

# Annotation of the *Kytococcus sedentarius* Genome from DNA Coordinates 07010 to 07030

Fiona Sullivan, Natalie Zapfel and Paul Weber

Amherst High School and the Western New York Genetics in Research Partnership

## Abstract

A group of consecutive 2 genes from the microorganism *Kytococcus sedentarius* (Ksed\_07010– Ksed\_07030) 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, the possibility of horizontal gene transfer, and the production of an RNA product. The Genbank proposed gene product name 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 in the nr database.

## 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 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 antibiotic kasimonsin 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 Dermacoccaceae within the actinobacterial suborder Micrococcales, which has yet to have been thoroughly studied utilizing bioinformatics (Sims et al., 2009).

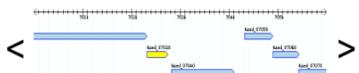


Figure 1. The locus of the *K. sedentarius* genome annotated in this research.

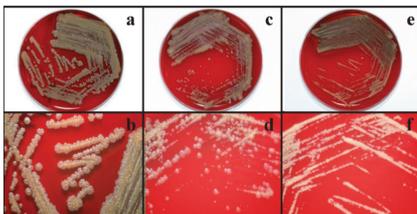


Figure 2. Morphology of a close relative of the *K. sedentarius* (*Kytococcus schroeteri*) on Blood Agar Plates (2).

## 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?

## Results

### *Kytococcus sedentarius*07010:

The initial proposed product of this gene by GENI-ACT was a protein of unknown function. Top BLAST hits in the nr database were all hypothetical proteins. However, the sequence had COG, TIGRFAM and PFAM hits suggestive of possible function. SignalP and Phobius predict the presence of a signal peptide, both TMHMM and Phobius predict transmembrane helices, though TMHMM predicts 3 and Phobius predicts only 1. Psorb predicts Ksed\_07010 to be a cytoplasmic membrane protein. At the very least Ksed\_07010 should be annotated as a membrane protein of unknown function until further information can be obtained.

### *Kytococcus sedentarius*07030:

The initial proposed product of this gene by GENI-ACT was a protein with tetrapyrrole methyltransferase and pyrophosphatase domains. 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 and the cellular localization of the amino acid sequence. The top Swiss-Prot blast hit was Nucleoside triphosphate pyrophosphohydrolase/pyrophosphatase MazG; Short=NTP-PPase. The top COG hit in the conserved domain database was COG3956: YabN; YabN is an uncharacterized conserved protein, containing tetrapyrrole methylase and MazG-like pyrophosphatase domains. The top PFAM hit was pfam03819: MazG. This domain is about 100 amino acid residues in length. It is found in the MazG protein from *Escherichia coli*. It contains four conserved negatively charged residues that probably form a divalent metal binding site. This domain is found in isolation in some proteins as well as associated with pfam00590. This domain is clearly related to pfam01503 another pyrophosphohydrolase involved in histidine biosynthesis (PFAM site description). As such, the proposed annotation is a MazG-like protein with tetrapyrrole methyltransferase and pyrophosphatase domains.

Sequences producing significant alignments:

Alignments	Max score	Total score	Query cover	E value	Ident	Accession
<input type="checkbox"/> hypothetical protein (Kytococcus sedentarius)	1377	1377	100%	0.0	100%	WP_022802126.1
<input type="checkbox"/> hypothetical protein (Shiella asiatica)	354	354	84%	3e-108	43%	WP_020538818.1
<input type="checkbox"/> hypothetical protein (Shiella asiatica)	343	343	83%	2e-103	43%	WP_020538808.1
<input type="checkbox"/> hypothetical protein (Shiella asiatica)	338	338	85%	3e-101	42%	WP_020538872.1
<input type="checkbox"/> hypothetical protein (Dactylospira ferax)	338	338	86%	3e-101	41%	WP_021456621.1
<input type="checkbox"/> hypothetical protein (Dactylospira ferax)	331	331	87%	1e-98	38%	WP_020229961.1

Figure 3. List of top BLAST hits for Ksed\_07010. All are hypothetical.

Conserved domains on [g1256687999g1ACV0576.1]

Name	Accession	Description	Internal	Start	End
TMHMM	g1256687999g1ACV0576.1	TMHMM: Predicts transmembrane helices. The domain is about 100 amino acid residues in length.	13-107	13-107	13-107
Phobius	g1256687999g1ACV0576.1	Phobius: Predicts transmembrane helices. The domain is about 100 amino acid residues in length.	13-107	13-107	13-107
SignalP	g1256687999g1ACV0576.1	SignalP: Predicts signal peptides. The domain is about 100 amino acid residues in length.	13-107	13-107	13-107
PSORT	g1256687999g1ACV0576.1	PSORT: Predicts protein localization. The domain is about 100 amino acid residues in length.	13-107	13-107	13-107

Figure 4. CDD results for Ksed\_07010.

Prediction of Ksed\_07010

ID	Start	End	Function
FT SIGNAL	1	40	
FT TMHMM	1	21	H-REGION
FT PSORT	22	33	H-REGION
FT PSORT	24	40	C-REGION
FT TIGRFAM	41	215	NON-CYTOPLASMIC
FT TRANSFORMER	216	235	
FT TIGRFAM	236	716	CYTOPLASMIC

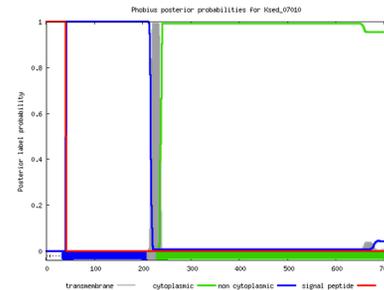


Figure 5. Phobius results for Ksed\_07010. A signal peptide and a single transmembrane helix are predicted. TMHMM predicts 2 additional helices near the carboxy terminus (not shown).

Download - GenPept Graphics

nucleoside-triphosphate pyrophosphatase [Clostridium sp. CAG-813]

Sequence ID: emb|CDE97846.1| Length: 268 Number of Matches: 1

Range	Expect	Method	Identities	Positives	Gaps
131 bits(330)	2e-35	Compositional matrix adjust.	64/130(49%)	89/130(68%)	2/130(1%)

Query 5 ERPGASLARVVEIAHRLRAGDCAWFRQTHASLAPFVLEAEAEVVEIAGDDPAHIAEE 64  
 E+ +E ++++++LK +GC W RQTH SD P +LEEA E +AI E+D AII EE  
 Sbjct 2 ESKYKLEELIDVVAKLAPDQCPDRQRTFTSLFNNLEAEYAVDAIDENHARLEE 61

Query 65 LGDGLVQVVVHAGLAEAEGFTIDVVAELKARLRFVFPDVPVNTAAEVVPLMR 124  
 LGD+L QV+H+Q+R+E+ EPT+DDV EL+ EL+ RRPVIV +A +H+ W  
 LGD+LQV+H+Q+R+E+ EPT+DDV EL+ EL+ RRPVIV +A +H+ W  
 Sbjct 62 LGDGLVQVVVHAGLAEAEGFTIDVVAELKARLRFVFPDVPVNTAAEVVPLMR 119

Query 125 AAKAREAKR 134  
 K+ E+ E+  
 Sbjct 120 KLSSEKTER 129

Figure 6. Top Swiss-Prot BLAST hits for Ksed\_07030.

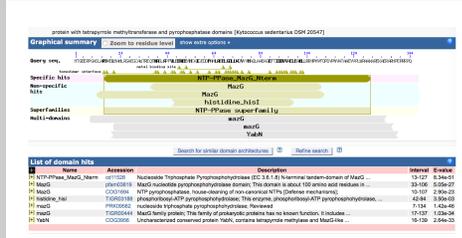


Figure 7. CDD results for Ksed\_07030

Localization Scores:

Cytoplasmic	7.50
CytoplasmicMembrane	1.15
Cellwall	0.62
Extracellular	0.73
Final Prediction:	
Cytoplasmic	7.50

Figure 8. Psorb results for Ksed\_07030. A cytoplasmic location is predicted.

## Conclusion

The GENI-ACT proposed gene product for each of the genes in the table below.

Locus Tag	GENI-ACT Name/Function	Proposed Name/Function
Ksed_07010	protein of unknown function	membrane protein of unknown function
Ksed_07030	protein with tetrapyrrole methyltransferase and pyrophosphatase domains	MazG-like protein with tetrapyrrole methyltransferase and pyrophosphatase domains

## References

- Sims et al. (2009). Complete genome sequence of *Kytococcus sedentarius* type strain (541T). *Standards in Genomic Sciences*, 12-20.
- Infection (2012) 40:567-573

## Acknowledgments

Dr. Rama Dey- Rao, Dr. Stephen Koury  
 Supported by NSF ITEST Strategies Award Number 1311902