

Annotation of the *Kytococcus sedentarius* Genome from DNA Coordinates 512848 to 515799

Macayla Barnes, Lily Hamill, Taven Romesser, Madhuri Vihani and Elizabeth Winters

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

Abstract

A group of consecutive 7 genes from the microorganism *Kytococcus sedentarius* (Ksed_05160 – Ksed_05190) 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 oligoketide 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 *Dermaoococcaceae* within the actinobacterial suborder *Micrococcineae*, 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_05160		
Basic Information	DNA Coordinates	512848..513330
	DNA Length	483 bp
	Amino Acid Sequence Length	160 aa
Sequence Based Similarity	COGs	COG1006 MnhC
	TCOFFEE	Well conserved in the beginning
	WEBLOGO	Well conserved
Cellular Localization	Transmembrane Helices	3
	Signal Peptide	none
	PSORTB Final Prediction	Cytoplasmic membrane
Alternative ORFs		none
Structure Based Evidence	TIGRFAMs	(TIGR00941)
	PFAMs	Oxidored_q2 (PF00420)
	HMM Logo Key Residues	LIMVF27, MILFV28, MILFV33
	PDB	n/a

Kytococcus sedentarius 05160:

T-Coffee and Weblogo alignments show that the protein product of Ksed_05160 is well conserved. Data supports this protein has not been changed much through time. COG results indicate this protein to be an sodium anti-porter, which is inferred to be electrogenic (produces a change in electrical potential of a cell). Cellular localization data supports an integral transmembrane protein because it has three helices, and no signal peptides.

Ksed_05170		
Basic Information	DNA Coordinates	513327..514892
	DNA Length	1566 Base Pairs
	Amino Acid Sequence Length	521 Amino Acids
Sequence Based Similarity	COGs	COG0651: HyfB
	TCOFFEE	Well conserved in the middle
	WEBLOGO	Well conserved in the middle from amino acid 185 to 447.
Cellular Localization	Transmembrane Helices	14
	Signal Peptide	No signal peptide
	PSORTB Final Prediction	Cytoplasmic Membrane
Alternative ORFs		None
Structure Based Evidence	TIGRFAMs	(TIGR00944)
	PFAMs	PF00361
	HMM Logo Key Residues	E12, K92, W101, K168, A172, S174, F212, F260
	PDB	Crystal structure of the membrane domain of respiratory complex I from E. Coli

Kytococcus sedentarius 05170:

Using the WEBLOGO research, it was found that the protein sequence of Ksed_05170 was very well conserved in the middle, specifically from amino acid number 185 through 447. This means that the gene sequence at this point is very similar to that of other species, which indicates an important function. Ksed_05170 is most likely part of an antiporter subunit. It is possible that the gene is an integral transmembrane protein, responsible for transferring Na⁺ and H⁺ ions through the cell membrane by active transport.

Ksed_05180		
Basic Information	DNA Coordinates	514889..515509
	DNA Length	621bp
	Amino Acid Sequence Length	399aa
Sequence Based Similarity	COGs	COG1863 Na ⁺ /H ⁺ ion transporter
	TCOFFEE	Amino acids are well conserved
	WEBLOGO	Amino acid conservation mainly in beginning
Cellular Localization	Transmembrane Helices	1-2 Transmembrane Helices
	Signal Peptide	none
	PSORTB Final Prediction	Cytoplasmic Membrane
Alternative ORFs		none
Structure Based Evidence	TIGRFAMs	No TIGRFAM found
	PFAMs	MnhE family (PF01899)
	HMM Logo Key Residues	WYF9, GA24, AGSC28, EDQ66, MILV67, AS70, NSH71, VIM74, PA87, LVMI96, IVML109, TSC110, P113, G114, TS115, IVLM129, HYF130, MILF132, EDQ149,
	PDB	No results found.

Kytococcus sedentarius 05180:

The sequence based similarity tests indicate a fairly well-preserved protein. Results from the cellular localization tests shows this protein to have 1-2 transmembrane helices supports an integral transmembrane protein. This gene in other organisms serves as a sodium/hydrogen ion transporter molecule.

Ksed_05190		
Basic Information	DNA Coordinates	515509..515799
	DNA Length	291 Base Pairs
	Amino Acid Sequence Length	96aa
Sequence Based Similarity	COGs	COG2212: MnhF Multisubunit Na ⁺ /H ⁺ antiporter, MnhF subunit [Inorganic ion transport and metabolism]
	TCOFFEE	Well conserved in the middle
	WEBLOGO	Mostly well conserved in the middle of the protein, little to no conservation at the C and N terminal.
Cellular Localization	Transmembrane Helices	3
	Signal Peptide	None
	PSORTB Final Prediction	Cytoplasmic Membrane
Alternative ORFs		None
Structure Based Evidence	TIGRFAMs	No hits above threshold
	PFAMs	PF01899 (Only one)
	HMM Logo Key Residues	WYF9, GA24, AS70, VIM74, PA87, TSC110, P113, G114, TS115, HYF130, MILF132, EDQ149
	PDB	No Results

Kytococcus sedentarius 05190:

COG data supports Ksed_05190 to be a Multisubunit Sodium/Hydrogen ion antiporter that uses the hydrogen ion gradient to expel sodium ions from the cell. It assists in metabolism and maintains pH within the cell. The presence of transmembrane helices supports an integral transmembrane protein. Due to the lack of signal peptide, it is suggested that the protein is not secreted from *Kytococcus*. The PSORTB final prediction proved it to be a cytoplasmic membrane protein with a score of 10. This protein is also found in eukaryotic cells in the kidneys to regulate cell volume and cell migration.

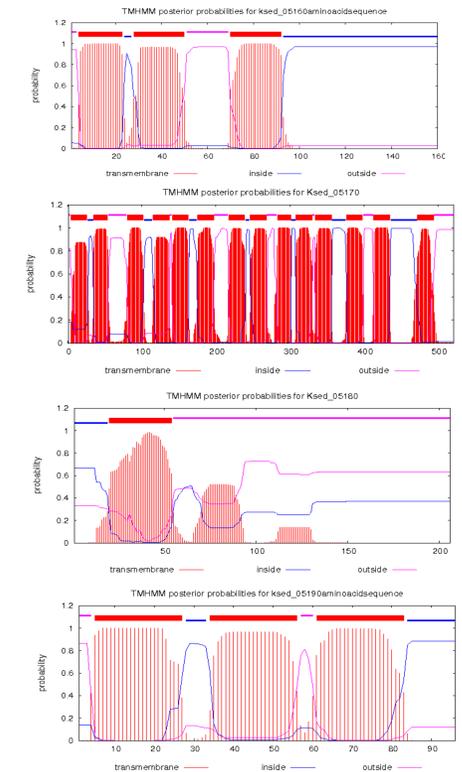


Figure 1 – Cellular Localization Data for *Kytococcus sedentarius* 05160, 05170, 05180, and 05190.

All are predicted to be transmembrane proteins. Due to the predicted functions, it is probable that these proteins are involved in reactions within the membrane as well as inside and outside 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. These results support a transmembrane protein that functions as a Na⁺/H⁺ transporter. More tests would need to be done to see if these genes are possibly part of an operon system.

Gene Locus	Proposed Annotation
05160	Sodium/Hydrogen Ion Antiporter
05170	Sodium/Hydrogen Ion Antiporter
05180	Sodium/Hydrogen Ion Antiporter
05190	Sodium/Hydrogen Ion Antiporter

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