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

Annotation of the *Kytococcus sedentarius* Genome from DNA Coordinates 821129 to 824062

Sherrie Wang*, Pakhi Tripathi*, Marilena Papavassiliou*, Stephanie Terhaar, Leena Sen, Mrs. Kathleen Kramer

Williamsville East High School and the Western New York Genetics in Research Partnership

Abstract

The annotation of 5 consecutive genes from the microorganism Kytococcus sedentarius (Ksed_08070 - Ksed_08110) was completed with the use of the 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, and the possibility of horizontal gene transfer. The annotation of *Kytococcus sedentarius* by computer software was verified through manual annotation with results recorded in the GENI-ACT lab notebook.

Introduction

Kytococcus sedentarius is a gram positive, strictly aerobic bacterium that resides in various environments such as human skin and marine habitats. The bacterium resides in temperatures of 25-37 °C and is a chemoorganotrophic organism, requiring methionine and other amino acids to grow. It thrives in sodium chloride concentrations less than 10% (w/v), and it is a spherical coccoid that occurs predominantly in tetrads. In addition, the organism is non-motile, non-encapsulated, and non-endospore forming.

Kytococcus sedentarius was studied due to the fact that it has several advantageous uses in biotechnology. It is known for the production of polyketide antibiotics (monesin A and B) as well as its role as an opportunistic pathogen. It may be harmful to humans because it can cause pitted keratolysis, valve endocarditis, and hemorrhagic pneumonia.

Ksed_08070 is a transcriptional regulator of the ArsR family.
Ksed_08080 is a arabinose efflux permease family protein.
Ksed_08090 is a transposase with a molecular function of nucleic acid binding and serves in the biological process of DNA integration.
Ksed_08100 is a transposase that serves in DNA binding and DNA mediating.
Ksed_08110 is a transcriptional regulator of the TetR family.

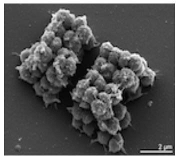

Figure I - scanning electron micrograph of *Kytococcus sedentarius* (Manfred Rohde, Helmholtz Centre for Infection Biology, Braunschweig).

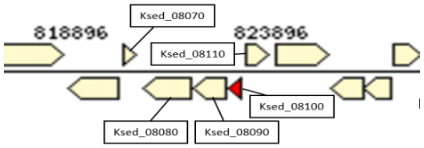

Figure II - Gene neighborhood from IMG/EDU. The five consecutive genes that were studied are labeled.

Methods and Materials

Modules of 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 6- Enzymatic Function	KEGG, MetaCyc, E.C. Number,	In what process does my protein take part?
Module 7- Gene Duplication/Gene Degradation	Paralog, Pseudogene	Are there other forms of my gene in the bacterium? Is my gene functional?
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 08070:

The initial proposed product of this gene by GENI-ACT was a transcriptional regulator of the ArsR family. This gene product proposal was supported by the top BLAST hits and the cytoplasmic location of the product as predicted by PSORT-B. The amino acid sequence of this gene tends to be well-conserved across species. No signal peptide was predicted and the probability of any transmembrane regions was low. The start and stop codons were correct as called, and there is a Shine-Dalgarno sequence 5-15 base pairs upstream of the start codon. The domain of the gene product according to the database Pfam is HTH 20, and the clan is the helix-turn-helix clan with the clan number being CL0123. No active site was predicted for this gene product; it is most likely not an enzyme. There is no evidence that this gene is a pseudogene. There is some evidence for horizontal gene transfer.

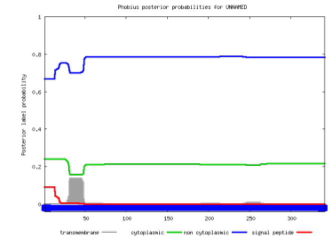

Figure III: Phobius results for Ksed_08070 demonstrating the probabilities for the location of the protein. Although there appears to be a high probability for the location to be non-cytoplasmic, PSORT-B reported a cytoplasmic score of 7.5, predicting that the protein is located in the cytoplasm.

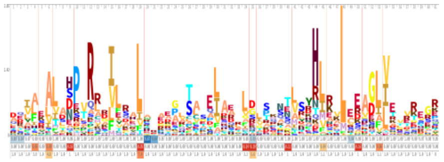

Figure IV – Ksed_08070 HMM LOGO results show that there were key functional residues at L19, H44, and L48 within the amino acid sequence.

Kytococcus sedentarius 08080:

The initial proposed product of this gene by GENI-ACT was a arabinose efflux permease family protein. The top BLAST hit for the amino acid sequence was arabinose ABC transporter permease. Arabinose efflux permease is an integral component of the cell membrane and aids in transport across the membrane. Through the analysis of T-Coffee and WebLogo, it can be concluded that the protein was very conservative throughout the species, indicating its important functionality. The domain of the gene product according to the database Pfam is PF06779, and the clan is the Major Facilitator Superfamily clan with the clan number being CL0015. There is no evidence to conclude that this is a pseudogene.

Kytococcus sedentarius 08090:

The initial proposed product of these genes by GENI-ACT was a transposase. BLAST provided evidence to support this proposed analysis, and showed top hit gene products of integrase present in the organisms *Propionibacterium freudenreichii* and *Micrococcus luteus*. The gene is well conserved throughout the species as seen with the wide, large letters present in the Web Logo. According to the web logo, many hydrophobic regions were also present, indicating the possibility of lipid makeup. The results of PSORT-B determine the likely location of the protein to be cytoplasmic. Pfam shows the protein belong to the helix-turn-helix clan (CL0123). By analyzing the results of PDB, the Cyro-EM reveals a novel octameric integrase structure for betaretroviral intasome function. There are no unique ligands, a largely helical structure is formed, and the macromolecule content is integrase protein. No evidence for the gene being a pseudogene is provided. Ksed 08090 remains consistent with the initial proposed product of transposase.

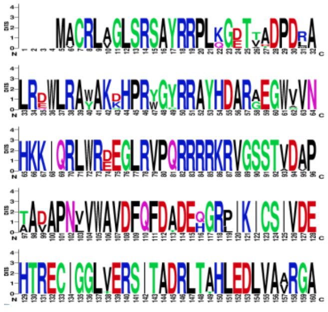

Figure V: *Kytococcus sedentarius* 08090 Web Logo revealing a large amount of conservation throughout species and a sufficient amount of hydrophobic regions.

Kytococcus sedentarius 08100

The initial proposed product of these genes by GENI-ACT was a transposase. This gene product proposal was supported by BLAST as there were significant matches with similar proteins in *Microbacterium laevaniformans* and *Micrococcus luteus*. T-Coffee and WebLogo analysis revealed that the protein was well conserved, indicating important functionality. Results of PFAM show that the protein comes from a family with similar structures and functions. TMHMM, SignalP, and Phobius do not show a signal peptide nor any transmembrane helices. The cellular localization of the protein is unclear and needs further investigation.

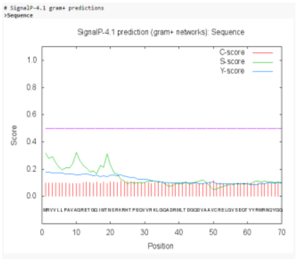

Figure VI - SignalP graph for Kytococcus sedentarius_08100 showing the C-score, S-score, and Y-score. The results indicate that there is no signal peptide and thereby no cleavage sites.

Kytococcus sedentarius 08110:

The initial proposed product of this gene by GENI-ACT was a transcriptional regulator. This gene product was supported by the top BLAST hit, which was also a transcriptional regulator protein. PFAM analysis showed the gene produces bacterial regulatory proteins of the TetR family. Moreover, the amino acid sequence is well conserved across species according to the T-Coffee and WebLogo analyses. Both LIPO-P and PSORTB analyses suggest that the protein has cytoplasmic placement over a noncytoplasmic placement. TMHMM, Signal P, and Phobius showed no signal peptide or transmembrane helices. 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 Gene Products	Proposed Annotation
08070	Transcriptional Regulator	Transcriptional Regulator
08080	Arabinose Efflux Permease	Arabinose Efflux Permease
08090	Transposase	Transposase
08100	Transposase	Transposase
08110	Transcriptional Regulator	Transcriptional Regulator

References

Sims et al. (2009). Complete genome sequence of Kytococcus sedentarius type strain (541T). *Standards in Genomic Sciences*, 12 - 20.
"Kytococcus Sedentarius DSM 20547." *BacMap*. Web. 09 May 2016.
"JGI IMG Home." *JGI IMG Home*. Web. 09 May 2016.
"Western New York Genetics in Research Partnership." *Western New York Genetics in Research Partnership*. Web. 09 May 2016.
"Output Format." *SignalP 4.1*. Web. 09 May 2016.

Acknowledgments

Supported by NSF ITEST Strategies Award Number 1311902
Dr. Rama Dey-Rao Clinical Assistant Professor, University at Buffalo Dr. Stephen Koury Research Associate Professor, University at Buffalo