

Annotation of the *Kytococcus sedentarius* Genome from Locus Tag Ksed_07330 to Ksed_07370

Gabrielle Gelinas*, Thomas Griffin*, Ryan Supples*, Nina Vadala*, Jacob Eisensmith*, and Jason Rodemeyer

West Seneca West Senior High School 3330 Seneca Street, West Seneca, NY 14224 and the Western New York Genetics in Research Partnership *equal contribution by gene annotation authors

Abstract

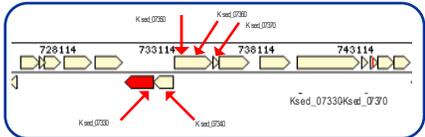
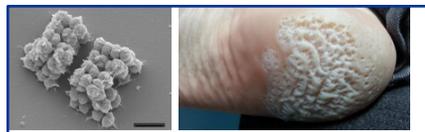
A group of five consecutive genes from the microorganism *Kytococcus sedentarius* (Ksed_07330 – Ksed_07370) were annotated using the collaborative genome annotation website GENI-ACT. The Genbank proposed gene product names for each gene was assessed in terms of the general genomic information, amino acid sequence-based similarity data, structure-based evidence from the amino acid sequence, cellular localization data, potential, or absence of gene duplication and degradation, and enzymatic function. The Genbank proposed gene product names did not differ significantly from the proposed gene annotations for four of the five genes in the group, which appear to be correctly annotated by the database. The fifth gene, Ksed_07370, had conflicting results.

Introduction

First discovered in 1944, *Kytococcus sedentarius* is a gram positive, marine dwelling organism. As an opportunistic pathogen, this bacterium can easily infect a wounded or weakened host. It lays dormant, not displaying pathogenic qualities until a viable host is discovered. *K. sedentarius* cannot grow unless a multitude of pre-existing amino acids, such as methionine, are present in the environment and the temperature ranges between 28 and 36 degrees Celsius. *K. sedentarius* can produce two different extracellular enzymes that can breakdown natural, insoluble human callus.

Strains of this species are frequently found residing in human skin. It is non-motile, non-encapsulated, and non-endospore forming. *K. sedentarius* cells are spherical or cooidal in nature and are typically arranged in tetrads, which, in turn, can be arranged similar to that of a cube. Its genome is 2,785,024 bp long with its central, circular chromosome consisting of 71.6% GC content.

K. sedentarius primarily causes pitted keratolysis, a skin infection of the feet caused by wearing tight, restricting footwear and sweating profusely. It appears as crater-like crevasses in the feet and toes, primarily found in weight-bearing areas. People of any age, race, or gender can suffer from the disease though it is found more commonly in men. People who sweat and/or wash at an excessive rate are most at risk of contracting the disease.



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:

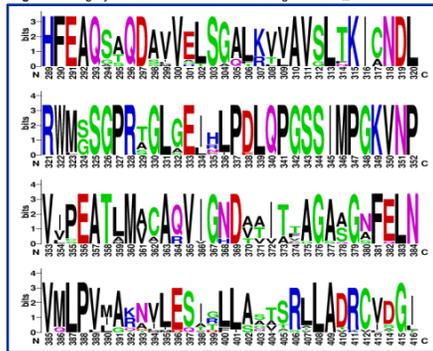
Modules	Activities	Questions Investigated
Module 1 - Basic Information Module	DNA Coordinates and Sequence, Protein Sequence	What is the sequence of my gene and protein? Where is it located in the genome?
Module 2 - Sequence-Based Similarity Data	Blast, CDD, T-Coffee, WebLogo	Is my sequence similar to 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, SignalIP, PSORTb, Phobius	Is my protein in the cytoplasm, secreted or embedded in the membrane?
Module 5 - Alternative Open Reading Frame	JMG Sequence Viewer for Alternate ORF Search	Has the amino acid sequence of my protein been called correctly by the computer?
Module 6 - 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 07330:

The initial proposed product of this gene by GENI-ACT was a fumarase. This gene product proposal was supported by accrued results performed using sequence-based similarity data and personal research. WebLogo for the amino acid sequence showed the significant presence of highly-conserved areas of amino acids in common with known cytosolic fumarase. This indicates that Ksed_07337 maybe cytosolic in nature and play an integral role in the citric acid metabolic pathway.

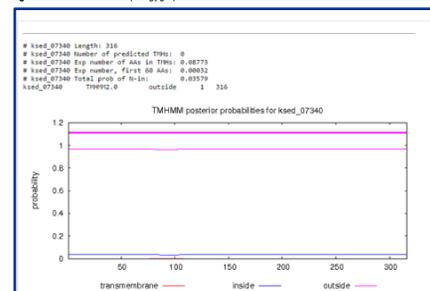
Figure 3: Highly conserved area in a WebLogo of Ksed_07330



Kytococcus sedentarius 07340:

Initial proposed results for Ksed_07340 is pantothenate kinase, which was determined by the Genebank database is cytoplasmic in nature. This gene does not possess transmembrane helices as determined by use of TMHMM and associated transmembrane topology graph. The data found would indicate a location inside the cytoplasm.

Figure 4: Transmembrane topology graph



Kytococcus sedentarius 07350:

No data acquired beyond initial BLAST results suggesting glutamine--fructose-6-phosphate transaminase, which is also suggested as the suspected protein by initial Genebank results.

Kytococcus sedentarius 07360:

The initial proposed product of these genes by GENI-ACT was phosphopantetheinyl-protein transferase. This gene product proposal was supported by the top BLAST hits for the amino acid sequences, the presence of well-curated functional domains within the amino acid sequences, the cellular location of the amino acid sequences, and the enzymatic function of the amino acid sequences. The crystalline structure of this protein in direct correlation with the proposed product and family of phosphopantetheinyl transferases.

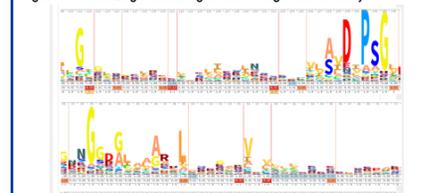
Figure 5: Crystal structure of the 4'-phosphopantetheinyl transferase slip-coenzyme A complex



Kytococcus sedentarius 07370:

Top BLAST hits, as well as results from T-Coffee and PD B suggest that this gene is responsible for producing repair proteins such as NAD(P)H-hydrate repair enzymes, which are most likely located in the cytoplasm. Phobius and Signal IP results are conflicting and suggest further research is needed to pinpoint the location as transmembrane or cytoplasmic. This enzyme allows for the repair of both epimers of NAD(P)HX, a damaged form of NAD(P)H that is a result of enzymatic or heat-dependent hydration. The yjeF-like protein selected by the initial Genebank results are similar in nature and cytosolic, but each with different roles in the metabolic pathway. Collected data indicates a need for more research. The HMM Logo below suggests the protein is an NAD(P)H-hydrate repair enzyme.

Figure 6: HMM Logo indicating areas of high conservancy



Conclusion

The products proposed by GENI-ACT did not differ significantly from the proposed gene annotation for four of the genes in the group and as such, those genes appear to be correctly annotated by the computer database. One gene, Ksed_07370, did show a slight difference from the proposed gene annotation through use of information extracted during the GENI-ACT research project.

Gene Locus	Geni-Act Products	Proposed Annotation
Ksed_07330	fumarate	fumarase
Ksed_07340	pantothenate kinase	pantothenate kinase
Ksed_07350	glutamine--fructose-6-phosphate aminotransferase	glutamine--fructose-6-phosphate transaminase
Ksed_07360	phosphopantetheinyl transferase	phosphopantetheinyl-protein transferase
Ksed_07370	bifunctional NAD(P)H-hydrate repair enzyme	yjeF-like protein, hydroxyethylthiazolekinase-related

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.

National Science Foundation ITES Strategies Award Number 1311902