

Annotation of the *Kytococcus sedentarius* Genome from DNA Coordinates 05560 to 05610

Miranda Allen, Tyler Edwards, Sara Maxon, Emily Blanchard, and Benjamin DeJonge
Albion High School and the Western New York Genetics in Research Partnership

Abstract

A group of 8 genes from the microorganism *Kytococcus sedentarius* (Ksed_05560 – Ksed_05640) 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, and the possibility of horizontal gene transfer. Two of the annotated genes were found to be transmembrane while the remaining six were cytoplasmic. There was no evidence supporting that the genes as called by Genbank were done so incorrectly.

Introduction

Kytococcus sedentarius, a Gram-positive bacteria first discovered in a marine environment, has also been found on human skin (Sigs, 2009). This bacteria is considered relatively harmless for most humans, however, work done by Longshaw et al. (2002), suggests that *Kytococcus sedentarius* has been found to produce two callus digesting enzymes. This may explain in part a cause of the human condition known as pitted keratolysis. In rare cases, *Kytococcus sedentarius* has been found to cause fatal pneumonia in immunosuppressed patients (Chaudary et al. 2010). Species of *Kytococcus* are normally found to be resistant to penicillin and methicillin (Chaudary et al. 2010).

A genome is a set of instruction produced by the DNA and uses a set of four letter sequence, adenine, cytosine, guanine, and thymine. Ksed was first sequenced in 2009 by D. Sims et al. The ksed genome is 2,758,024 bp long, single replicon genome and has 2639 protein-coding and 64 RNA genes. *Kytococcus* is a member of the *Dermaococcaceae* family. (Sigs, 2009). To better keep track of genes in a genome, researchers assign locus tags to indicate the potential starting position of each gene. Locus tags are systematic unique identifiers assigned to each gene. The genes studied in this project as part of the Western New York Genetics in Research Partnership include ksed_05560 through ksed_05640.

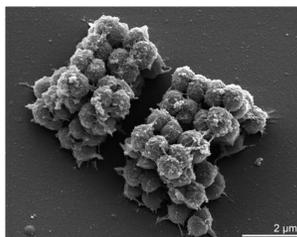


Figure I. Electron micrograph of *Kytococcus sedentarius*. (Sigs 2009)

Figure II. Gene neighborhood from IMG-edu. The genes studied are shown in red.



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?
Module 8- Evidence for Horizontal Gene Transfer	Phylogenetic Tree,	Has my gene co-evolved with other genes in the genome?
Module 9- RNA	RFAM	Does my gene encode a functional RNA?

Results

The results of using Kyoto Encyclopedia of Genes and Genomes showed that the called gene ksed_05620 most closely resembles the orotate phosphoribosyltransferase protein found in *Streptomyces thermophilinus* and *Dermaococcus nishinomiyaensis*. This enzyme has been given the EC number 2.4.2.10 and is involved in the metabolism of pyrimidine.

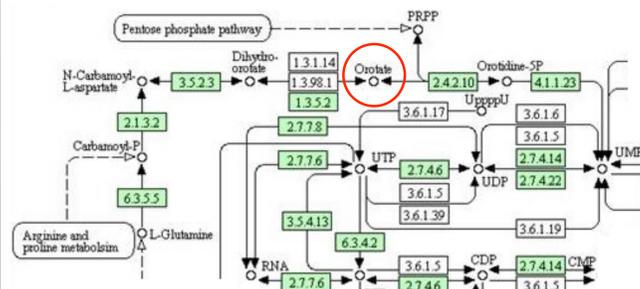


Figure III. Part of the pathway of pentose phosphate, the pathway orotate phosphoribosyltransferase is involved in. Target enzyme is circled in red.

Part of this research project included using TMHMM, PSORT, and SignalP to determine where in the bacteria the called genes might reside. Of the genes studied, most were found to be cytoplasmic proteins, the exception being ksed_05630.

The Markov Model prediction tool indicated that ksed_05630 most likely is a transmembrane protein. The annotation process revealed that the called gene ksed_05630 most likely is a member of the major facilitator superfamily MFS_1. This superfamily is known to exist in *Arthrobacter* and *Tetrasphaera japonica*.

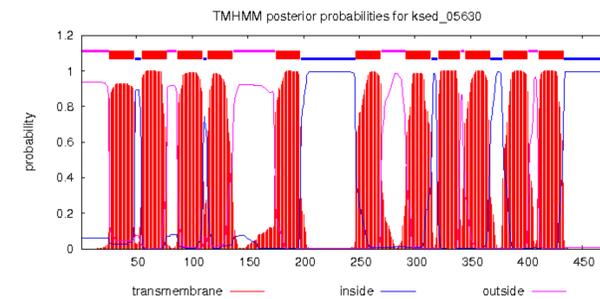


Figure IV. Results of the transmembrane helices Markov model for ksed_05630, indicating the transmembrane nature of the MFS_1 superfamily.

The results from T-Coffee, using Phylogeny.fr coupled with BLAST data, indicate that the ksed_05630 gene seems to stand alone. It should be noted that other modules of this bioinformatics project have indicated a similarity in genes to other bacteria. This discrepancy may be due to the Phylogeny.fr database lacking curated information about the *Kytococcus* genome.

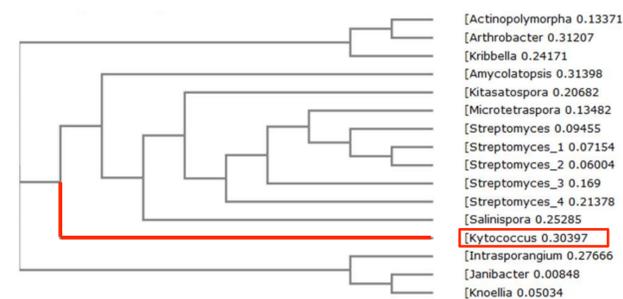


Figure V. The generation of a cladogram for ksed_05630 using alignment data from T-Coffee to examine the microbial evolution of this *Kytococcus* protein in relation to clusters of bacteria

The ksed_05590 gene appears to be well conserved as it appears in several other bacteria. The neighborhood of the gene locus was examined for these bacteria using data from IMG/EDU for evidence of horizontal gene transfer. The results indicate the ksed_05590 gene arose independently from other organisms and was not horizontally transferred because of the lack of similar genes on either side of the target gene when comparing it to multiple species. Vertical transfer suggest that since the same gene is found in multiple bacteria this gene is environmentally favorable.

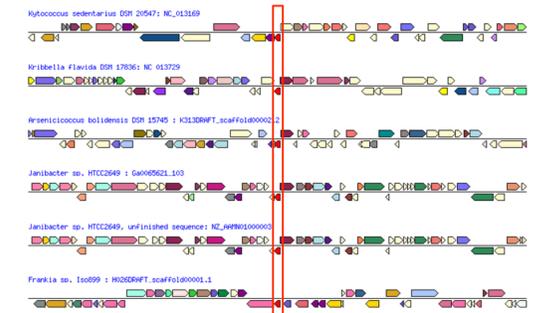


Figure VI. Results from IMG/EDU showing dissimilar neighborhoods of ksed_05590 and other bacterial. The red box contains the target gene.

Conclusion

With the exception of one gene, all were found to be cytoplasmic. Gene 5590 was found to be a conserved hypothetical protein, whose exact function is unknown.

Gene Locus Tag	Top Blast Species/ Gene Product Name	Cellular Localization
5560	<i>Serinicoccus profundi</i> Dehydrogenase	Cytoplasm
5580	<i>Knoellia flava</i> Exodeoxyribonuclease III	Cytoplasm
5590	<i>Tetrasphaera australiensis</i> Conserved Hypothetical Protein	Cytoplasm
5600	<i>Kribbella flava</i> Major facilitator superfamily MFS_1	Transmembrane
5610	<i>Intrasporangium chromatireducens</i> Short Chain Dehydrogenase	Cytoplasm
5620	<i>Dermaococcus nishinomiyaensis</i> Orotate Phosphoribosyltransferase	Cytoplasm
5630	<i>Arthrobacter sp. FB24</i> Major facilitator superfamily MFS_1	Transmembrane
5640	<i>Serinicoccus profundi</i> rRNA Methyltransferase	Cytoplasm

References

Sims et al. (2009). Complete genome sequence of *Kytococcus sedentarius* type strain (541T). *Standards in Genomic Sciences*, 12 - 20.
Longshaw et al. (2002). *Kytococcus sedentarius*, the organism associated with pitted keratolysis, produces two keratin-degrading enzymes. *Journal of Applied Microbiology*, 93 (5).
Chaudhary, D., Finkle, S.N. 2010. Peritoneal Dialysis-Associated Peritonitis Due to *Kytococcus sedentarius*. *Peritoneal Dialysis International* 30(2):252-253.

Acknowledgments

Supported by NSF ITEST Strategies Award Number 1311902
Dr. Rama Dey-Rao Clinical Assistant Professor, University at Buffalo
Dr. Stephen Koury Research Associate Professor, University at Buffalo