Annotation of the *Kytococcus sedentarius* Genomes at Locus Tags KSED_RS02465 and KSED_RS02445)

Michael Fernandez, Oscar Gaile, Jacky Hua, Akash Jacobus and Dr. Paul A. Weber
Amherst Central High School, Amherst, NY 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 group of 2 genes from the microorganism *Kytococcus sedentarius* (KSED_RS02465 & KSED_RS02445) 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. Our suggested name differs from the initially proposed ones, as the aforementioned methods suggest certain functions.

Introduction

Kytococcus sedentarius is a marine dwelling, Gram Positive bacteria, well known for the production of the polyketide antibiotic. It is strictly aerobic and can only grow when the necessary amino acids are provided. The completed 2,785,204 nucleotide sequence makes it perhaps the smallest of the Actinomycetes order. Usually found in nature as tetrads, irregular clusters or a combination of 8 bacterium, *Kytococcus sedentarius* is of catalase positive, oxidase positive, and exhibits optimum cell growth at 25-37 C. It is primarily isolated from the human skin, and is not known for harmful diseases in humans. The polyketide are a type of secondary metabolite, organic compounds not directly involved with the normal growth and development of the organism. As a source of natural antibiotics of which have the ability to be easily massed produce, *Kytococcus sedentarius* poses as a large interest to scientists as an easy and cost effective source of medicine.

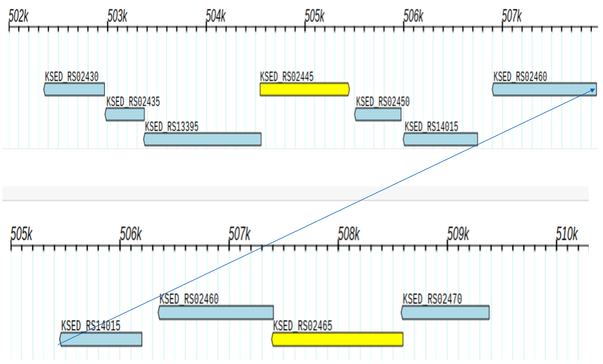


Figure 1. The locus tags and relative position of the genes under investigation in this research

Methods

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

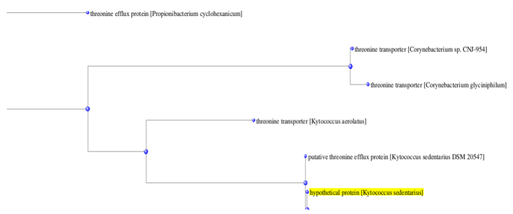
Modules	Activities	Questions Investigated
Basic Information	DNA Coordinates and Sequence, Protein Sequence	What is the sequence of the gene and protein? Where is it located in the genome?
Sequence-Based Similarity	Blast, CDD, T-Coffee, WebLogo	How similar is the protein under investigation to other proteins in GenBank?
Structure-Based Similarity	TIGRfam, Pfam, PDB	What functional domains are present in the protein under investigation?
Cellular Localization (Not functional)	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 (Not functional)	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 (Not functional)	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

KSED_RS02465:
Top BLAST hits propose the idea of KSED_RS02465 as acting as a peptidase for the particular bacteria, hydrolyzing pepsin bonds. This can be supported by the PH conditions of the specific gene, hanging around 4.84, suggesting it works on an acid.

KSED_RS02445:
Top BLAST hits propose the idea of KSED_RS02445 as acting as a translocator as a part of the LysE family. This is supported by Pfam structural analysis, which proposes the protein being part of only the LysE family. LysE family translocators export amino acids out of the cell to maintain homeostasis (A process known as transmembrane efflux). Top BLAST hits also suggest that this protein may act on threonine, though this is not as supported.

Figure 2– Phylogenetic tree



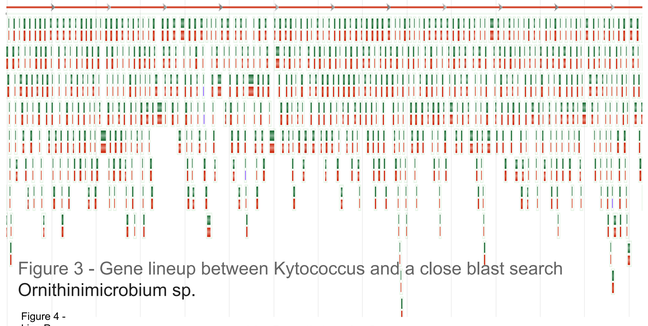


Figure 3 - Gene lineup between Kytococcus and a close blast search Ornithinimicrobium sp.

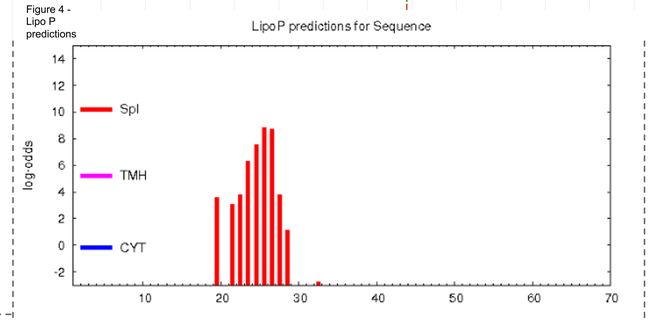


Figure 5– Domain Organization that include LysE family (Pfam)

KSED_RS02465:
The initial proposed product of this gene was not concluded. After research was done on the BLASTn, top results show well correlated functions in the nucleotide and amino acid sequences. One bacterium, Ornithinimicrobium sp, had the E-Value closest to the specific gene we are looking for yet develops no signs of correlation between function. The LipoP and SignalP websites either through human error or through other means did not give the research team enough data to prove the location of the given protein. Also, again through either human error of website malfunction, a phylogenetic tree was not able to be generated for either of the proteins studied.

KSED_RS02445:
Evidence exists to relate KSED_RS02445 to that of the LysE family of translocators. Some evidence also suggests this to specifically by a threonine translocator. However, Pfam domain organization (Fig. 5) suggests that threonine translocators have Cytochrome_B561 as well, though LysE makes the vast majority of it. As a result, the protein can neither be confirmed nor ruled out as a threonine translocator.

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.

Locus Tag	Pipeline Annotation Product Name	Proposed Annotation	Changes Proposed?
KSED_RS02465	CDS hypothetical protein	Peptidase	Yes
KSED_RS02445	Hypothetical Protein	LysE Family Translocator	Yes

References

Sims et al. (2009). Complete genome sequence of *Kytococcus sedentarius* type strain (541T). Standards Genomic Sciences,12 - 20.
National Center for Bioinformatics (Updated 2019). “(complete genome sequence)”

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