University at Buffalo *The State University of New York* | REACHING OTHERS

Annotation of the *Kytococcus sedentarius* Genome from DNA Coordinates 582517 to 585970

Joshua Guzman-Gaitan, Mercedes Velazquez, Kayla Sherk, Elijah Kirkland-Johnson and MaryAnn Roll

Science Department Western New York Maritime Charter School and Western New York Genetics in Research Partnership

Abstract

A group of four consecutive genes from the microorganism *Kytococcus sedentarius* (Locus tag: Ksed_05820 – Ksed_05850) were annotated using genome annotation modules in GENI-ACT (Genomics Educational National Initiative- Annotation Collaboration Toolkit). The objective was to determine if the genes had been annotated correctly in Genbank. The proposed gene product name for each gene was evaluated by exploring modules 1, 2, 3, and 4 that include the basic information of the gene, amino acid sequence based similarity data, cellular localization data, and structure based evidence. The proposed gene product names for all four genes 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 in the database.

Introduction

Kytococcus sedentarius is a gram positive bacteria that is responsible for the disease pitted keratolysis that can become serious in people with compromised immune systems. It is known that automated annotations include up to 35% error rates. This was one of the many reasons for undertaking manual annotation of the genes. Our objective was to make sure the gene was identified correctly and its function was predicted accurately.

According to S. Pospisil, et al. (1998) Kytococcus sedentarius is a producer of the antibiotic monensin, an antibiotic often used in animal food products. Kytococcus sedentarius is also suspected to be responsible for the death of a 55 year old man who had undergone chemotherapy. With chemotherapy, the man was severely immunocompromised, thus *K. sedentarius* broke its way into the blood stream, ultimately reaching and invading the lungs, causing major, and fatal, hemorrhagic pneumonia (Levenga, et al, 2003).

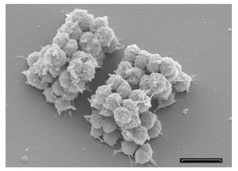

Figure I: Electron micrograph of Kytococcus sedentarius (Sigs, 2009)

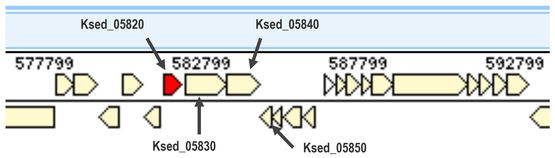

Figure II: Gene neighborhood from IMG/EDU (http://img.jgi.doe.gov/cgi-bin/edu/main.cgi). The four consecutive genes that were studied are labeled.

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- 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?

Table 1

Results

Ksed_05820:

The initial proposed product of this gene by GENI-ACT was Pyridoxamine 5'-phosphate oxidase. This gene product was supported by the BLAST hits along with CDD and T-Coffee results. There does not seem to be a difference in the product determined by both the proposed product and personal research. Web logo provides info that the gene also has a range of well-conserved amino acids. Using TMHMM cellular localization tool, there were no transmembrane helices predicted. SignalP tool indicated a "YES" with a cleavage site between position 16 and 17 for the presence of a signal peptide. However, the c, s and y plots in the graphical output do not support that conclusion. The D score however does exceed the cut-off value and thus the prediction is a "yes". The final prediction by PSORT-B with a score of 8.91, is extracellular. The results from Phobius predicts a high probability of a signal peptide. Thus, although the enzyme is indicated to be extracellular, we propose further analyses for confirmation.

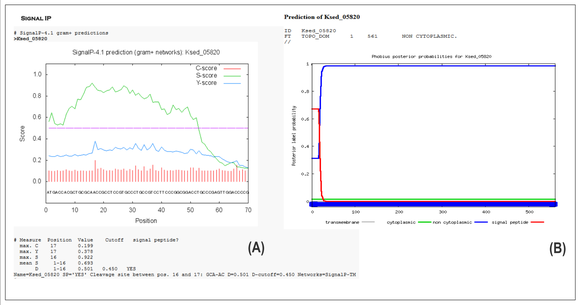

Figure III: (A) SignalP graph showing a predicted signal peptide with c,s,and y plots not confirming the prediction. The final measure D exceeds cutoff value. **(B)** Phobius graph also showing a high probability of the presence of a signal peptide (red)

Results

Ksed_05830

The initial proposed product of this gene by GENI-ACT was a phosphoribosylamine-glycine ligase. It is indicated to be correctly annotated due to significant matches with similar proteins in *Renibacterium* and *Arsenicicoccus* by BLAST search against the nr database. COG0151 also supports this claim and finds the enzyme to be involved in nucleotide transport and metabolism. The data from T-COFFEE and WebLogo show that the amino acids in the gene was very well conserved. Results of TIGRFAM and PFAM show that the protein comes from a family with similar structures and functions.

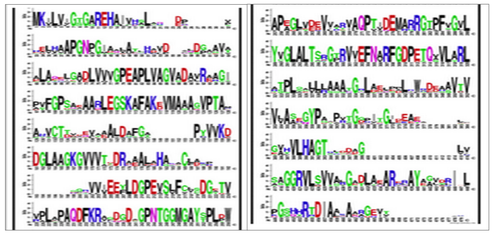

Figure IV - The height and number of amino acid stack in several positions in the Weblogo from N–C terminal indicates that amino acids in the protein are well conserved among the ten orthologs chosen for sequence alignment by T-coffee

Ksed_05840

The initial proposed product of this gene by GENI-ACT was a phosphoribosylaminoimidazole-succinocarboxamidesynthase. This gene product proposal was supported by BLAST and CDD searches (nr database). T-COFFEE analysis followed by WebLogo revealed that the protein was well conserved indicating important functionality.

Ksed_05850

The initial proposed product of this gene by GENI-ACT was a death-on-curing protein. BLAST search against the nr database returned a moderately high significant match to a similar protein in *Serinicoccus*. Multiple alignment of amino acid sequence from ten orthologs using T-Coffee followed by WEBLOGO indicates the protein to be well conserved. TMHMM, SignalIP and Phobius do not show any transmembrane helices or signal peptide. PSORTb shows the highest score for cytoplasmic membrane. The cellular localization of the protein is unclear and needs further investigation. According to the top hit of TIGRFAM this gene is in a Death-On-curing protein family.

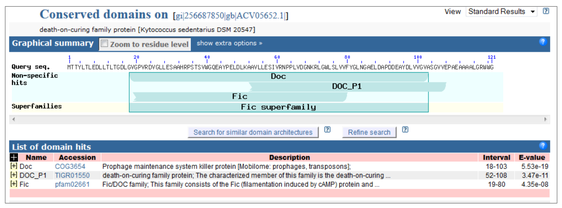

Figure V: The CDD search shows COG, TIGR and pfam hits with significant e-values. This protein is indicated to be part of the Fic (filamentation induced by cAMP)/ DOC (death on curing) protein family.

Conclusions

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 automated methods.

Gene Locus	GENI-ACT Products	Proposed Annotation
05820	Pyridoxamine 5' - phosphate oxidase	Pyridoxamine phosphate oxidase
05830	Phosphoribosylamine-glycine ligase	Phosphoribosylamine-glycine ligase
05840	phosphoribosylaminoimidazole-succinocarboxamidesynthase	phosphoribosylaminoimidazole-succinocarboxamidesynthase
05850	Death-on Curing Protein	Death-on-Curing Protein

Ksed_05820, although a signal peptide is indicated both by Signal IP and Phobius, the graphical output in Signal IP does not corroborate the textual data.

Ksed_05850 the output from PSORT-b indicates a cytoplasmic membrane protein which is not corroborated by either TMHMM or Phobius.

We recommend further research to determine cellular localization for genes in locus tags Ksed_05820 and Ksed_05850.

We learn that the ambiguities presented during genome annotation using bioinformatics tools *in silico* need elucidation by further research into literature as well as wet lab experimentations.

References

Levenga, et al. (2003). Fatal Hemorrhagic Pnuemonia caused by Infection due to Kytococcus Sedentarius- a Pathogen or Passenger? 447-449

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

Pospisil, et al. (1998). Kytococcus sedentarius (formly Micrococcus sedentarius) and Dermacoccus nishinomiyaensis (formly Micrococcus nishinomiyaensis) produce monensins, typical Streptomyces cinnamonensis metabolites. 1007-1011.

Acknowledgments

We would like to thank Drs. Rama Dey-Rao and Stephen Koury for assisting with the training and implementation of this project.

Supported by NSF ITEST Strategies Award Number 1311902