

Annotation of the *Kytococcus sedentarius* Genome from DNA Coordinates 474856 to 475971 and 476000 to 477214

Cheyenne Wick, Emma Herman and Melanie Roth

Warsaw Central School and the Western New York Genetics in Research Partnership

Abstract

Two consecutive genes from the microorganism *Kytococcus sedentarius* (Ksed_04800 and Ksed_08410) were annotated using the collaborative genome annotation website GENI-ACT. The Genbank proposed gene product for each gene was assessed in terms of the general genomic information, amino acid sequence, cellular localization data, potential alternative open reading frames, and enzymatic function and presence or absence of gene duplication and degradation. The Genbank proposed gene product name did not differ significantly from the proposed gene annotation by the computer database. Both genes were found to reside in the cytoplasm and are transferase enzymes.

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, as it requires methionine and several amino acids for growth. Originally isolated from a microscope slide submerged in sea water in 1944, *Kytococcus sedentarius* grows well in sodium chloride at concentrations less than 10% (w/v) (Sims et al., 2009).

According to Sims et al. (2009), *Kytococcus sedentarius* is a microorganism of interest for several reasons. This bacterium is a natural source of the oligoketide antibiotics monesin A and monesin B (Sims et al., 2009). *Kytococcus sedentarius* has been implemented as the etiological 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 *Dermaprococcaceae* within the actinobacterial suborder *Micrococccineae*, which has yet to have been thoroughly studied utilizing bioinformatics (Sims et al., 2009).

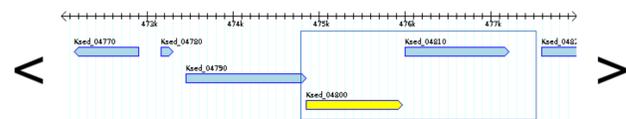
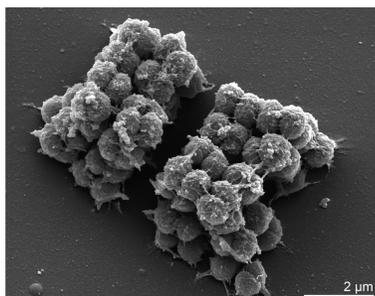


Figure 1: Gene neighborhood of Ksed_04800 and Ksed_04810

The manual genome annotation of *Kytococcus sedentarius* was carried out using GENI-ACT (Genomics Education National Initiative Annotation Collaboration Toolkit). Using PSORTb and MetaCyc, both genes were found to reside in the cytoplasm and to be transferase enzymes.



Kytococcus sedentarius
(<http://www.standardsgenomics.org/index.php/signet/article/viewFile/7/51/333>)

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

Ksed_04800	
Basic Information	DNA Coordinates: complement 474856..475971
	DNA Length: 1116 base pairs
Sequence Based Similarity	Amino Acid Sequence Length: 371 amino acids
	COGs: COG2021 Homoserine acetyltransferase [amino acid transport and metabolism]
Cellular Localization	T-Coffee: amino end is highly conserved
	WEBLOGO: most common amino acids are polar and hydrophobic
	Transmembrane Helices: none
Alternative ORFs	Signal Peptide: none
	PSORTB Final Prediction: cytoplasmic
Structure Based Evidence	TIGRFAMs: TIGR01392 homoserine O-acetyltransferase
	PFAMs: PF00561
	HMM Logo Key Residues: G164, S124, G160
Enzymatic Function	PDB: 2B61 homoserine transacetylase
	EC Number: 2.3.1.31
Duplication and Degradation	Paralogs: no paralogs found
	Pseudogenes: none



Figure 2: T-Coffee (2 panels, left and right) for Ksed_04800. Stretches of highly conserved amino acids are evident scattered through all but the most amino terminal portions of the alignment.

Ksed_04800 summary:

This protein is confirmed to be located in the cytoplasm because there were no transmembrane helices or signal peptides found. COG and TIGRFAM also concluded that the gene product is in the Homoserine O-acetyltransferase family. WEBLOGO showed conserved amino acids in scattered regions of the alignment (not shown). There were no paralogs found demonstrating no recent gene duplication.

Ksed_04810	
Basic Information	DNA Coordinates: 47600..477214
	DNA Length: 1215 base pairs
Sequence Based Similarity	Amino Acid Sequence Length: 404 amino acids
	COGs: COG 0620 methionine synthase [amino acid transport and metabolism]
Cellular Localization	T-Coffee: generally well-conserved
	WEBLOGO: highly conserved regions at the C-terminus well conserved region at residues 290-301
	Transmembrane Helices: none
Alternative ORFs	Signal Peptide: none
	PSORTB Final Prediction: cytoplasmic
Structure Based Evidence	TIGRFAMs: none
	PFAMs: PF01717 cobalamin-independent synthase
	HMM Logo Key Residues: P10, C218, C301
Enzymatic Function	PDB: -
	EC Number: 1.16.1.8
	EC Name: methionine synthase II (cobalamin-independent)

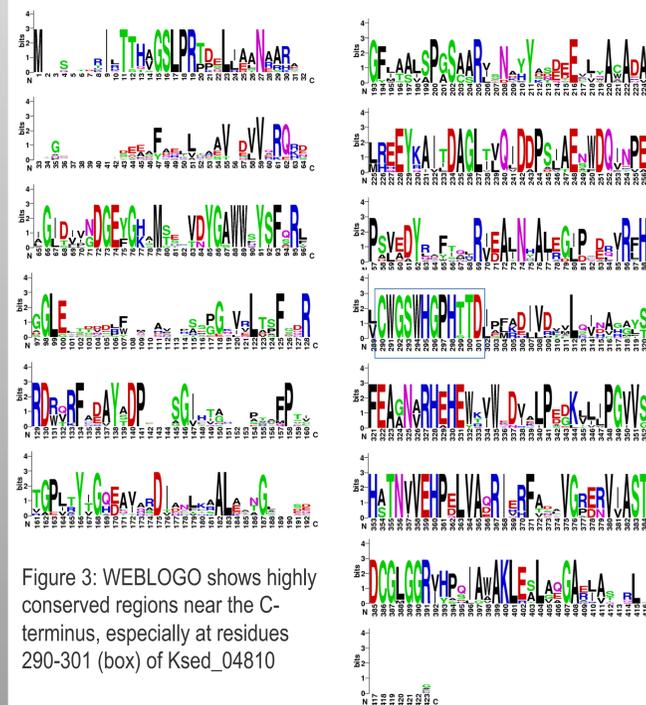


Figure 3: WEBLOGO shows highly conserved regions near the C-terminus, especially at residues 290-301 (box) of Ksed_04810

Ksed_04810 summary:

The gene product is proposed to be an enzyme, methionine synthase as shown through PFAM result PF01717 and MetaCyc. This protein, as shown through the tools in the cellular localization module, was found to be cytoplasmic, and it also lacks both transmembrane helices and a signal peptide. As shown in WEBLOGO and T-Coffee, the amino acid sequence is generally well-conserved. In the WEBLOGO results, segments that are highly conserved are shown throughout the sequence with residues 290-301 being the particularly notable.

Ksed_04800	Ksed_04810
Localization Scores:	Localization Scores:
Cytoplasmic: 9.67	Cytoplasmic: 7.50
CytoplasmicMembrane: 0.01	CytoplasmicMembrane: 1.15
Cellwall: 0.15	Cellwall: 0.62
Extracellular: 0.17	Extracellular: 0.73
Final Prediction: Cytoplasmic	Final Prediction: Cytoplasmic
	7.50

Figure 4: PSORTb output show that both genes are located in the cytoplasm

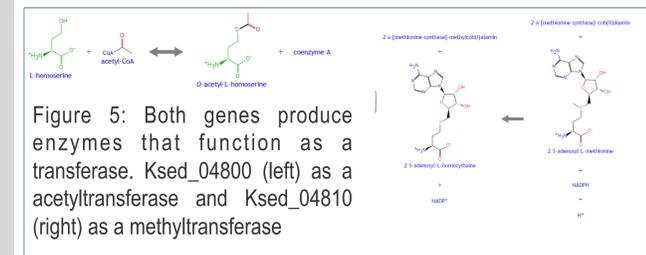


Figure 5: Both genes produce enzymes that function as a transferase. Ksed_04800 (left) as an acetyltransferase and Ksed_04810 (right) as a methyltransferase

Functions of an enzyme:

- Increase reaction rates without being consumed, altered or changing the chemical equilibrium
- Enzymes work to catalyze or speed up chemical reactions
- A transferase enzyme enacts the transfer of specific functional groups from one molecule to another
 - Ksed_04800 – transfers an acetyl group
 - Ksed_04810 – transfers a methyl group

Conclusion

The GENI-ACT proposed gene product did not differ significantly from the proposed gene annotation for each of the genes in the group, the genes appear to be correctly annotated by the computer database.

Gene Locus	Proposed Annotation
04800	Homoserine O-acetyltransferase
04810	Methionine synthase II

Student's benefit from this project:

- Looking into pursuing medical fields
 - Genetic engineering
- Research helped us to understand and prepare for what college and future jobs will be like
- This project allowed us to look into this area of medical career opportunities
 - Allowed us to learn more about the topic

References

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

Acknowledgments

Thank you to Dr. Stephen Koury, Dr. Rama Dey-Rao & Dr. Patricia Masso-Welch for their guidance, support and dedication to this project. Supported by NSF ITEST Strategies Award Number 1311902