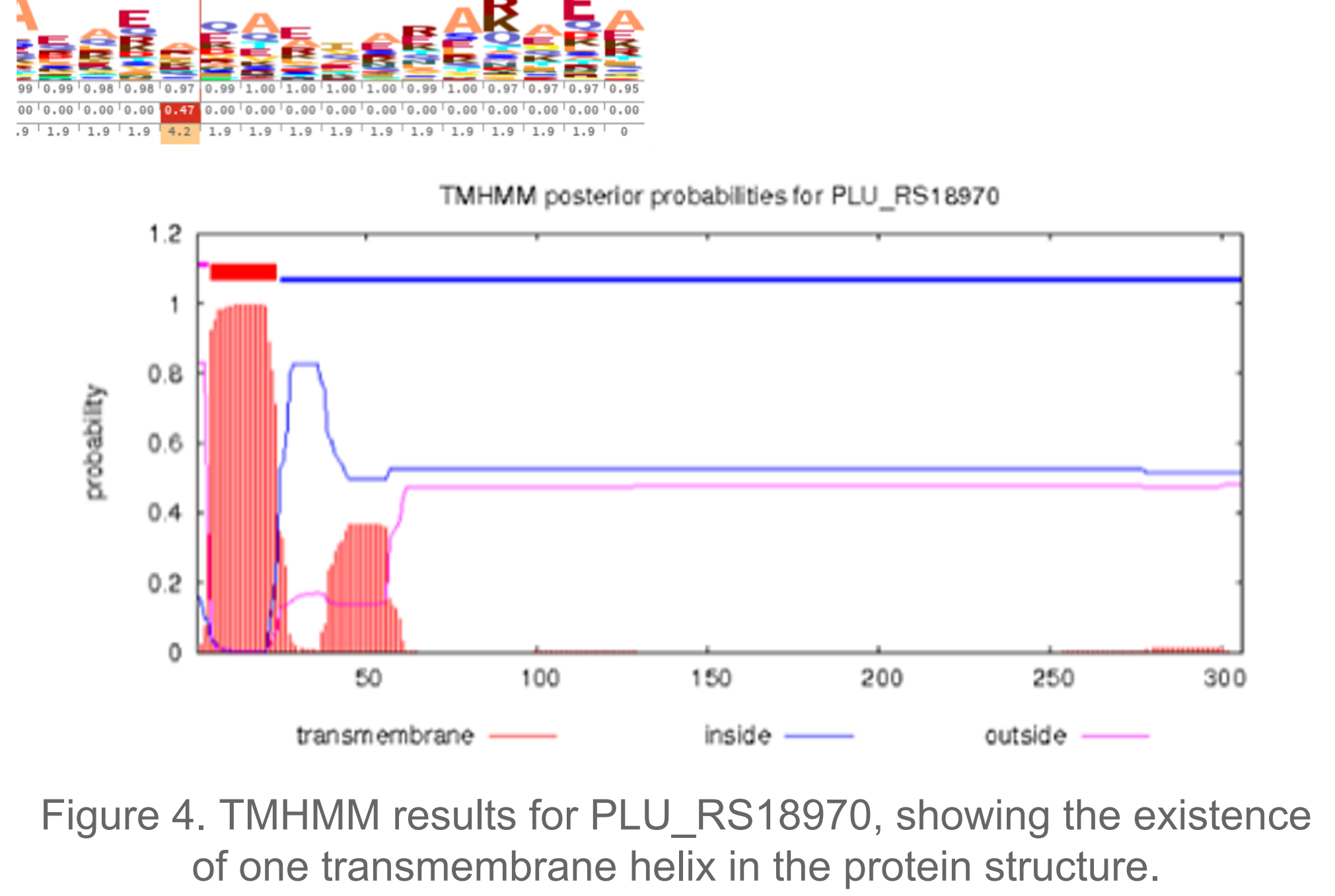


Figure 4. TMHMM results for PLU_RS18970, showing the existence of one transmembrane helix in the protein structure.

PLU_RS18970:
The computer pipeline proposed product of this gene was a SPFH/Band 7/PHB domain protein. This was supported by the top CDD search results 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 protein domain is SPFH Band 7 or PHB, a stomatin-like integral membrane domain conserved from protozoa to mammals. The exact function of the band-7 domain is not known, but examples from bacterial stomatin-type proteins demonstrate binding to lipids and the ability to assemble into membrane-bound oligomers that form putative scaffolds (Gehl et al.). Bacterial HflC protein also belongs to this family; this is significant, as the top COG and TIGRFAM hits indicated a close match to a HflC protein. This protein likely lies mostly within the cytoplasm. TMHMM, PSORT-B and Phobius database searches each predicted the presence of one transmembrane helix within the protein. PSORT-B predicts with a high certainty that the protein is located in the cytoplasm. The protein is partially within the cytoplasmic membrane due to the presence of a TMH, but lies mostly within the cytoplasm. This correlates with the prediction of the protein as an integral membrane protein, with a section of the protein anchored in the phospholipid bilayer.

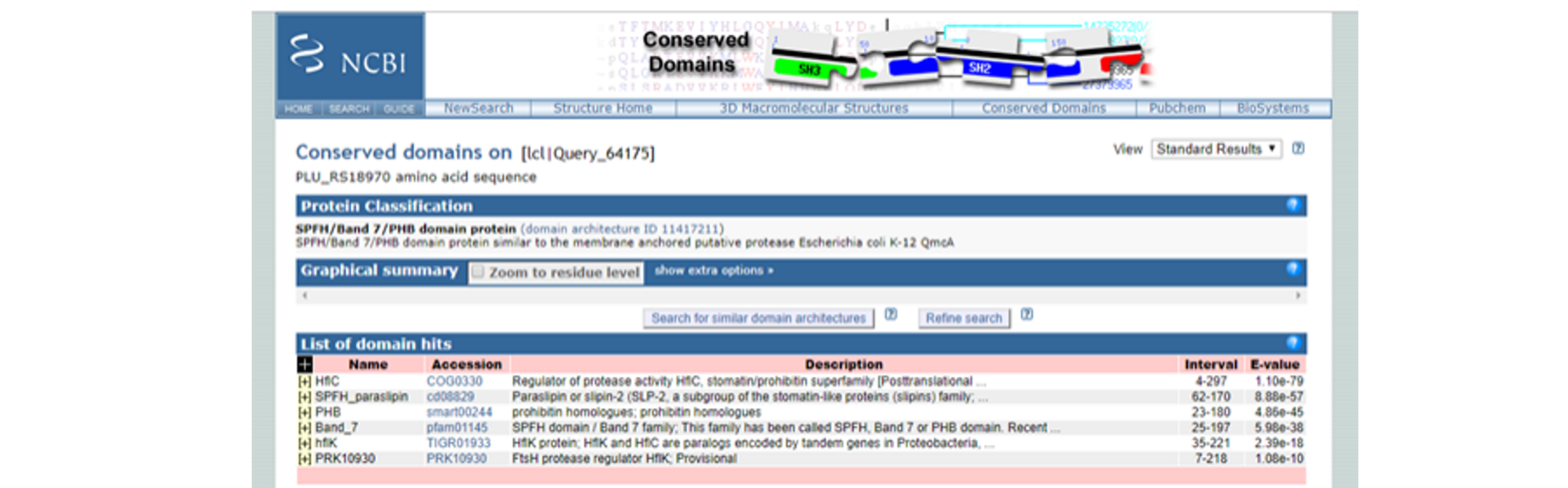


Figure 5. CDD search results for PLU_RS18970, supporting the pipeline annotation.

Conclusion

The GENI-ACT proposed gene product did not differ significantly from the proposed gene annotation for each of the genes in the group; as such, the genes appear to be correctly annotated by the computer database. Not all of the modules were completed; however, those which could be utilized to assess the accuracy of the proposed annotation supported the pipeline result.

Locus Tag	Pipeline Annotation Product Name	Proposed Annotation	Changes Proposed?
PLU_RS18945	ABC Transporter Permease	ABC Transporter Permease	No
PLU_RS18970	SPFH/Band 7/PHB domain protein	SPFH/Band 7/PHB domain protein	No

References

Dunchaud et al, 2013. “The genome sequence of the entomopathogenic bacterium *Photorhabdus luminescens*.” Nature Biotechnology.

Munch et al, 2008. “*Photorhabdus luminescens* genes induced upon insect infection.” BMC Genomics.

Gehl, Bernadette, and Lee J. Sweetlove, 2014. "Mitochondrial Band-7 Family Proteins: Scaffolds for Respiratory Chain Assembly?" *Frontiers in Plant Science*. NCBI PMC.

Acknowledgments

Dr. Stephen Koury
Supported by an NIH Science Education Partnership (SEPA) Award - R25GM129209