Annotation of the *Kytococcus sedentarius* Genome at Locus tags Ksed_01000, Ksed_00830, Ksed_02210 and Ksed_00960

Kasey Rubin, Katie Broikos, Mike DeLucia, Jack Andreucci, Kestrel Eisenstat, Luke O'Connor, MacKenzie Winn, Caroline Hanson, Elyse Barkstrom, Ryan Walpole, Morgan Levy, and Caitlin. Ullock
Pittsford Mendon High School and The Western New York Genetics in Research Partnership

University at Buffalo

Abstract

A group of 4 genes from the microorganism *Kytococcus sedentarius: Ksed_01000, Ksed_00960,* Ksed_00830 *and Ksed_02210* 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, 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. The Genbank proposed gene product name 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 in the r 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 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).

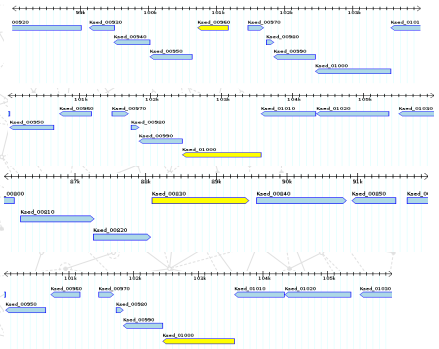

Figure 1- From top to bottom: Ksed_01000, Ksed_02210, Ksed_00830, Ksed_00960. Each tag in yellow is subjected to research

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_01000

The initial proposed product of this gene by GENI-ACT was a histidine kinase, which was ultimjately supported by our top BLAST hit for the amino acid sequence. PSORT-B predicted that our protein was most likely found within the cytoplasmic membrane of the cell; final results scored a 10.00 on PSORT-B. This prediction was also supported by the results from SignalP, TMHMM, and also Phobius. SignalP showed that there was no signal peptide, and TMHMM showed that there were at least 4 transmembrane helices, although it is quite possible that there may be 5. In conclusion, our protein most likely functions in signal transduction within the cytoplasmic membrane.

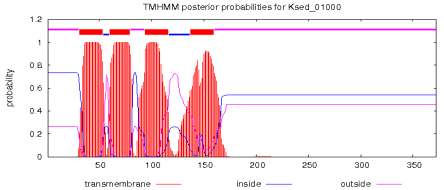

Figure 2: Transmembrane topology graph, for Ksed_01000.

Ksed_02210

The initial proposed product by GENI-Act was tryptophan 2,3-dioxygenase. Psort-b predicted this protein was found in the cytoplasm. TMHMM and Phobius did not find transmembrane helixes, meaning its not likely to be in the cytoplasmic membrane. SignalP and Phobius did not find signal peptides on the protein, meaning not likely to be secreted outside of the bacterium. Therefore it is found in the cytoplasm. There was no Shine-Dalgarno sequence originally so the coordinates may need to be changed after further study.

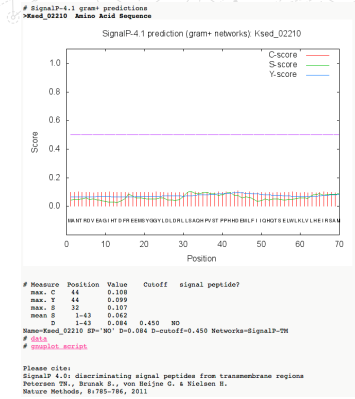

Figure 3: Signal peptide graph, Ksed_02210

Ksed_00830

The gene product is esterase/lipase and the organism name of the top blast hit was *Piscicoccus intestinalis*. The alignment length was 327 and the score 130 bits. The e-value is 1e-30. PsortB could not predict the location of the enzyme. The gram stain of the microbe is gram-positive as predicted by NCBI. TMHMM did not find any transmembrane helices so it is unlikely to be in cytoplasmic membrane. SignalP indicated signal peptides were present meaning it was likely secreted outside the bacterium. SignalP predicted one at 22-23. The signal peptide probabilities value is 0.574 and the probability must be greater than 0.450 for a signal peptide to be predicted. Ksd_00830 is threfore likely a noncytoplasmic protein. Our proposed DNA coordinates were likely correct because a Shine-Dalgarno sequence was found 5-15 spaces upstream of the proposed start codon.

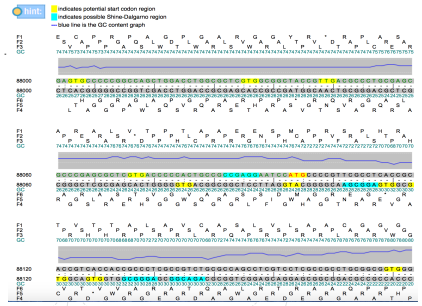

Figure 4 suggested start codons and Shine-Dalgarno sequence for Ksed_00830.

Ksed_00960:

The initial proposed product of this gene by GENI-ACT was a transcriptional regulator. The top BLAST hit supported this finding. PSORT-B and Phobius predicted this protein to be located inside the cytoplasm, with a score of 7.50. SignalP predicted that there is no signal peptide. TMHMM predicted that this protein has no transmembrane helixes. All three of these results conclude that this membrane has to be cytoplasmic. The weblogo suggests that the sequence is well conserved throughout.

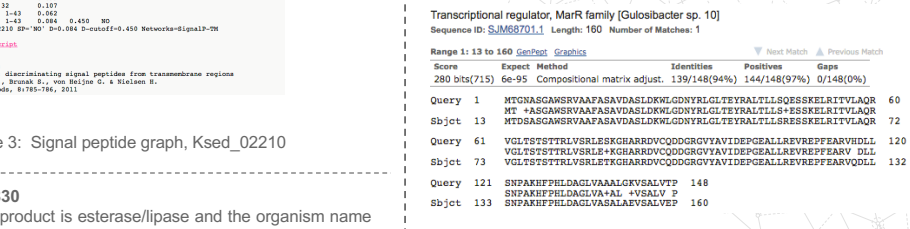

Figure 5: Pariwise alignment of top BLAST hit for Ksed_00960.

Conclusion

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 Product	Proposed Annotation
01000	Two component system, NarL family, histidine sensor kinase DesK	Histidine Kinase (HK1)
00830	Esterase/Lipase	Esterase/Lipase
02210	Catalyst, Tryptophan 2,3-dioxygenase	Tryptophan 2,3-dioxygenase
00960	Transcriptional Regulator, MarR family	Transcriptional Regulator

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