University at Buffalo The State University of New York | REACHING OTHERS

Annotation of the *Kytococcus sedentarius* Genome from DNA Locus Tags Ksed_07900 to Ksed_07960

Erin Migas, Joel Mercado, Melissa Morga, Bailey Kinn, Sarah Suarez, and Jean Baker

Frontier Central High School and the Western New York Genetics in Research Partnership

Abstract

A group of five sequential genes from the marine dwelling microorganism *Kytococcus sedentarius* (Ksed_07900 through Ksed_07960) 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, cellular localization data, potential alternative open reading frames, structure-based evidence from the amino acid sequence, enzymatic function, and the presence or absence of gene duplication and degradation. The Genbank proposed gene product name for the genes studied did not differ significantly from the proposed gene annotation and appear to be correctly annotated by the database.

Introduction

Kytococcus sedentarius (figure 1) is an opportunistic pathogen that is gram positive and used to produce the antibiotics monesin A and B, which are used extensively in the beef and dairy industries to improve and promote growth. (phosphil. et al., 1998.) It is known to cause Pitted Keratolysis (figure 2), a conditions that creates crater-like pits on the soles of the feet and forms a strong odor. It affects those who sweat profusely, especially in those who wear boots or shoes for an extended period of time. (DermNet,1998)

A team of five students from Frontier Central High School as part of a larger Western New York group were randomly chosen from a list of twelve volunteers to research this relatively obscure bacteria. Utilizing programs such as T-Coffee, WebLogo, TIGRfam, and Pfam, the students analyzed the genes from Ksed_07900 to Ksed_07960 discovering the gene products produced by each gene. The gene neighborhood is shown in figure 3. A better understanding of the gene product could lead to potential medical applications.

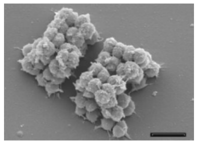

Figure 1: Kytococcus sedentarius is gram positive, opportunistic pathogen.

Figure 2: Pitted Keratolysis is the most common condition caused by Kytococcus sedentarius.

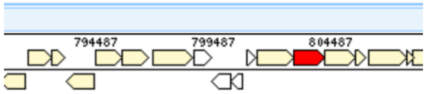

Figure 3: Gene Neighborhood of Ksed_07900 through Ksed_07960

Methods and Materials

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
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?
Module 6- Enzymatic Function	KEGG, MetaCyc, E.C. Number,	In what process does my protein take part?
Module 7- Gene Duplication/ Gene Degradation	Paralog, Pseudogene	Are there other forms of my gene in the bacterium? Is my gene functional?

Results

Kytococcus sedentarius_07900: Between amino acids 161 and 192, WEBLOGO results show a high level of conservation for Ksed_07900 in figure 4.

Figure 4: WebLogo for ksed_07900. WebLogo is a computerized database that generates sequence logos that are graphical representations of amino acid multiple sequence alignments. These consist of stacks of letters of various heights to represent the frequency of sequence conservation at different positions. Each letter represents an amino acid sequence with the larger letters showing greater conservation.

Kytococcus sedentarius 07930: The TIGRFAM name proposed for ksed_07930 is D-xylose ABC transporter, ATP-binding. ABC transporters are involved in the export or import of a wide variety of substrates ranging from small ions to macromolecules. The major function of ABC import systems is to provide essential nutrients to bacteria. They are found only in prokaryotes and their four constitutive domains are usually encoded by independent polypeptides (two ABC proteins and two TMD proteins). Prokaryotic importers require additional extracytoplasmic binding proteins (one or more per systems) for function. In contrast, export systems are involved in the extrusion of noxious substances, the export of extracellular toxins and the targeting of membrane components. They are found in all living organisms and in general the TMD is fused to the ABC module in a variety of combinations. Some eukaryotic exporters encode the four domains on the same polypeptide chain.(wikipedia)

Kytococcus sedentarius 07940:
The proposed annotation by Genbank for Ksed_07940 was a nucleoside ABC transporter membrane protein. Using BLAST search, this gene was found to be a significant match to an ABC transporter permease. This was confirmed by CDD and T-Coffee. PSORT-B found the cellular location of the gene to be in the cytoplasmic membrane. The Phobius probability graph in figure 5 indicates that there are several predicted transmembrane helixes and no predicted signal peptides in the gene.

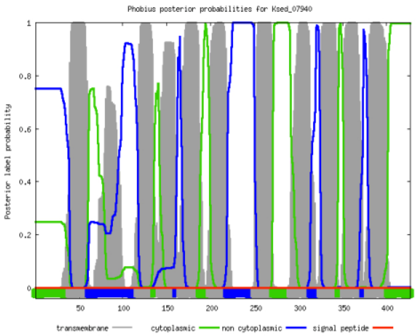

Figure 5: This Phobius probability graph showed that gene Ksed_07940 had several transmembrane helixes and no predicted signal peptides.

Kytococcus sedentarius 07950: The top gene product proposed by BLAST for Ksed_07950 is ABC transporter permease. This was the top hit with an E-value of 4e-157. The functions of this enzyme are similar to those of the proposed product of Ksed_07930. It most likely acts as a dimer and maybe located in the membrane of the organism itself or the organelles within *Kytococcus sedentarius*(InterPro). According to TMHMM, it has 10 predicted transmembrane helices. In addtion, as illustrated in the graph below, this gene has no observable signal peptides.

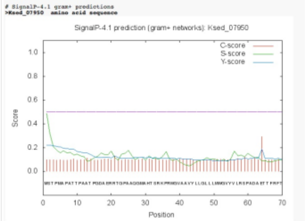

Figure 6: A SignalP prediction graph of Ksed_07950 indicated that there are no signal peptides in this gene.

Kytococcus sedentarius 07960:
The top BLAST hit was the gene product named Cytidine deaminase [Tetrasphaera japonica T1-X7] with an e value of 2e-57. This small, homotetrameric zinc metalloprotein is found in humans and most bacteria. A related, homodimeric form with a much larger subunit is found in E. coli and in Arabidopsis. Both types may act on deoxycytidine as well as cytidine.

WEBLOGO showed that in the beginning of the sequence there is weak conservation of amino acids. From 38 to 119, there is a strong conservation of the amino acid. At the carboxi-terminal there is also weak conservation of amino acid.

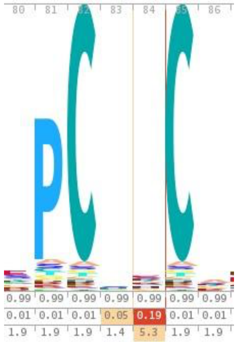

Figure 7: HMM Logo is used to visualize amino acids conserved among various sequences in the Pfam domain. The larger the letters the more conserved this residue is in the protein family. This figure shows the HMM Logo for Ksed_07960

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.

References

Sims et al. (2009). Complete genome sequence of Kytococcus sedentarius type strain (541T). *Standards in Genomic Sciences*, 12 - 20.

http://www.pittedkeratolysis.com/wp-content/uploads/2012/03/pitted-keratolysis-sticky-socks.jpg

http://genome.jgi.doe.gov/kytse/kytse.jpg

https://en.wikipedia.org/wiki/ATP-binding_domain_of_ABC_transporters

Pitted keratolysis (2016, March 16) http://www.dermnetnz.org/bacterial/pitted-keratolysis.html

https://www.ebi.ac.uk/interpro/entry/IPR011864

Acknowledgments

Supported by NSF ITEST Strategies Award Number 1311902.
Dr. Stephen Koury, University of Buffalo.