

Annotation of the *Kytococcus sedentarius* Genome from Locus Tags Ksed_07450 to Ksed_07490

Boris Chan, Kevin Grimes, Amhajah Jackson, and Omer Bitar

Lackawanna High School, and the Western New York Genetics in Research Partnership

Abstract

The investigation of three genes (Ksed_07450, Ksed_07480, and Ksed_07490) from the marine bacteria, *Kytococcus sedentarius*, have been manually annotated with the assistance of the GEN-ACT (Genomics Educational National Initiative – Annotation Collaboration Toolkit). The Genbank database uses an automatic method of predicting gene presence and function. Manual annotation is used to confirm, reject, or reassign the name and function of the investigated gene. Annotation is achieved by comparing proteins of other organisms by examining similarities in amino acid sequences (sequence based similarity data), structure based evidence, and cellular localization data. In most cases, the annotated genes did not differ from the Genbank predicted function. However, one hypothetical gene (Ksed_07450) may actually be alanine racemase.

Introduction

Kytococcus sedentarius is a strictly aerobic, non-motile, non-encapsulated, and non-endospore forming gram positive coccoid bacterium, found predominantly in tetrad formation. This organism is classified as a chemoheterotroph, a strict aerobe, and requires methionine and several other amino acids for growth. Originally isolated from a microscope slide submerged in sea water in 1944, *Kytococcus sedentarius* grows well in sodium chloride concentrations less than 10% (w/v).

According to Sims et al. (2009), *Kytococcus sedentarius* is a microorganism of interest for several reasons. This bacterium is a natural source of the oligopeptide antibiotics monensin A and monensin B (Sims et al., 2009). *Kytococcus sedentarius* has been implemented as the etiologic agent of a number of opportunistic infections including valve endocarditis, hemorrhagic pneumonia, and pitted keratolysis (Sims et al., 2009). Finally, the phylogeny of this microorganism is a source of interest, as it is a member of the family *Dermaoocaceae* within the actinobacterial suborder *Micromicrococinae*, which has yet to have been thoroughly studied utilizing bioinformatics (Sims et al., 2009).



Figure I. Gene neighborhood from IMG-edu. The genes studied are in the red ovals

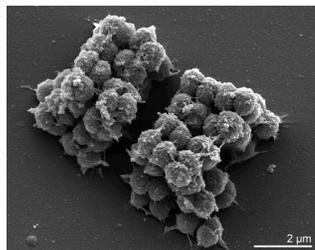


Figure II. Scanning electron micrograph of *Kytococcus sedentarius* (Sims et al., 2009)

Methods and Materials

Modules of the GEN-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?

Results

*Kytococcus sedentarius*07450:

The initial proposed product of this gene by GEN-ACT was a hypothetical protein. Upon further investigation, the second BLAST hit for the amino acid sequence, the presence of well-curated protein functional domains within the amino acid sequence, and the cellular location of the amino acid sequence were all indicative of **alanine racemase** as the protein. As such, the proposed annotation is a **alanine racemase**.

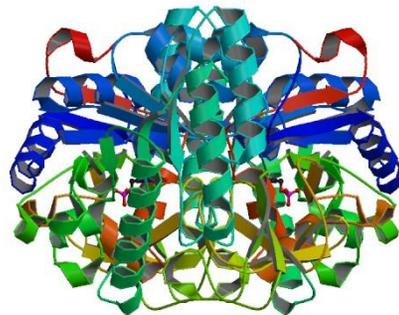


Figure III. 3D image of alanine racemase.

*Kytococcus sedentarius*07480:

The initial proposed product of this gene by GEN-ACT was **tRNA threonylcarbamoyladenosine biosynthesis protein TsaB**. This gene product proposal was supported by the top BLAST hits for the amino acid sequence, the presence of well-curated functional domains within the amino acid sequence, the cellular location of the amino acid sequence, and the enzymatic function of the amino acid sequence. As such, the proposed annotation is a **tRNA threonylcarbamoyladenosine biosynthesis protein TsaB**.

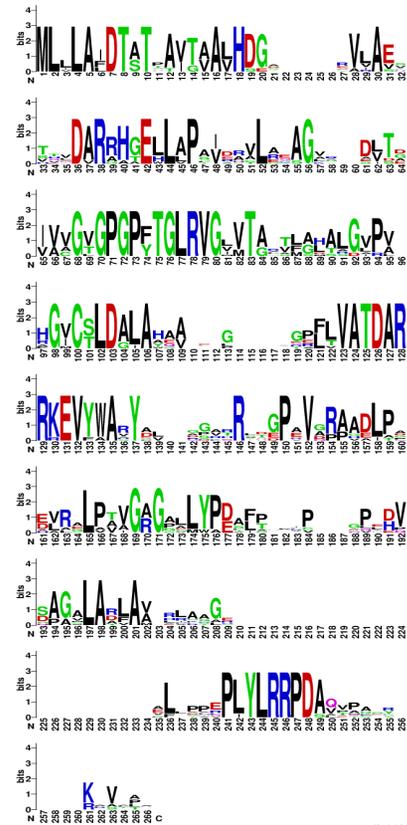


Figure IV. WebLogo of T-Coffee Multiple Sequence Alignment output for Ksed_07480 showing regions of high conservation of amino acid sequences.

*Kytococcus sedentarius*07490:

The initial proposed product of this gene by GEN-ACT was **ribosomal-protein-alanine N-acetyltransferase RimI**. This gene product proposal was supported by the top BLAST hits for the amino acid sequence, the presence of well-curated functional domains within the amino acid sequence, the cellular location of the amino acid sequence, and the enzymatic function of the amino acid sequence. As such, the proposed annotation is **ribosomal-protein-alanine N-acetyltransferase RimI**.

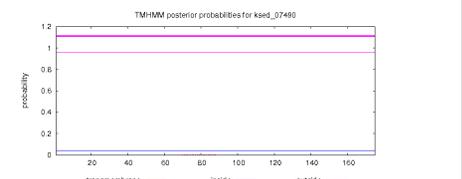


Figure V. TMHMM output displaying the absence of a transmembrane protein in Ksed_07490.

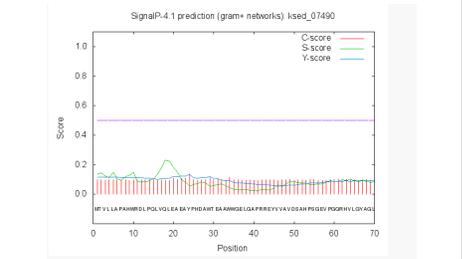


Figure VI. SignalP output displaying the absence of signal peptide in Ksed_07490.

Conclusion

Locus Tag	BLAST Prediction	Proposed Annotation
Ksed_07450	Hypothetical protein	Alanine Racemase
Ksed_07480	tRNA threonylcarbamoyladenosine biosynthesis protein TsaB	tRNA threonylcarbamoyladenosine biosynthesis protein TsaB
Ksed_07490	ribosomal-protein-alanine N-acetyltransferase RimI	ribosomal-protein-alanine N-acetyltransferase RimI

References

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

Acknowledgments

Many thanks to Dr. Stephen Koury, Dr. Rama Dey-Rao, and Danise Wilson for their guidance, support and dedication to this project. Supported by NSF ITEST Strategies Award Number 1311902