Annotation of the *Kytococcus sedentarius* Genome at Locus Tags Ksed_10160, Ksed_10320, Ksed_10520, and Ksed_19790

Rebecca Avorkliyah, Sydney Ball, Cecilia Daugherty, Katie Eipert, Connor Fitzgerald, Annie Hamer, Olivia Keagle, Jeremy Nguyen, Patrick O'Brien, Patrick Sullivan, David Thompson, and Caitlin Ullock

Pittsford Mendon High School located in Pittsford, New York and the Western New York Genetics in Research Partnership

University at Buffalo

Abstract

4 genes belonging to the organism *Kytococcus sedentarius* (Ksed_10160, Ksed_10320, Ksed_10520, Kesd_19790) were run though the collaborative gene-annotating database GENIACT. In the GenBank name of each gene was judged based on sequence based similarities, structural evidence, reading frame data, and amino acid based similarities. From this analysis the program was able to annotate the gene and come to a conclusive proposed gene name for each locus tag.

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 microscope slide submerged in sea water in 1944, *Kytococcus sedentarius* grows well in sodium chloride at concentrations less than 10% (w/v).

According to Sims et al. (2009), *Kytococcus sedentarius* is a microorganism of interest for several reasons. This bacterium is a natural source of the oligoketide 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 *Micrococcineae*, which has yet to have been thoroughly studied utilizing bioinformatics (Sims et al., 2009).

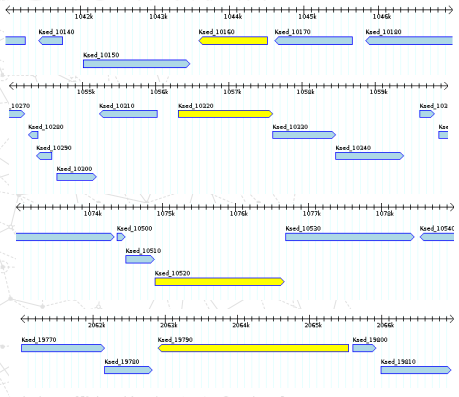

Figure 1- From top to bottom: Ksed_10160, Ksed_10320, Ksed_10520, Ksed_19790. Each tag in yellow is subjected to research.

Methods

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- 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?

Results

Ksed_10160

The annotation of the ksed_10160 that is computed by the computer is correct. When the amino acid sequence was entered into the TIGRFAMS database, it was able to be confirmed that the amino acid sequence did code for an enzyme called exodeoxyribonuclease III. The name of the enzyme was further confirmed when the PFAM database, marked the ksed_10160 as a part of the Endonuclease/Exonuclease/phosphatase family. This enzyme catalyzes the removal of mononucleotide bases from the end of a polynucleotide chain. This enzyme is found in the cellular membrane (according to the PSORTB database) to confirm this new finding the amino acid sequence was entered into the LIPTOP database and was predicted to be located in the cytoplasm of a cell. Considering that there are multiple databases agreeing on the type of enzyme and the location of it in a cell, it is fair to accept that the annotation of the Ksed_10160 is correct as computed by the computer.

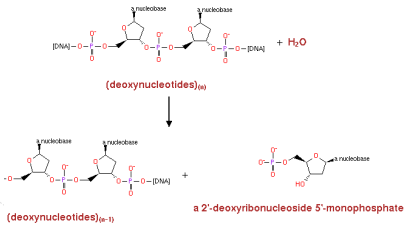

Figure 2 – Ksed_10160 exodeoxyribonuclease III catalyzes the breaking down of phosphodiester bonds.

Ksed_10320

Geni-act originally proposed this gene to be a pilus assembly protein/ATPase CpaF. Blast matched the gene to pilus assembly proteins as well. There was no Shine-Dalgarno sequence upstream of the proposed start codon, so the predicted start was may actually 5' to the predicted start codon where a Shine-Delgarno sequence was immediately upstream. COG also defined it as a CpaF. The gene is conserved well in the center but poorly towards the beginning and the end. No helixes, single peptides, or cleavage sites were observed. Both Lipo-P and PSORT-B found the protein to be likely located in the cytoplasm, with a score of 7.5. ExPASy ENZYME concluded that the protein is a Potassium-transporting ATPase.

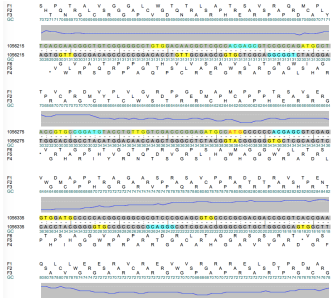

Figure 3 – Ksed_10320 suggested start codons and Shine-Dalgarno sequences

Ksed_10520

The proposed product by GENI-ACT was a DNA/RNA Helicase. It was originally predicted to be a hypothetical protein but we found a name for it in the COG search. The gene product proposal was not supported by the original BLAST search because the only results were hypothetical proteins, the COG search resulted in a name. We did have to edit the coordinates because it had a different start codon than originally predicted. The stop codon remained the same. This protein was conserved throughout the sequence. In some places it was well conserved but there were numerous times it was not. Psort-b suggested that the protein was located in the cytoplasm however TMHMM did not find any transmembrane helixes which means it wouldnt be likely to be found in the cytoplasmic membrane. Since SignalP and Phobius did not find signal peptides on the protein so it most likely is not secreted outside of the cell, meaning it is found in the cytoplasm as originally suggested. The SignalP graph pictured below shows the protein having no signal peptides, supporting the location of the protein. The location is supported by the protein having a cytoplasmic score of 7.50.

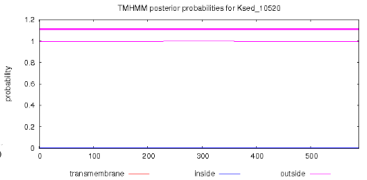

Figure 4 - Ksed_10520 does not possess a signal peptide according to SignalP

Ksed_19790

The gene annotation proposed by GENI-ACT was DNA ligase D/DNA polymerase LigD. After performing different tests and collecting data, the computer's gene annotation was correct. This was based on many tests such as BLAST, CDD, TIGRFAM, and Pfam. The BLAST resulted in bifunctional non-homologous end joining protein LigD, the CDD resulted in LigD, TIGRFAM resulted in ligD_pol: DNA polymerase LigD, polymerase domain, and Pfam resulted in LigD_N. To find where the protein is located, we used a variety of tests. TMHMM predicted zero transmembrane helices. Then however, PSORT-B predicted that it was located in the cytoplasm while Phobius showed no evidence of transmembrane helixes or a signal peptied. Ksed_19780 therefore is likely a cytoplasmic protein.

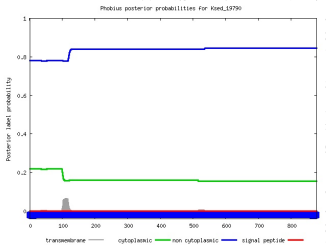

Figure 5 –The Phobius results for Ksed_19790 suggests that the protein does not have a transmembrane helix and is likely found in the cytoplasm.

Conclusions

There is no significant difference between the GENI-ACT proposed gene product and the proposed gene annotation for all of the genes in the group. Therefore, there is reason to believe the genes are correctly annotated by the computer.

Gene Locus	Geni-Act Gene Products	Proposed Annotation
Ksed_10160	Exodeoxyribonuclease III	Exodeoxyribonuclease III
Ksed_10320	Pillus Assembly Proten/ATPase	Protein Transporting ATPase
Ksed_10520	DNA/RNA Helicase	DNA/RNA Helicase
Ksed_19790	DNA Ligase D/DNA Polymerase LigD	DNA Ligase D/DNA Polymerase LigD

Reference

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

Acknowledgments

Supported by an NSF Innovative Technology Experiences for Students and Teachers (ITEST) Award - 1311902

www.buffalo.edu