

Annotation of the *Kytococcus sedentarius* Genome from Locus Tag Ksed_06010 to Ksed_06030

Samantha Maley*, Lydia Petricca*, Emily Zhang* and Stacy Rominger* (*indicates equal contribution)

Orchard Park High School, 4040 Baker Road, Orchard Park, NY 14127 and the Western New York Genetics in Research Partnership

Abstract

A group of consecutive 3 genes from the microorganism *Kytococcus sedentarius* (Ksed_06010 – Ksed_06030) were annotated using the collaborative genome annotation website GEN-ACT. Through a series of progressive modules performed on Genbank data regarding 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, the possibility of horizontal gene transfer, and the production of an RNA product was obtained for all three genes. For Ksed_06010 and Ksed_06020, the proposed gene product name matched the proposed gene annotation for each of the genes in the group. For Ksed_06030, the initial proposed gene product name was a hypothetical protein, and annotation findings reaffirmed this.

Introduction

The advent of Bioinformatics technology in recent years has provided a channel and mechanism to analyze amino acid and protein sequence so far specified genes. The significance of research projects designed around this focus is that it provides information regarding navigation of gene information including the location, role, and function of various loci of sequences. In the case of *Kytococcus sedentarius*, obtaining this data can progress knowledge about the locus in its entirety and its role as an opportunistic pathogen.

Previous research about *Kytococcus sedentarius* has revealed that it is pertinent in the production of oligo ketide antibiotics as an opportunistic pathogen, precipitating the onset of endocarditis, hemorrhagic pneumonia, and pitted keratolysis. Moreover, *Kytococcus* undergoes cellular respiration anaerobically and only has the capacity to develop when a medium of several amino acids is provided. Studies have shown that it is oxidase positive, emphasizing the presence of oxygen as a terminal electron acceptor in respiration, and catalase positive, affirming the presence of an enzyme that is produced by bacteria respiring using oxygen in order to provide defense against H₂O₂. However, there is still a lack of understanding about *Dermaoocaceae*, the location and family of *Kytococcus*, it is imperative to facilitate further research, such as the studies performed in these modules.

This study of *Kytococcus sedentarius* has provided information about the role, function, and location of three locus tags of the genome (Ksed_06010, Ksed_06020, and Ksed_06030). Guided by a series of eight objective-based modules, all three participants reaffirmed that the initial proposed product of the gene offered by GEN-ACT was in fact the product supported by these module findings.



Methods and Materials

Modules of the GEN-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?
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?

Results

*Kytococcus sedentarius*06010:

The initial proposed product of this gene by GEN-ACT was threonine dehydrogenase-like Zn-dependent dehydrogenase. An alcohol dehydrogenase is responsible for metabolizing the ethanol consumed and their activities contribute to the rate of ethanol elimination from the blood. Using the amino acid sequence, I ran a search through BLAST and received two significant results. Both results did confirm the proposed gene product by GEN-ACT of an alcohol dehydrogenase. I found that my protein likely resides in the cytoplasm of *Kytococcus*, based on TMHMM, PSORTb, and SignalP. TIGRFAM again confirmed the proposed gene product of a dehydrogenase. My results from TIGRFAM gave me two significant hits, both saying it is a dehydrogenase, removing a toxic metabolite. PFAM gave me a more precise result, saying it was a Zn-c-binding dehydrogenase. After gathering all of this information, the initial proposed gene product is confirmed. The proposed annotation is an alcohol dehydrogenase.

*Kytococcus sedentarius*06020:

The initial proposed product of this gene by GEN-ACT was arginine deiminase, an enzyme that plays a significant role in energy metabolism (MetaCyc). This gene product proposal was supported by the evidence yielded from NCBI blasts and CDD results. Both of my top blast hits and the COG name were arginine deiminase. In addition, the top TIGRFAM hit was *arcA*: arginine deiminase, and the top Pfam hit was Amidotransferase, which includes arginine deiminase. All of these findings supported the computer annotation of Ksed_06020 as arginine deiminase.

*Kytococcus sedentarius*06030:

The initial proposed product of this gene by GEN-ACT was a hypothetical protein. This gene product proposal was supported by the evidence yielded from NCBI blasts (all proteins are hypothetical with high E values), thus solidifying the conclusion that this gene is common with other genomes, but there is not enough information to establish the gene's name and function. Further investigation in wet lab, such as placing this gene in bacterium and observing the results, is necessary to determine the role of this gene, beyond a hypothetical basis.

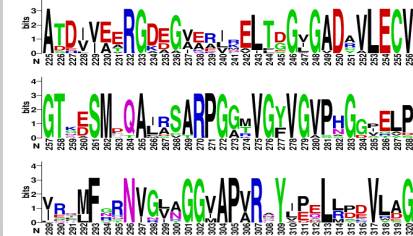


Figure I- Throughout the amino acid sequence for Ksed_06010, there are many sites that have only one letter meaning they all have the same amino acid. Majority of amino acids in common are represented as green meaning they are polar, or black meaning they are hydrophobic.

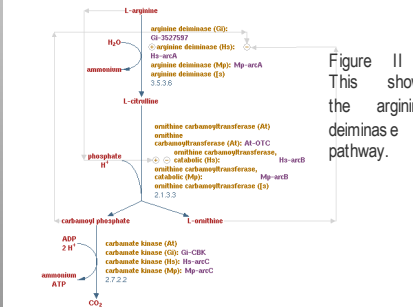


Figure II – This shows the arginine deiminase pathway.

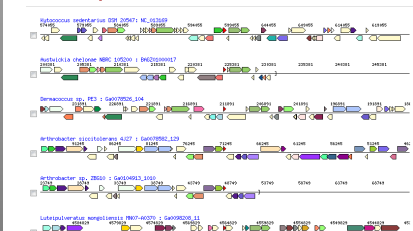


Figure III - Ortholog Neighborhood Region of organisms proposed to have a common ancestor with *Kytococcus*. Because the image above shows no discernible similarity in gene neighborhood clustering, it can be concluded that the relative evidence in support of horizontal gene transfer. Only small sections of the neighborhood are conserved.

Conclusion

For each locus, the accuracy of the GEN-ACT proposed gene product differed. In regards to genes Ksed_06020 and Ksed_06030, the proposed gene annotation was supported by the various tests performed. Therefore, these genes appear to be correctly annotated by the computer database. However, the GEN-ACT proposed gene product for Ksed_06010 differed significantly from what the tests suggested.

Gene Locus	Geni-Act Gene Products	Proposed Annotation
06010	Threonine Dehydrogenase	Alcohol Dehydrogenase
06020	Arginine Deiminase	Arginine Deiminase
06030	Hypothetical Protein	Hypothetical Protein

References

"Arginine Deiminase Pathway." *MetaCyc*. N.p., 2009. Web. 12 May 2016.
 "Catalase Test - Virtual Interactive Bacteriology Laboratory." *Catalase Test*. Michigan State University Board of Trustees, 2010. Web. 12 May 2016.
 DW, Crabb, Chang D, and Matsumoto M. "Result Filters." *National Center for Biotechnology Information*. U.S. National Library of Medicine, 2004. Web. 12 May 2016.
 "Oxidase Test - Virtual Interactive Bacteriology Laboratory." *Oxidase Test - Virtual Interactive Bacteriology Laboratory*. Michigan State University Board of Trustees, 2010. Web. 12 May 2016.
 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, Stephen Koury, Ph.D., Rama Dey-Rao Ph. D., Patricia Masso-Welch Ph. D., Danise Wilson, Mr. Jonathan Wolf