Annotation of the *Pseudomonas aeruginosa* LES431 Genome from Locus Tags PA0029 to PA0034

Megan Clemens, Jonathan Riedel, Daniel Schiffhauer, Natalie Standish, Isabella Thomas, Samuel Weber and Samantha Evans
Iroquois 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

The students of Iroquois High School proofed 6 consecutive genes from the microorganism *Pseudomonas aeruginosa* (PA0029 – PA0034). These genes were annotated using the collaborative genome annotation website GENI-ACT. The Genbank proposed gene product name for each gene was assessed in terms of general genomic information, amino acid sequence-based similarity data, and structure-based evidence from the amino acid sequence. Via the comparison across several genetic databases, the students were able to enhance or confirm the suggested function of the gene that the computer generated.

Introduction

Pseudomonas aeruginosa belongs to the Gamma Proteobacteria class and the Pseudomonadaceae family. It is a Gram-negative bacteria that is rod shaped (see Figure 1). It is a free-living bacterium commonly found in water, damp soil, and occasionally on the surface of animals. *Pseudomonas aeruginosa* is a true pathogen of plants, however it has been acknowledged as a pathogen of clinical relevance. It uses aerobic respiration and has requires little nutritional sustenance. The bacteria is complicit to various habitats, which contributes to its ecological success.

It is imperative to study genes, especially infectious bacteria, in order to have knowledge on how to progress our studies across gene research. For example, *Pseudomonas aeruginosa* is extremely resistant to antibiotics, which therefore makes it dangerous to humans and other animals. It is resistant due to its strong outer membrane and its natural immunity to antibiotics, caused by its contact with naturally occurring antibiotics within the soil.

Pseudomonas aeruginosa is one of the worst bacterial pathogens that can infect humans. It can infect the majority of tissues within humans. It can cause urinary tract, soft tissue, bone, joint, gastrointestinal and respiratory systems infections. In addition, people who are suffering from a *Pseudomonas aeruginosa* infection are often also suffering from severe burns, AIDS or cancer. This can cause a great problem within hospitals, with an infection rate of 4 out of 1000 hospital patients and a death rate of about 50 percent of those infected (Todar, 2012).

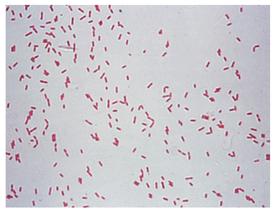

Figure 1: Gram Stain of *Pseudomonas aeruginosa* cells

Methods

Modules of the GENI-ACT (http://www.geni-act.org/) were used to complete *Pseudomonas aeruginosa* 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?

Results

PA0029:
Based on the top two BLAST hits, the probable gene coded was a sulfate transporter. The top COG hit was named SUL1, which is a sulfate permease or related transporter. The TIGRfam hit was named sulP: Sulfate permease and the PFAM hit was named Sulfate transporter. Through all of these findings, the probable enzyme coded for is permease, which is used to transport sulfate.

PA0030:
PA0030 was predicted to be of an unknown function by the computer pipeline annotation. Based on my research and comparisons in modules 1 through 3, it is now likely that the gene codes for a binding protein. Based on the similar organisms found in the BLAST and their CDD's, the protein had a function similar to the proline/glycine betatine transporter. The TIGRFAM suggests that it may have been a choline binding transporter. PFAM determined the gene to code for a protein of the periplasmic binding protein clan, which is consistent with the data collected throughout the modules. In conclusion, the gene PA0030 likely codes for a choline or betatine transporter with a main function of binding and transporting materials in the cell.

PA0031:
Supported by BLAST, PA0031 was found to be choline-sulfatase which functions as a catalyst. The only COG hit showed another sulfatase related enzyme. For TIGRfam, the only hit resulted in choline-sulfatase. Both Pfam hits showed choline-sulfatase. This gene has been found to code for a catalyst.

PA0032:
PA0032 has consistently been regarded as, or related to a transcriptional activator (BLAST hits) or transcriptional regulator (COG hit). It is confirmed as a Periplasmic Binding Protein. This is consistent with findings from other modules, as the term "transcriptional" has been used to describe it numerous times.

PA0033:
Based upon the data, the most probable function of this protein is likely an Htp or Histidine phosphotransferase domain protein, rather than being a hypothetical protein as predicted by the computer. This is supported by the CDD and BLAST results. Both show the protein being closely related to other Htp proteins. These proteins possess a phosphorylatable histidine residue and are responsible for transferring a phosphoryl group from an aspartate residue on an intermediate "receiver" domain. These proteins are observed more in bacteria due to its function, while hardly ever been observed in multicellular animals

PA0034:
By using BLAST, PA0034 was found to be a putative transcriptional regulator. There was only one hit for COG which was a DNA-binding response regulator. The two TIGRFAM hits supported this data. The first hit being a Phosphate regulon transcriptional regulator, and the second hit being a Transcription factor. Only one hit was present when PFAM was used and it was a Response regulator receiver domain. This gene codes for a transcriptional activator or regulator.

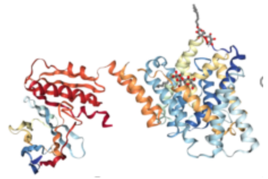

Figure 2: Structure of the SLC26 transporter SLC26Dg in complex with a nanobody

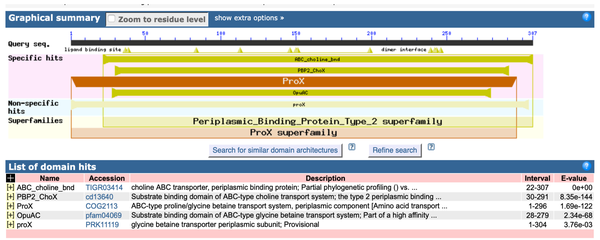

Figure 3: CDD Results for PA0030

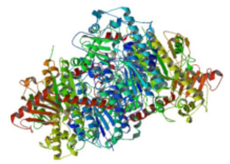

Figure 4: Crystal structure of a choline sulfatase from *Sinorhizobium melliloti*

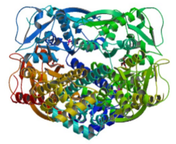

Figure 5: Structure of the full-length CcmR complexed with 2-OG from Synechocystis PCC6803 (transcriptional protein)

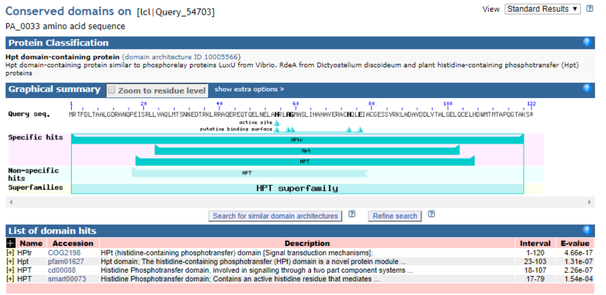

Figure 6: CDD Results for PA0033

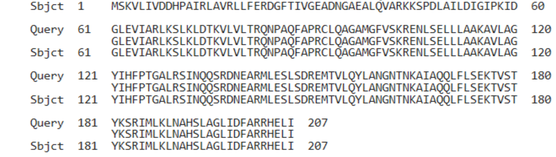

Figure 7: Top BLAST hit for PA0034

Conclusion

The GENI-ACT proposed gene product did not differ significantly from the proposed gene annotation for 4 of the 6 genes in the group. There were two hypothetical proteins for which we now propose alternative annotations through our research. A table is shown below to summarize original pipeline annotations and final proposed annotations.

Locus Tag	Geni-Act Product Name	Proposed Annotation
PA0029	Sulfate Transporter	Sulfate Transporter
PA0030	Hypothetical Protein	Choline or Betatine Transporter
PA0031	Choline Sulfatase	Choline Sulfatase
PA0032	Transcriptional Regulator	Transcriptional Activator/Regulator
PA0033	Hypothetical Protein	Histidine Phosphotransferase Domain Protein
PA0034	Two-Component Response Regulator	Putative Transcriptional Regulator

References

Todar (2012). *Pseudomonas aeruginosa*. Todar's Online Textbook of Bacteriology,1-4.

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 - R25ODO10536