Annotation of the *Kytococcus sedentarius* Genome at Locus Tags Ksed_00430, Ksed_00200, Ksed_00740, Ksed_00310)

Charles Shi , Colleen Eckl, Casey Woodward, Kelly Knickerbocker, Lulu Tong, Devon Tayman, Brennan Mooney, Ava Roncone, Alexa Golubjatnikov, Evan Realbuto and Caitlin Ullock
Pittsford Mendon High School, Pittsford, NY

University at Buffalo

Abstract

A group of 4 genes from the microorganism *Kytococcus sedentarius* (Ksed_00430, Ksed_00200, Ksed_00310 and Ksed_00740) 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. 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 the nr database.

Introduction

Kytococcus sedentarius is a strictly aerobic, non-motile, non-encapsulated, catalase and oxidase positive, 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. *Kytococcus sedentarius* grows well in sodium chloride at concentrations less than 10% (w/v). It is primarily isolated from human skin and plays no role in disease.

Kytococcus sedentarius is a marine dwelling Gram positive bacterium in the genus *Kytococcus*. It is known for the production of polyketide antibiotics as well as for its role as an opportunistic pathogen. It is strictly aerobic and can only grow when amino acids are provided.It is found in tetrads, irregular clusters, and cubical packets of eight. It is catalase positive, oxidase positive, and exhibits strictly aerobic metabolism. Optimum growth temperature is 25-37 C. It is primarily isolated from human skin, and plays an opportunistic role in disease. *Kytococcus sedentarius* was once considered a species of the genus *Micrococcus*

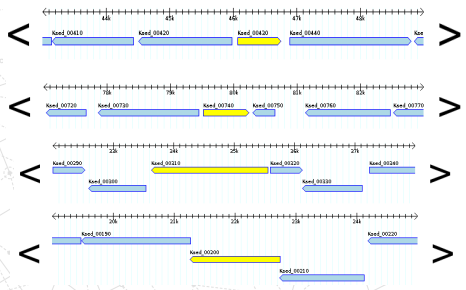

Figure 1. The locus tags and relative position of the genes under investigation in this research

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	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 Sequence	Has the amino acid sequence of my protein been called correctly by the computer?
Module 5- Structure-Based Evidence	TIGRfam, Pfam, PDB	Are there any functional domains in my protein?
Module 6- Enzymatic Activity	Enzymatic Pathway,	What is the function of my gene product?

Results

Ksed_00430:
The initial proposed product of this gene by GENI-ACT was a transcriptional regulator. This gene product proposal was supported by the top BLAST hits for the amino acid sequence, the presence of well-curated protein functional domains within the amino acid sequence, the transmembrane topography of the amino acid sequence, and the cellular location of the amino acid sequence. As such, the proposed annotation is a transcriptional regulator. However, in the Alternative Reading Frame, there was no Shine Dalgarno region immediately upstream of the proposed start codon so we proposed a new start codon at complement(45994..46758).

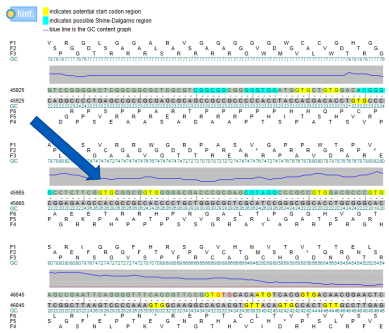

Figure 2 – The IMG database gene page for Ksed_00430. The computer generated start codon had no Shine-Dalgarno sequence 5-15 bases upstream, so a possible new proposed start codon is where the arrow is pointing because there is a Shine-Dalgarno sequence in close proximity

Ksed_00310:
The initial proposed product of this gene by GENI-ACT was glutaminase II. We also found the same result that the product of this gene is glutaminase II based on our top BLAST hits. Based on SignalP there was no signal peptide found. Also, PSORT-B predicted that this protein is located in the cytoplasm with a final prediction of 7.50. In support of the psortB findings, on the graphs of Phobius and TMHMM posterior probabilities for this gene, we found no evidence of a signal peptide or transmembrane helixes.

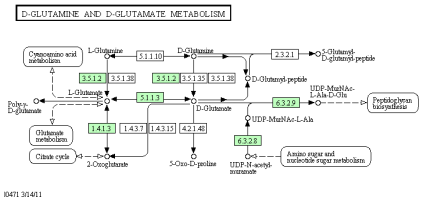

Figure 3 - This is a pathway map that \Ksed_00310 was found to be part of.

Ksed_00740
The initial proposed product of this gene by GENI-ACT was peptidase E. This gene product proposal was supported by the top BLAST hits. Based on the Phobius, TMHMM, SignalP, and PSORT-B tests, the location of this enzyme within the cell is uncertain. For instance, the lack of transmembrane helixes from the TMHMM test shows that this enzyme is not located within the membrane. However, it is possible that it could be on the inner or outer most part of the membrane. Since all the scores from the PSORT-B test were equal, the prediction for the enzyme's location was deemed "unknown". According to the SignalP test, the gene lacks any signal peptides as well. Additionally, we may want to propose new coordinates for the start of the enzyme on the gene itself. The new proposed coordinates are from DNA positions 79600 to 79613 since it was the only part of the gene that had a start codon preceded by a Shine-Dalgarno sequence that was 5-15 bases upstream. Lastly, the closest organisms to the *Kytococcus sedentarius* are *Thermoscrispum municipale* and *Serinicoccus chungangensis*

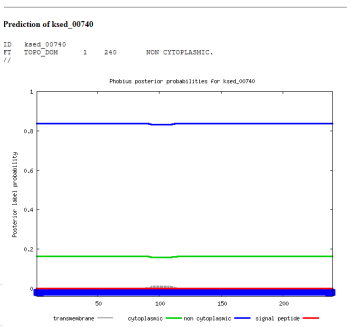

Figure 4– Phobius results and prediction for Ksed_00740

Ksed_00200:
The initial proposed product of this gene by GENI-ACT was the penicillin-binding protein called transpeptidase. This product was supported by the evidence from the BLAST results which verified the amino acid sequence, the LipoP results which declared the gene most likely to be a "Single peptidase I," and the DNA coordinates from the Alternative Open Reading Frame, which contained a Shine-Dalgarno sequence 5-15 bases upstream of the codon, matched the original coordinates. Also, in the Structure-Based Evidence module, the TIGRFAM tool verified that the structure was that of a "penicillin-binding protein" which, upon further research, are DD-transpeptidases. Furthermore, the Pfam tool produced one search result which was transpeptidase, as well. Therefore, the initial proposed product was proven to be named correctly.

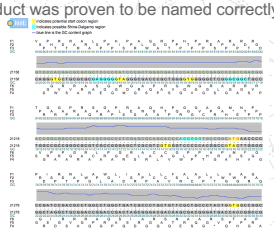

Figure 5- The IMG database gene page for Ksed_00200. The computer generated start codon had a Shine-Dalgarno sequence 5-15 bases upstream, so it was called correctly.

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 Products	Proposed Annotation
Ksed_00430	transcriptional regulator	transcriptional regulator
Ksed_00310	glutinamase	glutaminase
Ksed_00740	peptidase E	peptidase E
Ksed_00200	cell division protein FtsI/penicillin-bindingprotein 2	cell division protein FtsI

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