

Annotation of the *Kytococcus sedentarius* Genome From DNA Coordinates 576728 to 580387

Micah Narraway, Katharina Heitzman, Trinity Clark and Kirk Winters

Pioneer Central High School and The Western New York Genetics in Research Partnership

Abstract

A group of 3 genes from the microorganism *Kytococcus sedentarius* (*Ksed_05760*, *Ksed_05770* and *Ksed_05780*) 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, and potential alternative open reading frames. The Genbank proposed gene product name did not differ significantly from the proposed gene annotation (Based on Modules 1-5) for some of the genes in the group and as such, most genes appear to be correctly annotated by the computer database.

Introduction

Kytococcus sedentarius is a strictly aerobic, non-motile, non-encapsulated, and non-endospore forming gram positive coccoid bacterium, found predominantly in tetrad formation. This organism is classified as a chemoheterotroph, as it requires methionine and several other amino acids for growth. Originally isolated from a playground slide submerged in sea water in 1944, *Kytococcus sedentarius* grows well in sodium chloride at concentrations less than 10% (w/v) (Sims et al., 2009).

According to Sims et al. (2009), *Kytococcus sedentarius* is a microorganism of interest for several reasons. This bacterium is a natural source of the oligopeptide antibiotics monensin A and monensin B (Sims et al., 2009). *Kytococcus sedentarius* has been implemented as the etiological agent of a number of opportunistic infections including valve endocarditis, hemorrhagic pneumonia, and pitted keratolysis (Sims et al., 2009). Finally, the phylogeny of this microorganism is a source of interest, as it is a member of the family *Dermacoccaceae* within the actinobacterial suborder *Micrococccineae*, which has yet to have been thoroughly studied utilizing bioinformatics (Sims et al., 2009)

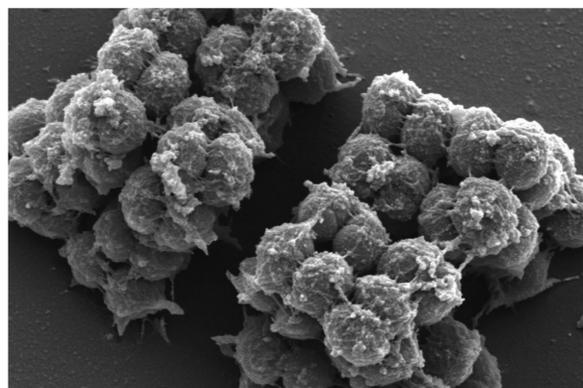


Figure I. A scanning electron micrograph of *Kytococcus sedentarius* showing the characteristic cuboidal growth pattern of the individual bacterial cells (Sims et al., 2009)

Methods and Materials

Modules of the GENI-ACT (<http://www.geni-act.org/>) were used to complete *Kytococcus sedentarius* genome annotation (For this project only modules 1-5 were used). 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?

Results

Ksed_05560

The initial proposed product of this gene by GENI-ACT was a DNA/ RNA helicase, superfamily 1. This gene product proposal was not supported by the top BLAST hits for the amino acid sequence, but well-curated protein functional domains within the amino acid sequence were present. There were no transmembrane helices predicted by TMHMM or Phobius, and the cellular location of the amino acid sequence was predicted to be cytoplasmic. Interestingly, evidence was found that further research into the start codon call for this gene should be further investigated. The sequence viewer for alternative open reading frame search in IMG/edu did not highlight the start codon used in Genbank and several upstream alternative start codons were observed (see figure VI). The annotation was only completed through modules 1-5.

Ksed_05770:

The initial proposed product of this gene by GENI-ACT was acetyltransferase or ribosomal protein N-acetylase. This gene product proposal was supported by the top BLAST hits for the amino acid sequence, the TMHMM scan (no transmembrane helices predicted), the LipoP prediction (cytoplasmic), and the PsortB prediction (cytoplasmic). The annotation was completed from modules 1-5.

Ksed_05780:

The initial proposed product of this gene by GENI-ACT was an NAD-dependent protein deacetylase, SIR2 family. This gene product proposal was supported by the top BLAST hits for the amino acid sequences, as well as PFMA and COG hits matching SIR2.

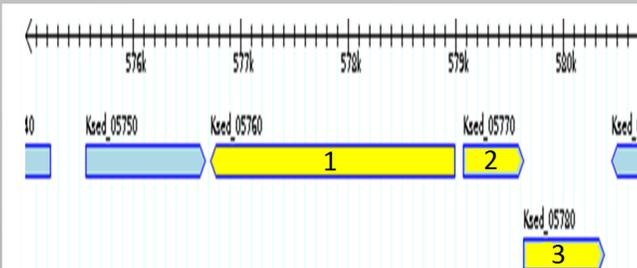


Figure II – *Ksed_05760* (1), *Ksed_05770* (2) and *Ksed_05780* (3) gene neighborhood delineating relationship of the three genes under investigation.

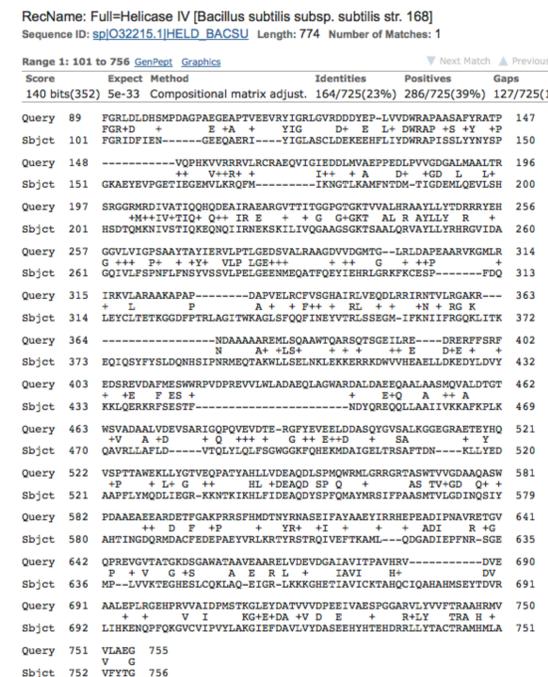


Figure III – *Ksed_05760* top BLAST search result. Notice how it still has helicase, part of its predicted gene product. This is a close relation.

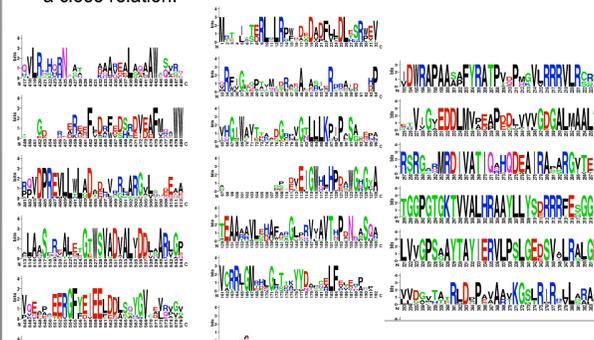


Figure IV – *Ksed_05760*, *Ksed_05770*, and *Ksed_05780* respectively. These are the WebLogo sequences that indicate how well conserved the gene is. The large amount of letters shown here indicate good gene conservation.

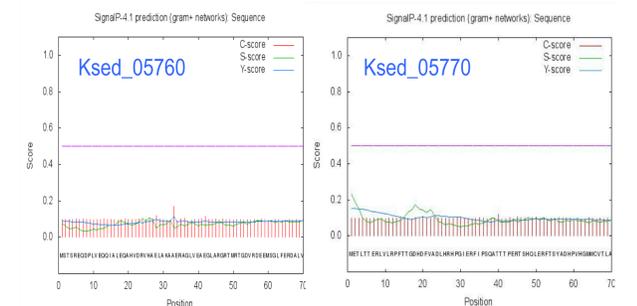


Figure V. These are the SignalP graphs for *Ksed_05760*, *Ksed_05770*, and *Ksed_05780*. Based on the fact that two of the graphs never go over the purple cutoff line to make it a signal protein, and the third graph doesn't go over enough, none of these genes include a signal peptide.

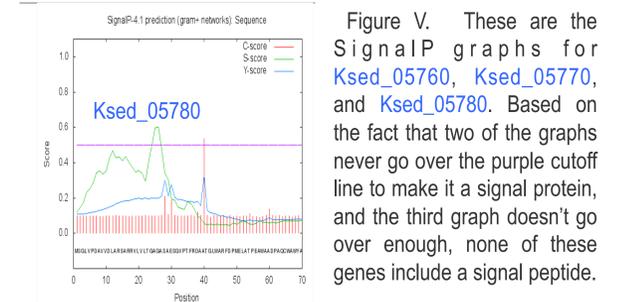
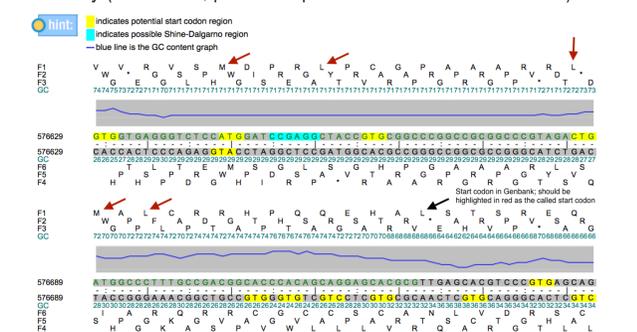


Figure VI. Sequence Viewer for Alternative Open Reading Frame Search. This site on IMG/edu is used to determine if start and stop codons were called correctly. The results for *Ksed_05760* (below) warrant further study to determine if the start codon was called correctly (red arrows, potential upstream alternate start codons).



Conclusion

The computer database might not be entirely accurate in the annotation of these two of the three genes investigated using the first 5 modules of GENI-ACT. Though the names and functions appear to be correct for all, further study of the start codon call for *Ksed_05760* will need to be undertaken to determine if the start codons for these two genes have been called correctly.

References

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