

# Annotation of the *Kytococcus sedentarius* Genome from DNA Coordinates 468392 to 471684

Michael Hagelberger, Shawn Kirsch, Joshua Mincer, Elisabeth Rebmann, Lindsay VonOhlen and Laura O'Donnell

Attica High School and The Western New York Genetics in Research Partnership

## Abstract

A group of consecutive 5 genes from the microorganism *Kytococcus sedentarius* (Ksed\_04710 – Ksed\_04750) 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, and potential alternative open reading frames. 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 by the computer 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, as it 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 at 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 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 *Dermacoccaceae* within the actinobacterial suborder *Micrococccineae*, which has yet to have been thoroughly studied utilizing bioinformatics (Sims et al., 2009).

## 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

Ksed_04710		
Basic Information	DNA Coordinates	complement(468392..468940)
	DNA Length	549 BP
	Amino Acid Sequence Length	182 AA
Sequence Based Similarity	COGs	No putative conserved domains have been detected
	TCOFFEE	-
	WEBLOGO	-
Cellular Localization	Transmembrane Helices	-
	Signal Peptide	-
	PSORTB Final Prediction	unknown
Alternative ORFs		none
Structure Based Evidence	TIGRFAMs	No hits
	PFAMs	No hits
	HMM Logo Key Residues	No hits
	PDB	No hits

### *Kytococcus sedentarius* 04710:

There were no hits in sequence based similarity, cellular localization, and structure based evidence. No alternate open reading frames were discovered. With the lack of evidence discovered, it is possible that this gene was called wrong. It may be referred to as a hypothetical protein.

Ksed_04720		
Basic Information	DNA Coordinates	complement(468942..469481)
	DNA Length	540BP
	Amino Acid Sequence Length	179aa
Sequence Based Similarity	COGs	COG1881 Phospholipid-binding protein E value 7.83e-47 No other COG hit was found
	TCOFFEE	It is more conserved in the middle.
	WEBLOGO	It is more conserved in the middle.
	Transmembrane Helices	none
Cellular Localization	Signal Peptide	none
	PSORTB Final Prediction	Cytoplasmic
	Alternative ORFs	
Structure Based Evidence	TIGRFAMs	TIGR00481: Raf kinase inhibitor-like protein
	PFAMs	Phosphatidylethanolamine-binding protein PBP (PF01161)
	HMM Logo Key Residues	G17, P30, D42, D44, HY58, ND63, PK65, YWF90, GP92, PC93, P95, HK102, VL1108,
	PDB	Structure of Rv2140c, a phosphatidylethanolamine binding protein from <i>Mycobacterium tuberculosis</i>

### *Kytococcus sedentarius* 04720:

The protein product for this gene binds ATP, opioids and phosphatidylethanolamine. This gene is thought to be a serine protease inhibitor, the cleavage of peptide bonds in proteins. The gene is proposed to inhibit kinase activity of RAF1 by inhibiting its activation and by dissociating the RAF1/MEK complex. It's also thought to be a cytoplasmic protein. This gene is proposed to be a phospholipid binding protein. This gene was found to exist in eukaryotes as well.

Ksed_04730		
Basic Information	DNA Coordinates	469652..470242
	DNA Length	591bp
	Amino Acid Sequence Length	196aa
Sequence Based Similarity	COGs	No cog hits
	TCOFFEE	Amino acid well conserved except for the beginning and the end
	WEBLOGO	well conserved except for 3-10, 87-90, 163-172, 200-203 highly conserved from 35-86, 125-145
Cellular Localization	Transmembrane Helices	none
	Signal Peptide	none
	PSORTB Final Prediction	Cytoplasmic
Alternative ORFs		none
Structure Based Evidence	TIGRFAMs	family protein (TIGR03086) uncharacterized Actinobacterial (TIGR03083)
	PFAMs	MDMPI_N (PF11716)
	HMM Logo Key Residues	w29, h37, e131, h135, d138
	PDB	no hits

### *Kytococcus sedentarius* 04730:

The protein product predicted for this gene is a cytoplasmic enzyme based on its TIGR03083. TIGR03083 is a family of uncharacterized proteins that are predicted to have enzymatic functions. Its PFAM, PF11716 is a member of the super family DinB\_2 which is an uncharacterized family of potential enzymes. It is also predicted to be cytoplasmic because it contained no transmembrane helices or signal peptides. There were no COG hits and according to TCOFFEE and WEBLOGO it was fairly well conserved.

Ksed_04740		
Basic Information	DNA Coordinates	470246..471361
	DNA Length	1116 bp
	Amino Acid Sequence Length	371aa
Sequence Based Similarity	COGs	COG3616 [Amino acid transport and metabolism]
	TCOFFEE	Amino acid conservation is not evenly distributed, conservation seems to improve towards the C-Terminus
	WEBLOGO	Amino acid conservation is not evenly distributed, see above
Cellular Localization	Transmembrane Helices	none
	Signal Peptide	none
	PSORTB Final Prediction	Cytoplasmic
Alternative ORFs		none
Structure Based Evidence	TIGRFAMs	No significant hits found.
	PFAMs	<i>Ala_racemase_N</i> (PF01168)
	HMM Logo Key Residues	K28, E57, ILVM115, RK123, G125, ST192, R208, G210
	PDB	D-threo-3-hydroxyaspartate dehydratase

### *Kytococcus sedentarius* 04740:

This gene product is proposed to be an enzyme, and most likely a racemase enzyme, as shown through its Pfam of *Ala racemase N* and its COG hit of COG3616. Racemase belongs to a family of isomerases, which change one isomer to another. Racemase specifically acts upon amino acids. The gene was consecutively found to be most similar to different enzymes, such as D-threo-3-hydroxyaspartate dehydratase and aldolase. The low amount of conservation of this large gene could also lead to the assertion that this gene has a specific function pertaining to *Kytococcus sedentarius*. The resultant protein was found to be cytoplasmic.

Ksed_04750		
Basic Information	DNA Coordinates	471415-471684
	DNA Length	270bp
	Amino Acid Sequence Length	89aa
Sequence Based Similarity	COGs	COG0147 Anthranilate/para-aminobenzoate synthases component I
	TCOFFEE	Poorly conserved except at end
	WEBLOGO	Poorly conserved except at end
Cellular Localization	Transmembrane Helices	none
	Signal Peptide	none
	PSORTB Final Prediction	Cytoplasmic
Alternative ORFs		none
Structure Based Evidence	TIGRFAMs	none
	PFAMs	00425 Chorismate bind None
	HMM Logo Key Residues	P 69, E 70, 89 G, K 112, E 116, P 185, E 197, R 201, R 224, E 250, E 253
	PDB	The crystal structure of Aminodeoxychorismate Synthase from <i>Stenotrophomonas maltophilia</i> 4GRH

### *Kytococcus sedentarius* 04750:

The COG hits for Ksed\_04750 suggest that it may be Anthranilate/para-aminobenzoate synthases component I. Tcoffee and WEBLOGOs are poorly conserved except at the very end of the sequence. The PSORTB final prediction says it is a cytoplasmic gene. PDB suggests that it is derived from the crystal structure of Aminodeoxychorismate Synthase from *Stenotrophomonas maltophilia* 4GRH.

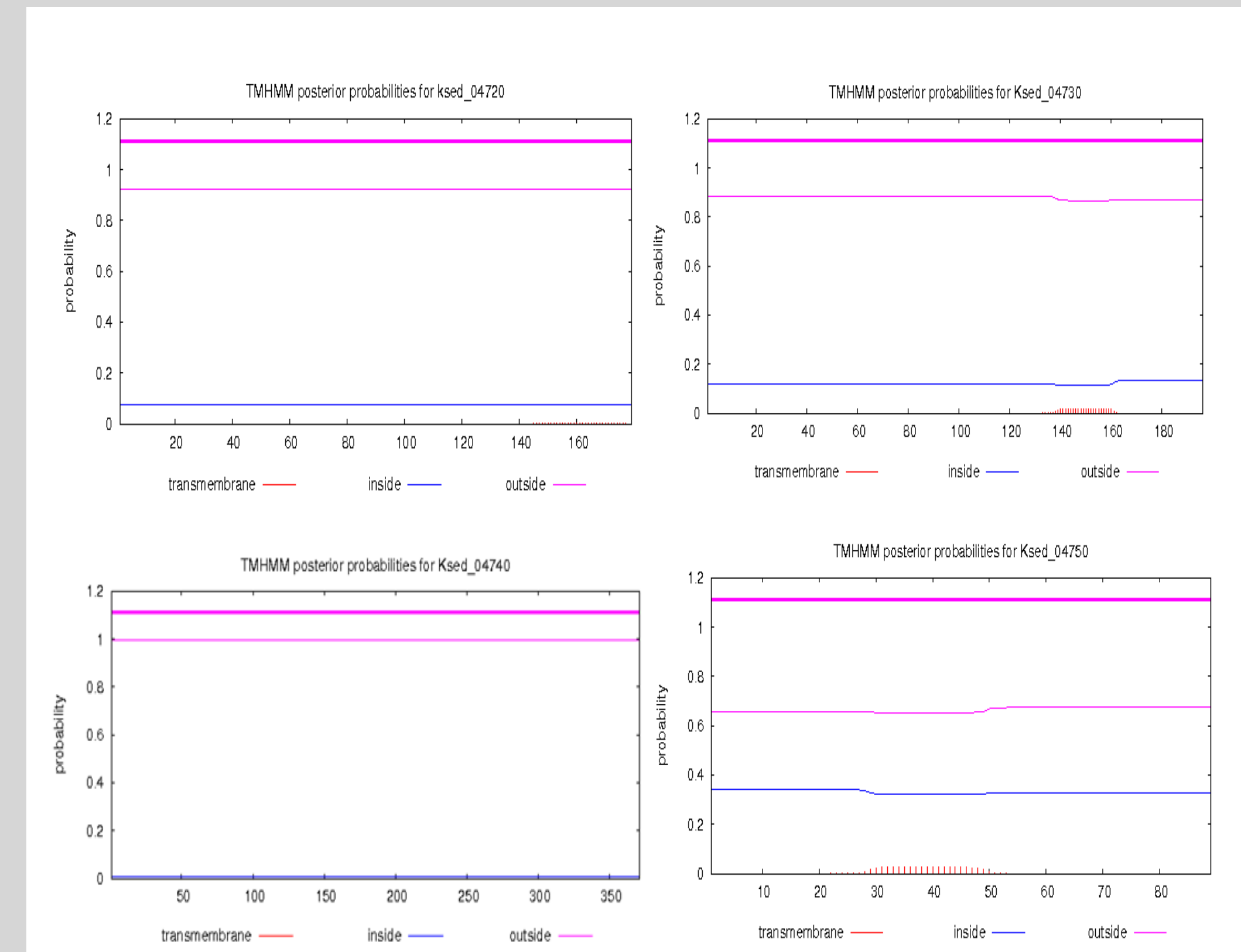


Figure 1 – Cellular Localization Data for *Kytococcus sedentarius* 04720, 04730, 04740, and 04750.

All are predicted to be cytoplasmic proteins. No significant cellular localization data was available for the gene locus 04710, which is predicted to be a hypothetical protein. Due to the predicted enzymatic functions, it is probable that these enzymatic pathways are involved in reactions within the cytoplasm of the bacterial cell.

## Conclusion

The Geni-Act proposed gene product 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 by the computer database. The exception to this is the 04710 gene locus which may be a hypothetical protein.

Gene Locus	Proposed Annotation
04710	Hypothetical Protein
04720	Serine Protease Inhibitor
04730	Uncharacterized family of potential enzymes
04740	Racemase Enzyme
04750	Anthranilate/para-aminobenzoate Synthase

## References

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

## Acknowledgments

Supported by NSF ITEST Strategies Award Number 1311902