

# Annotation of the *Kytococcus sedentarius* Genome from Locus Tags Ksed\_07250 to Ksed\_07310

Patrick Harrington\*, Andrew Jerabek\*, Kiandra Kormos\*, Isabella Mancino\*, Michael Janak\* and Gregory Jubulis

West Seneca West Senior High School, 3330 Seneca Street, West Seneca, NY 14224 and the Western New York Genetics in Research Partnership (\* equally contributed to gene annotation)

## Abstract

The group of genes in the locus tag interval Ksed\_07250 to Ksed\_07310 from the bacterium *Kytococcus sedentarius* were annotated using GENI-ACT, a compilation of multiple genomic data bases that use FASTA sequences to deduce a gene's final protein function and name. The proposed gene name was assessed using the following GENI-ACT modules: Basic Information, Sequence Based Similarity, Structure Based Evidence, Cellular Localization Data, Evidence for Horizontal Gene Transfer, and Alternative Open Reading Frame. The final annotation of the gene deduced by the aforementioned modules was consistent with the proposed gene annotation made by GENI-ACT. The genes were called correctly.

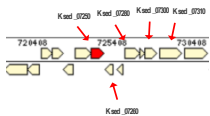
## Introduction

*Kytococcus sedentarius* is a marine dwelling, gram positive organism, that is free-living and non-motile, strictly aerobic, and can only grow when multiple amino acids are provided in a growth medium of an experimental locale. As for its structure, this gram positive strain can mostly be found in tetrads, irregular clusters, and cubical packets of eight.

The bacterium naturally produces oligopeptide antibiotics which suggests a potential application in biotechnology. The antibiotics the bacterium yields are to ward off competition and other pathogens; however, *Kytococcus sedentarius* is also an opportunistic pathogen causing numerous infections. Valve endocarditis is an infection in the lining of the heart and is dangerous when it reaches the valves because the lack of blood flow hinders an immune response. Another infection which is commonly caused by this bacterium is hemorrhagic pneumonia, which is marked by a pulmonary hemorrhage that leads to acute respiratory failure; it is mostly fatal. Pitted keratolysis is the most commonly caused disease by *Kytococcus sedentarius*. This is marked by pungent smelling feet and white clusters of punched-out holes in the heel or sole of the foot. Several clinical studies of the bacterium have been recorded, and they prove that the bacterium releases two extracellular enzymes that independently degrade natural, insoluble human callus. (Sims et al 2009)

Genes of *Kytococcus sedentarius* can be studied using bioinformatics. Specific locus tags of the bacterium's DNA are analyzed to determine their function, shape, location in the cell, and their genetic background which characterizes the bacterium. The primary component of the study is the compiling of the data found in multiple databases and comparing them to each other in order to surmise the true intention of a particular gene. Each student was assigned a unique locus tag in the range Ksed\_07250-07310.

Figure I. Gene Neighborhood of Ksed\_07250-07310



## 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, PSORT, 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?

Table 1. Modules employed for gene annotations of Ksed\_07250-Ksed\_07310

## Results

***Kytococcus sedentarius*; Ksed 07250:** Based on the GENI-ACT prediction this protein is a ABC transporter efflux protein. The web logo showed that this protein has high conservation in many locations. This is a gram positive microbe with 6 trans membrane helices. This means that the GENI-ACT prediction is likely correct. This protein is used to transfer material in and out of the cell. It has a Cytoplasmic Membrane score of 10.00.

***Kytococcus sedentarius*; Ksed 07260:** The initial proposed product of this gene by GENI-ACT was a transcription elongation factor, GreA/GreB family [transcription]. The web logo of this specific gene shows multiple contingencies where high conservation is observed between organisms and may show similar function throughout. The protein made by the transcription of this gene is not exported out of the cell, is not on the cellular membrane, has no evidence of transmembrane helices, the protein functions in the cytoplasm. The gene that encodes the explicit protein is seventy-eight percent Cytosine and Guanine, and does not indicate that any horizontal transfer occurred.

***Kytococcus sedentarius*; Ksed 07280:** BLAST illustrated that the top hit of this gene was mycobacterial conjugate amidase. Web logo showed that there were high levels of conservation in this sequence throughout many locations. TMHMM indicates that this protein is not in the cell membrane. This statement was further supported by SignalP, which stated that this protein is not exported out of the cell. Alternative open reading frame is the section in the notebook which confirmed that this sequence codes for a real gene. Finally, Gene Context and the Chromosome Viewer GC Heat Map displayed similar neighborhoods and similar distribution of GC, which means there is no evidence for horizontal gene transfer.

***Kytococcus sedentarius*; Ksed 07300:** Comparison of the amino acid sequence to the non-redundant protein database using protein BLAST resulted in the gene isoprenyl transferase as the top hit. The top hit has an E-value of 2e-119 which suggests that the compared protein sequences are very similar. In addition, the top Pfam-A match was the protein family Prenyltransferase which was identified as a san undecaprenyl diphosphate synthase, meaning this is a collection of related proteins containing the same domain as the search sequence.

***Kytococcus sedentarius*; Ksed 07310:** Based on the information provided by the SignalP and P-Sort B databases, there is no evidence that this gene is located in the cell membrane or transported out of the cell, therefore this gene most likely functions in the cytoplasm. Comparison of the amino acid sequence using non-redundant protein database produced a top hit of Janibacter sp. HTCC2649. This hit has an E-value of 0.0 which suggests that the compared amino acid sequences are very similar, if not exactly the same.

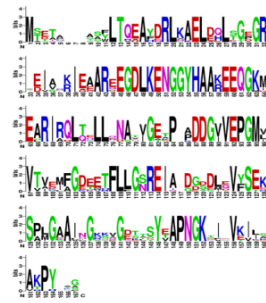


Figure II. Ksed\_07260 Web logo. The multiple large letters represent the multiple contingencies where the gene has been conserved across numerous species and may show similar function.

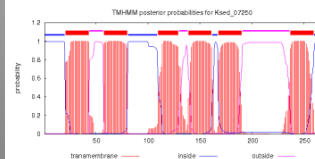


Figure III. TMHMM output predicting the presence of 6 transmembrane helices in Ksed\_07250.

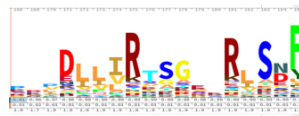


Figure IV. Ksed\_07300 HMM Logo showing a highly conserved region of the protein suggesting a functional role.

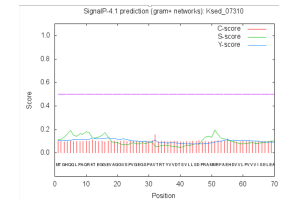


Figure V. Ksed\_07310 SignalP graph indicating that this gene does not code for a signal peptide and therefore will most likely be found in the cytoplasm.



Figure VI. Gene neighborhood of Ksed\_07280 looks similar in closely related bacteria suggesting there is no evidence for horizontal gene factor.

## 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.

GENE LOCUS	GENI-ACT PRODUCTS	PROPOSED ANNOTATION
07250	ABC Transporter Efflux Protein	ABC Transporter Efflux Protein
07260	Transcription Elongation Factor	Transcription Elongation Factor
07280	Mycobacterial Conjugate Amidase	Mycobacterial Conjugate Amidase
07300	Undecaprenyl Pyrophosphate Synthetase	Undecaprenyl Pyrophosphate Synthetase
07310	PhoH Family Protein	PhoH Family Protein

## References

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

## Acknowledgments

Special thank you to SUNY at Buffalo, Dr. Stephen Koury, and Dr. Rama Dey-Rao for the opportunity and their assistance with this project. Supported by NSF ITES Strategies Award Number 1311902.