Annotation of the *Photorhabdus luminescens* Genome at Locus Tag PLU_RS18960 and PLU_RS18955)

Matthew Izzo, Colden Ball, and Lawrence Grisanti
East Aurora High School, East Aurora, NY and The Western New York in Genetics in Research and Health Care Partnership

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

Abstract

A gene for the microorganism *Photorhabdus luminescens* using the collaborative genome annotation website GENI-ACT. The gene under investigation, PLU_RS18960 and PLU_RS18955, 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. The investigated function of PLU_RS18960 appears to be that of oxidoreductase and PLU_RS18955 appears to be a arylesterase, a member of the enzyme family of hydrolases. The gene does appear to be correctly annotated by the GENI-ACT database.

Introduction

Genome annotation is a scientific look at a specific organisms gene. More specifically, the genes location and function. Using modern technology, machines can read and annotate genes automatically however they are not always one hundred percent correct. The purpose of this research is to determine the correctness of the computer annotation of PLU_RS18960 and PLU_RS18955 by investigating even further.

Based on an article from Athina Rodou, Denis O. Ankrah, and Christos Stathopoulos on the NCBI database. *Photorhabdus luminescens* is a bioluminescent bacterium that belongs to the family *Enterobacteriaceae*. It is also nematode-symbiotic and gram negative. This bacterium has been found to be an alternative source of insecticides, and as an emerging human pathogen (Rodou et al., 2010)

The annotation of the gene PLU_RS18960 and PLU_RS18955 is an investigation to find the location of the protein generated from this gene within the cell, the function of the protein, and whether or not the computer annotation is correct. Using various methods involving online databases and programs, these variables can be determined.

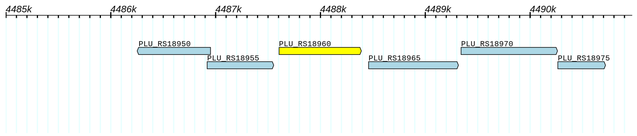


Figure 1. A small map of the surrounding genes. The highlighted gene (PLU_RS18960) is the gene under investigation.

Methods

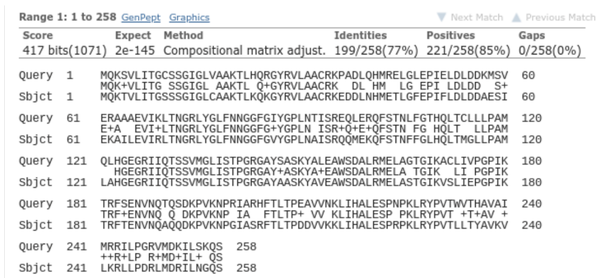
Modules of the GENI-ACT (http://www.geni-act.org/) were used to complete *Photorhabdus luminescens* genome annotation. The modules used are described below:

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

PLU_RS18960

Using the various methods shown above, the results can be generated. In the Sequence-Based Similarity module proteins are compared to others based upon their amino acid sequence. BLAST is a program that takes the desired sequence and compares it to others on a desired database.



Range 1: 1 to 258 GenPept Graphics ▼ Next Match ▲ Previous Match

Score	Expect	Method	Identities	Positives	Gaps
417 bits(1071)	2e-145	Compositional matrix adjust.	199/258(77%)	221/258(85%)	0/258(0%)

```
Query  1    MQKSVLITGCSSGIGLVAAKTLHQRGYRVLAACRKPADLQHMRELGLEPIELDLDDDKMSV  60
            MQK+VLITG SSGIGL AAKTL Q+GYRVLAACRK  DL HM  LG EPI LDLDD  S+
Sbjct  1    MQKTVLITGSSSGIGLCAAKTLKQKGYRVLAACRKEDDLNHMETLGFEPIFLDLDDAESI  60

Query  61   ERAAAEVIKLTNGRLYGLFNNGGFGIYGPLNTISREQLERQFSTNLFGTHQLTCLLLPAM  120
            E+A  EVI+LTNGRLYGLFNNGGFG+YGPLN ISR+Q+E+QFSTN FG HQLT  LLPAM
Sbjct  61   EKAILEVIRLTNGRLYGLFNNGGFGVYGPLNAISRQQMEKQFSTNFFGLHQLTMGLLPAM  120

Query  121  QLHGEGRIIQTSSVMGLISTPGRGAYSASKYALEAWSDALRMELAGTGIKACLIVPGPIK  180
            HGEGRIIQTSSVMGLISTPGRGAY+ASKYA+EAWSDALRMELA TGIK  LI PGPIK
Sbjct  121  LAHGEGRIIQTSSVMGLISTPGRGAYASKYAVEAWSDALRMELASTGIKVSLIEPGPIK  180

Query  181  TRFSENVNQTQSDKPVKNPRIARHFTLTPEAVVNKLIHALESPNPKLRYPVTWVTHAVAI  240
            TRF+ENVNQ Q DKPVKNP IA  FTLTP+ VV KLIHALESP PKLRYPVT +T+AV +
Sbjct  181  TRFTENVNQAQQDKPVKNPGIASRFTLTPDDVVKKLIHALESPRPKLRYPVTLLTYAVKV  240

Query  241  MRRILPGRVMDKILSKQS  258
            ++R+LP R+MD+IL+ QS
Sbjct  241  LKRLLPDRLMDRILNGQS  258
```

Figure 2. One of the top results from a BLAST search on the nr database. An oxidoreductase from the organism *Proteus Mirabilis*.

The genes found on BLAST proved to be very similar to PLU_RS18960 indicated by good e-values. This gives evidence leading to the legitimacy of the proposed gene. Results of blast also give evidence towards the function of the gene by analyzing the function of genes from the BLAST results, showing multiple oxidoreductase.

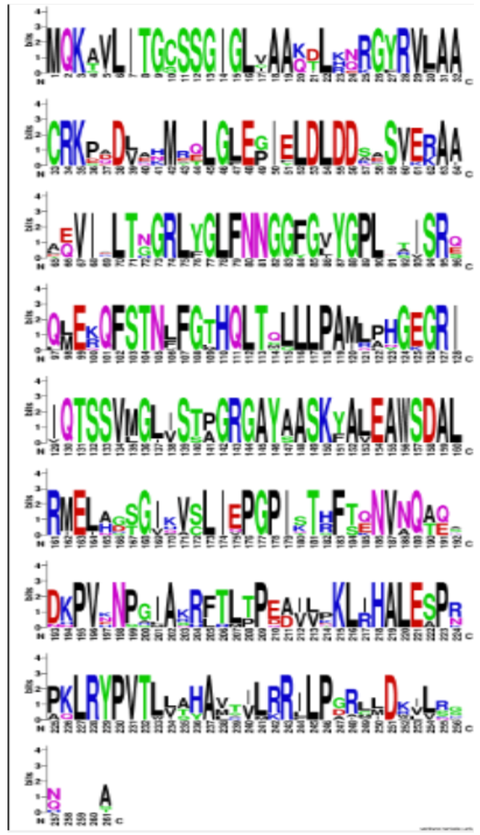


Figure 3. WebLogo made with the assistance of T-Coffee alignment. Made using 10 sequences including PLU_RS18960.

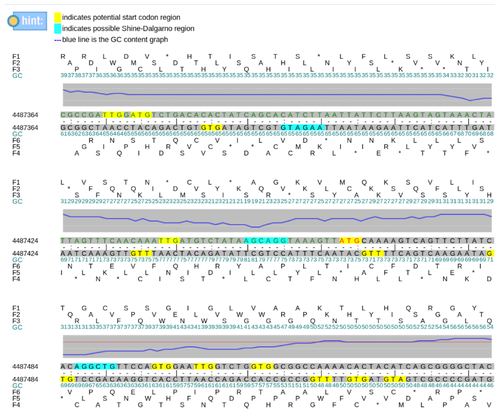
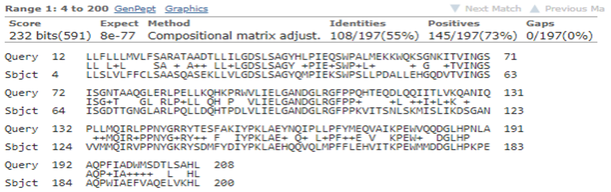


Figure 4. Open reading frame graphical output. Called start codon is highlighted in red.

The called start codon on section 2 of the graphical output can be shown highlighted in red with a corresponding Shine-Dalgarno sequence 7 bases upstream. The called coordinates appear to be correct as they hold the best e-values and scores when put into BLAST. A shorter alternate ORF was identified at coordinates 4487935…..4488248 with a corresponding Shine-Dalgarno sequence however the E values increased and the scores decreased.

PLU_RS18955

Using the various methods shown above, the results can be generated. In the Sequence-Based Similarity module proteins are compared to others based upon their amino acid sequence. BLAST is a program that takes the desired sequence and compares it to others on a desired database. With minimal pre-analyzed data available, BLAST provided the most significant function information of PLU_RS18955.



Range 1: 4 to 200 GenPept Graphics ▼ Next Match ▲ Previous Ma

Score	Expect	Method	Identities	Positives	Gaps
232 bits(591)	8e-77	Compositional matrix adjust.	108/197(55%)	145/197(73%)	0/197(0%)

```
Query  12   LLFLLLMVLFSARATAADTLLILGDSLSAGYHLPIEQSWPALMEKKWQKSGNKITVINGS  71
            LL L+L    SA + A++ LL+LGDSLSAGY +PIE+SWP+L+     + G  +TVINGS
Sbjct  4    LLSLVLFFCLSAASQASEKLLVLGDSLSAGYQMPIEKSWPSLLPDALLEHGQDVTVINGS  63

Query  72   ISGNTAAQGLERLPELLKQHKPRWVLIELGANDGLRGFPPQHTEQDLQQIITLVKQANIQ  131
            ISG+T   GL RLP+LL QH P  VLIELGANDGLRGFPP+    +L ++I+L+K +
Sbjct  64   ISGDTTGNGLARLPQLLDQHTPDLVLIELGANDGLRGFPPKVITSNLSKMISLIKDSGAN  123

Query  132  PLLMQIRLPPNYGRRYTESFAKIYPKLAEYNQIPLLPFYMEQVAIKPEWVQQDGLHPNLA  191
             ++MQIR+PPNYG+RY+++ F  IYPKLAE+ Q+ L+PF++E V  KPEW+  DGLHP
Sbjct  124  VVMMQIRVPPNYGKRYSDMFYDIYPKLAEHQQVQLMPFFLEHVITKPEWMMDDGLHPKPE  183

Query  192  AQPFIADWMSDTLSAHL  208
            AQP+IA++++  L  HL
Sbjct  184  AQPWIAEFVAQELVKHL  200
```

Figure 5. One of the top results from a BLAST search on the Swiss-Prot database. An arylesterase from the organism *Vibrio mimicus*

The genes found on BLAST of the organism proved to be very similar to PLU_RS18955 indicated by significant e-values and high query percentages. This gives evidence leading to the legitimacy of the proposed gene. Results of blast also give evidence towards the function of the gene by analyzing the function of genes from the BLAST results, showing multiple oxidoreductase. After analyzing the sequencing and function of the gene in the organism *Vibrio mimicus*, it was found that this similar sequence was a arylesterase. Arylesterase is an enzyme that catalyzes the chemical reaction, phenyl acetate + H_2O → phenol + acetate. This enzyme belongs to the family hydrolase, a class of enzyme that performs as a biochemical catalysts that use water to break chemical bonds.

Conclusion

The GENI-ACT proposed annotation is accurate after further investigation from BLAST and the open reading frame. PLU_RS18960 is a member of the oxidoreductase family and is located within the cytoplasm. PLU_RS18960 is an arylesterase, a member of the hydrolase family and is located in the cell wall of the bacteria.

References

Rodou, A.; Ankrah, D.O.; Stathopoulos, C. Toxins and Secretion Systems of *Photorhabdus luminescens*. Toxins 2010, 2, 1250-1264.

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
Dr. Stephen Koury