# Annotation of the Pseudomonas aeruginosa Genome from Locus Tags T223_00130 to T223_00145

Andrew Herberger, Ethan Rio, Samantha Phillips, Cole Werner, Michael Oldziej and Mary Keymel
Avon Central 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

University at Buffalo

## Abstract

A group of consecutive 3 genes from the microorganism *Pseudomonas aeruginosa* (T223_00130, T223_135, and T223_00145) 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, potential alternative open reading frames, 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 in the database.

## Introduction

*Pseudomonas aeruginosa* is a rod shaped,mono flagellated, and asporogenous forming Gram negative opportunistic human pathogen, found in environments such as soil, water, humans, animals, plants, sewage, and hospitals. Organic growth factors are not required, and it can use more than seventy-five organic compounds for growth. *Pseudomonas Aeruginosa* can use both aerobic and anaerobic respiration. *Pseudomonas aeruginosa* is an opportunistic human pathogen. It plays a large role in cystic fibrosis infections and other hospital infections.

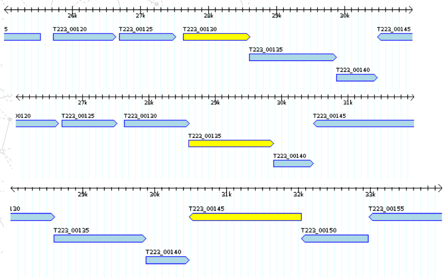

Figure 1: The locus tags under investigation in this research.

Figure 2. A picture of the bacteria Pseudomonas Aeruginosa.

## Methods

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

| Modules | Activities | Questions Investigated |
|---|---|---|
| Module 1- Basic Information Module | DNA Coordinates and Sequence, Protein Sequence | What is the sequence of my gene and protein? Where is it located in the genome? |
| Module 2- Sequence-Based Similarity Data | Blast, CDD, T-Coffee, WebLogo | Is my sequence similar to other sequences in Genbank? |
| Module 3- Cellular Localization Data | Gram Stain, TMHMM, SignalP, PSORT, Phobius | Is my protein in the cytoplasm, secreted or embedded in the membrane? |
| Module 4- Alternative Open Reading Frame | IMG Sequence Viewer For Alternate ORF Search | Has the amino acid sequence of my protein been called correctly by the computer? |
| Module 5- Structure-Based Evidence | TIGRfam, Pfam, PDB | Are there functional domains in my protein? |

## Results

**T223_00130**

The initial proposed product of this gene by GENI-Acts was a Zinc Dependent phospholipase C. The top hit from blast was the organism Enterobacter cloacae. It's gram negative bacteria. The TMHMM results showed that the proposed product has no transmembrane helices, but does have signal peptide according to signal IP.

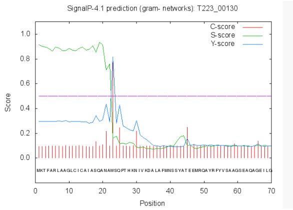

Figure 3. Signal Peptide Graph. The most likely cleavage site is between amino acid positions 22 and 23.

**T223_00135**

The initial proposed product of this gene by GENI-Acts as a foldase protein prsA 1 precusor. This product was found in the organism Enterobacter cloacae. This was supported by my blast that showed this organism as the top hit. This organism was also found to be gram negative. TMHMM predicted 0 helices for the gene product and Signal IP predicted no signal peptide. The PSORTB predicted that this protein may have multiple locations within the cell.

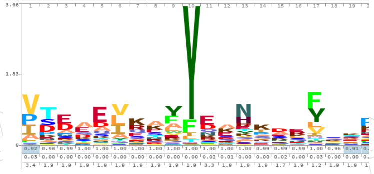

Figure 4. This is a segment of the HMM logo from organism T223_00135

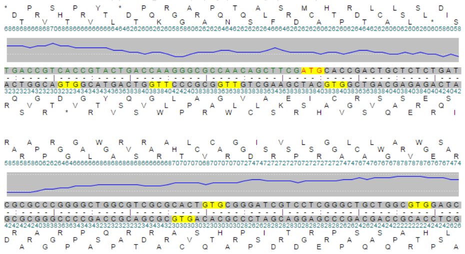

Figure 5. Alternate Open Reading Frame module results. The coordinates appeared to be correctly called. Another alternative open reading frame with a Shine-Dalgarno sequence was investigated but the E value was increased and the score was decreased

**T223_00145**

The gene product I received was a sulfate transporter. This was supported by being the top result when I did a BLAST search. The organism was a *Enterobacter cloacae*. Also in the Cellular Localization Data module I learned that the *Pseudomonas Aeruginosa* bacteria was Gram-Negative. The sulfate transporter is a transmembrane protein and TMHMM showed that it has11 helices.

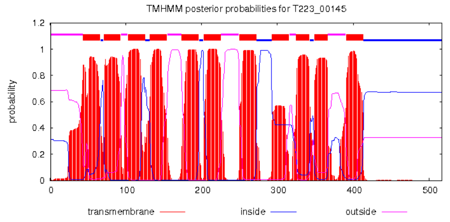

Figure 6. The TMHMM results which shows evidence of 11 helices in the gene product.

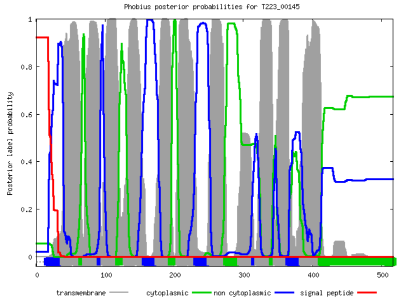

Figure 7. Phobius results for T223_00135 confirming the presences of 11 helices

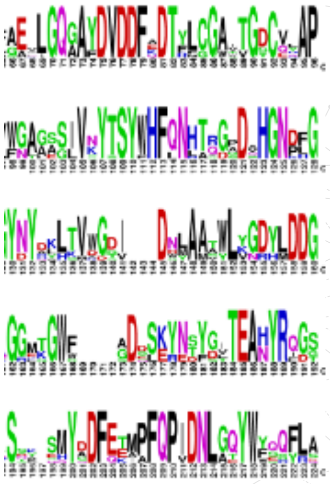

Figure 8. This is a Sequence from WebLogo with minimal gaps and some variation.

## 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.

| Gene Locus | Geni-Act Gene Products | Proposed Animation |
|---|---|---|
| T223_00130 | Zinc dependent phospholipase C | Zinc dependent phospholipase C |
| T223_00135 | Foldase protein prsA 1 precusor (Entrobacter cloacae) | Foldase protein prsA 1 precusor (Entrobacter cloacae) |
| T223_00145 | Sulfate Transporter | Sulfate Transporter |

## References

blast.ncbi, weblogo.edu, ncbi.gov/pubmed, cbs.dtu.dk/services/TMHMM, www.Healthline.com

## Acknowledgments

Dr. Stephen Koury and Dr. Rama Dey-Rao .
Supported by an NIH Science Education Partnership (SEPA) Award - R25OD10536-1 A1
**www.buffalo.edu**