University at Buffalo The State University of New York | REACHING OTHERS

Annotation of the *Kytococcus sedentarius* Genome for Ksed_07190- Ksed_07240

Kayla Murray*, Joseph Butler,* Benjamin Scheifla,* Jordan Hauser*, Carl Weinreber,* Brianna Morath, Lynn Barrett

Department of Biotechnical and Clinical Laboratory Sciences

West Seneca West Senior High School 3330 Seneca Street, West Seneca NY 14224 (* indicates students equally contributed to gene annotation)

Abstract

A group of consecutive 6 genes from the microorganism *Kytococcus sedentarius* (Ksed_07190 – Ksed_07240) 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 and the possibility of horizontal gene transfer. The Genbank proposed gene product name did not differ significantly from the proposed gene annotation for most of the genes in the group. Ksed_07190 is most likely a hypothetical gene. Otherwise, the genes appear to be correctly annotated by in the database.

Introduction

The *Kytococcus sedentarius* bacterium is an aerobic, free-living, non-motile bacterium. It has a High GC content of 72 %. That is gram positive and isolated from an marine environment. Methionine, along with other amino acids in the medium, are essential for growth. *Kytococcus sedentarius* is found in the *Dermacoccaceae* family. It is known for the production of oligoketide antibiotics. It is an opportunistic pathogen; causing valve endocarditis, hemorrhagic pneumonia, and pitted keratolysis. The optimal growth temperature is around 25-35 degrees celsius in a NaCl (sodium chloride) concentrations up to 10 %. It does not form spores, and its GPC in the tetrads.

The main purpose of the study was to annotate the assigned genes (Ksed_0719-07240). Through annotation, cell location and establishment of whether the genes were hypothetical or functioning, were determined. Another purpose of this project was to add further knowledge regarding *Kytococcus sedentarius* to the science community.

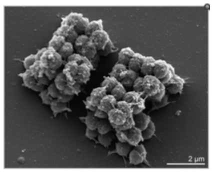
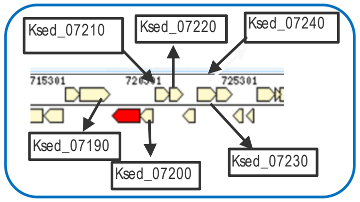


Figure 1: A scanning electron micrograph of Kytococcus sedentarius showing the characteristic cuboidal growth pattern of the individual bacterial cells (Sims et al., 2009).

Figure 2: Gene neighborhood of Ksed_07190 - Ksed_07240



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 - Structure-Based Evidence	TIGRfam, Pfam, PDB	Are there functional domains in my protein?
Module 4 - Cellular Localization Data	Gram Stain, TMHMM, SignalP, PSORTb, Phobius	Is my protein in the cytoplasm, secreted or embedded in the membrane?
Module 5 - 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 8 - Evidence for Horizontal Gene Transfer	Phylogenetic Tree	Has my gene co-evolved with other genes in the genome?
Final Annotation	Review data from all modules	Does the student proposed name of the gene agree with that proposed by the automated computer annotation? Are any changes proposed to the pipeline annotation?

Results

Kytococcus sedentarius 07190: Protein BLAST resulted in a top hit indicating a hypothetical protein; an organism known as [*Tessaracoccus sp. SIT6*]. BLAST hit #2 agreed with the predicted gene product by GeniAct; an ABC transporter permease protein. Pfam results indicated the gene was part of the Ftsx family with 4 clan members. THMM had 10 predicted transmembrane helices. PSORT-B showed a cytoplasmic membrane score of 10.0, the highest possible score. Phobius generated a graph illustrating mostly transmembrane. The absence of a signal peptide predicted in SignalP indicating it does not leave the cell.

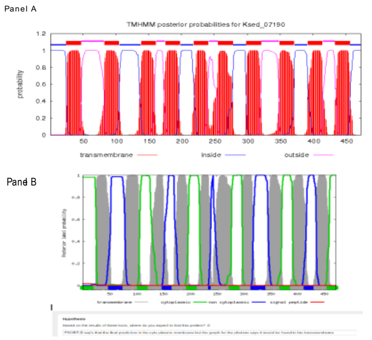
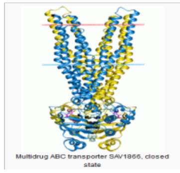


Figure 3: Panel A, showing the transmembrane domains. Panel B, showing external domains for Ksed_07190, a hypothetical gene.

Kytococcus sedentarius 07200: According to BLAST the top hit is a macrolide ABC transporter ATP-binding protein and the organism is Tessaracoccus sp. SIT6. The second top hit is macrolide ABC transporter ATP-binding protein and the organism is Propionibacterium propionicum. The WebLogo sequence logo for this gene is well conserved. The Gram stain of the microbe is Gram-Positive. PSORT-B final prediction is Cytoplasmic membrane.

Figure 4: 3-D structure of ABC transporter (ATP-binding domain of ABC transporters from Wikipedia, 2016)



***Kytococcus sedentarius* 07210:** The predicted gene product by GENIACT was histidine kinase. BLAST results from SwisPort correlate with the predicted gene. Signal transduction histidine kinase was the top hit from COG.

Kytococcus sedentarius 07220: The original gene predicted by GENIACT was a response regulator with a Che Y-like receiver domain and HTH DNA-binding domain. The Pfam prediction was a response regulator receiver domain; bacterial regulatory proteins, luxR family. The assessment of the BLAST resulted in the top gene product hit of DNA-binding response regulator. The COGs are in the same family as the original genes predicted. The WEBLOGO was well conserved indicating a match.

Figure 5: Top COG hit Ksed_07220

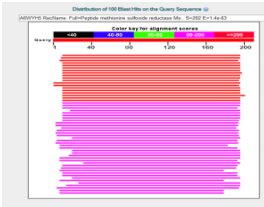

COG2771: CsgD

DNA-binding transcriptional regulator, CsgD family [Transcription]

linked to 3D-structure

Kytococcus sedentarius 07230: GENI-ACT initially proposed methionine-S-sulfoxide reductase as the gene product. BLAST results correlated with the initial gene product protein. The Web Logo of this specific gene was highly conserved; indicating a good indication of a matching protein function. Pfam, COG hits and the HMM logo all concur with the proposed gene product initially stated by GENI-ACT.

Figure 6: BLAST results for Ksed-07230 NR database



Kytococcus sedentarius 07240: Is located on the top strand of the DNA. The top BLAST hit for the amino acid sequence identified the gene as an organism called *Demetria terragena.* The gene product name is daunorubicin resistance protein DrrA family ABC transporter ATP-binding protein. Kytococcus sedentarius is a gram-positive bacteria with no transmembrane helices according to the TMHMM results. This protein is most likely Non Cytoplasmic. Coordinates 723990..725008 have a Shine-Dalgarno region in front of them, this shows a high probability of it having a start codon unlike the coordinates proposed by IMG/EDU. The gene Ortholog Neighborhood Region of Ksed_07240 using IMG/ER looks similar to the same neighborhood in closely related bacteria with no evidence for horizontal gene transfer. It had 72% GC - similar distribution with no indication of horizontal gene transfer.

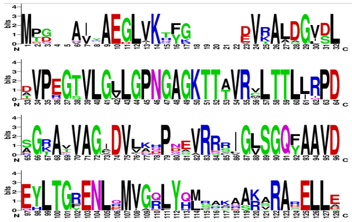


Figure 7: Kytococcus sedentarius 07240 WebLogo illustrating highly conserved amino acid sequences. Daunorubicin resistance protein DrrA family ABC transporter ATP-binding protein.

Conclusion

The GENI-ACT proposed gene product did not differ significantly from the proposed gene annotation for most of the genes in the group and as such, the genes appear to be correctly annotated by the computer database. Ksed_07190 is the exception and may be a hypothetical gene.

Gene Locus	Geni-Act Products	Proposed Annotation
Ksed _7190	Predicted permease	hypothetical protein [Tessaracoccus sp. SIT6]
Ksed_7200	ABC-type antimicrobial peptide transport system,ATPase component	macrolide ABC transporter ATP-binding protein
Ksed_7210	histidine kinase	histidine kinase. *Limited data researched.
Ksed_7220	response regulator containing a CheY-likereceiver domain and an HTH DNA-binding domain	DNA-binding transcriptional regulator, CsgD family
Ksed_7230	methionine-S-sulfoxide reductase	Peptide methionine sulfoxide reductase MsrA
Ksed_7240	daunorubicin resistance ABC transporter ATP-binding subunit	daunorubicin resistance ABC transporterATP-binding subunit

References

ATP-binding domain of ABC transporters from Wikipedia, 2016

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

Acknowledgments

Dr. Stephen Koury and Dr. Rama Dey-Rao

Supported by NSF ITEST Strategies Award Number 1311902