

Annotation of the *Kytococcus sedentarius* Genome from Ksed_08920 to Ksed_08950

Joshua Ninan, Erin Washburn*, Madison Crowley* and Meghan Janora
Cattaraugus Little Valley School and the Western New York Genetics in Research Partnership

Abstract

Four genes from the microorganism *Kytococcus sedentarius* (Ksed_08920 – Ksed_08950) were annotated using the collaborative genome annotation website GENI-ACT. The Genbank proposed gene product name for two of the genes was assessed using the proposed modules. Each gene product was assessed at different levels in terms of their 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, and enzymatic function. Thus far, the Genbank proposed gene product names did not differ significantly from the proposed gene annotation for the two annotated genes in the group, and as such, the genes appear to be correctly annotated by the database for the modules that were completed on each.

Introduction

Kytococcus sedentarius, the only known producer of the antibiotics monensin A and B, has been isolated from varying environments, including human skin, groundwater, and even airline cabins. It can be a human opportunistic pathogen. Strain DSM 20547, the type strain, is a free-living, nonmotile, Gram-positive bacterium, originally isolated from a marine environment in about 1944. It grows as spherical cocci and occurs predominantly in tetrads which can be arranged in cubical packets. It is non-encapsulated and does not form endospores, is strictly aerobic and chemoorganotrophic, requires methionine and other amino acids for growth, and grows well in NaCl concentrations up to 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 *Micrococineae*, which has yet to have been thoroughly studied utilizing bioinformatics (Sims et al., 2009).

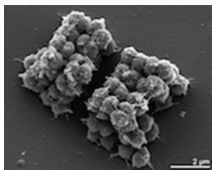
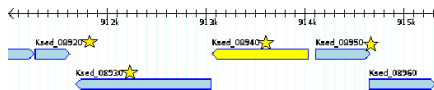


Figure 1. Scanning electron micrograph of *Kytococcus sedentarius*

Figure 2. Location of the locus tags analyzed in this project



Methods and Materials

Modules of the GENI-ACT were used to complete *Kytococcus sedentarius* genome annotation.

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?

Results

Ksed_08940		
Basic Information	DNA coordinates	913066..914037
	DNA length	972
	Amino Acid Sequence Length	323
Sequence-Based Similarity Data	COGS	COG1946 Lipid transport and metabolism
	TCOFFEE	The amino acids are well conserved
	WEBLOGO	Amino acid conservation mostly in the middle and end
Structure Based Evidence	TIGRFAMs	TIGR00189
	PFAMs	4HBT3
	Thioesterase-like superfamily	
	HMM Logo Key Residues	G13, G14, F40, W150, D170, W208, and Q242
	PDB	Crystal structure of acyl-CoA thioesterase tesB from <i>Yersinia pestis</i> in complex with coenzyme A
Cellular Localization Data	Transmembrane Helices	zero transmembrane helices
	Signal Peptide	No signal peptide predicted
	PSORTB Final Prediction	cytoplasmic
Alternative Open Reading Frame		n/a
Enzymatic Function	KEGG	

Ksed_08950		
Basic Information	DNA coordinates	914111..914662
	DNA length	552
	Amino Acid Sequence Length	183
Sequence-Based Similarity Data	COGS	COG0824
	TCOFFEE	Well conserved
	WEBLOGO	Highly conserved at positions 5, 24, 25, and 40.
Structure Based Evidence	TIGRFAMs	n/a
	PFAMs	n/a
	HMM Logo Key Residues	n/a
	PDB	n/a
Cellular Localization Data	Transmembrane Helices	zero transmembrane helices
	Signal Peptide	
	PSORTB Final Prediction	

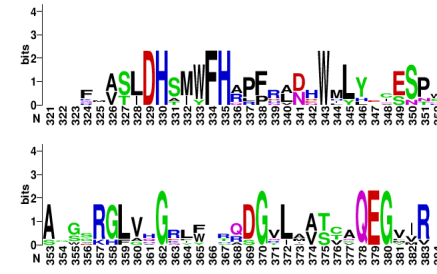


Figure 3 – The WebLogo N321-N384 for *Kytococcus sedentarius* 08940 where most conservation is shown. Other conserved portions not shown were at N49, N76, and N77.

SeqID: Ksed_08940 amino acid sequence

Analysis Report:	
CMSVM+	Unknown
CWSVM+	Unknown
CytoSVM+	Cytoplasmic
ECSVM+	Unknown
ModHM+	Unknown
Motif+	Unknown
Profile+	Unknown
SCL-BLAST+	Unknown
SCL-BLASTe+	Unknown
Signal+	Unknown
Localisation Scores:	
Cytoplasmic	7.50
CytoplasmicMembrane	1.15
Cellwall	0.62
Extracellular	0.73
Final Prediction:	
Cytoplasmic	7.50

Figure 4 – *Kytococcus sedentarius* 08940 Psortb results showing the score of 7.5 as cytoplasmic.

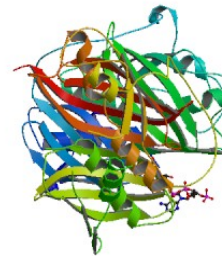


Figure 5 – Crystal structure of acyl-CoA thioesterase tesB from *Yersinia pestis* in complex with coenzyme A. The proposed product of gene 08940 of *Kytococcus sedentarius* may have similar structure.

Kytococcus sedentarius 08940:

The initial proposed product of this gene by GENI-ACT was an acyl-CoA thioesterase. This gene product proposal was supported by the top BLAST hit for the amino acid sequence, acyl-CoA thioesterase II. Sequence Based Similarity supports the proposal by what is shown in the semi-conserved portions of the WebLogo. Not shown in Figure 3 are similarly conserved regions at the beginning of the sequence as well. The CDD (Conserved Domain Database Search) showed a first hit of TesB, Acyl-CoA thioesterase which functions in lipid transport and metabolism. The TIGRFAM database resulted in TIGR00189 tesB: a cyl-CoA thioesterase II. This suggests the gene tesB produces the protein Acyl-CoA thioesterase II which can hydrolyze a broad range of acyl-CoA thioesters. Its physiological function may not be known. Figure 3 shows the PSORTb data to predict where if the protein is soluble in the cytoplasm, and integral membrane protein, or secreted by the cell. The score of 7.50 predicts the location of the protein is cytoplasmic, however, because none of the scores are greater than 7.5 suggests the protein may exist in multiple sites. Figure 4 is an image of the protein top hit for the PDB (Protein Data Bank). It is the protein crystal structure and is classified as a hydrolase. Such, the proposed annotation is an acyl-CoA thioesterase with possible function in lipid transport and metabolism.

Kytococcus sedentarius 08950:

The initial proposed product of this gene by GENI-ACT was a predicted thioesterase. This gene product proposal was supported by the second BLAST hit for the amino acid sequence which was a thioesterase in *Knoellia sinensis*. Supporting the proposal is also the COG hit 0824 FadM which is an Acyl-CoA thioesterase FadM functioning in lipid transport and metabolism. Thus far, the proposed annotation is a thioesterase.

Conclusion

The GENI-ACT proposed gene products did not differ significantly from the proposed gene annotations for each of the genes in the group that were annotated. 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.

Acknowledgments

Supported by NSF ITTEST Strategies Award Number 1311902