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

Annotation of the *Kytococcus sedentarius* Genome from DNA Coordinates 676559 to 679774

Zachary Adams, Rebecca Corby and Dennis Bauer

Amherst High School and the Western New York Genetics in Research Partnership

Abstract

A group of consecutive 2 genes from the microorganism *Kytococcus sedentarius* (Ksed_06850 and Ksed_06860) were annotated using the collaborative genome annotation website GENI-ACT. The Genbank proposed possible results of each gene 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. Gene 06850 is responsible for work associated with DNA repair, intramolecular recombination, transcription and transport of nucleotides. Gene 06860 is responsible for breaking down ATP to gain energy for transportation. The Genbank proposed gene product name did not differ significantly from the proposed gene annotation for either of the genes in the group and as such, the genes appear to be correctly annotated by in the database.

Introduction

Kytococcus sedentarius is an aerobic bacteria and pathogen which can cause hemorrhagic pneumonia, pitted keratolysis, and valve endocarditis (Sims et al., 2009). It is gram-positive and nonmotile, and is capable of producing polyketide antibiotics (Pospísil et al., 1998). Its enzymes which degrade human callus may also be of commercial use (Longshaw, Wright, Farrell, Holland, 2002). The organism's genome was sequenced and annotated by the DOE Joint Genome Institute due to its lack of close relatives as well as this potential medical use. While the error rate of sequencing is "less than 1 in 100,000" (Sims et al., 2009), annotation can be less accurate, which is why students have annotated genes manually with the help of databases and search engines such as Pfam and BLAST.

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?
Module 6- Enzymatic Function	KEGG, MetaCyc, E.C. Number,	In what process does my protein take part?
Module 7- Gene Duplication/ Gene Degradation	Paralog, Pseudogene	Are there other forms of my gene in the bacterium? Is my gene functional?
Module 8- Evidence for Horizontal Gene Transfer	Phylogenetic Tree,	Has my gene co-evolved with other genes in the genome?
Module 9- RNA	RFAM	Does my gene encode a functional RNA?

Results

Ksed_06850:

Ksed_06850 is a gene formed from 2142 Nucleotides producing 713 Amino Acids. According to Sequence-based Similarity Data of a sequence blast, this gene is most similar to the Full=Probable ATP-dependent DNA helicase RecQ in Bacillus subtilis subsp. subtilis str. 168. Due to the low E value of 6e-48, it can be inferred that Ksed_06850 performs the same or a similar action that Full=Probable ATP-dependent DNA helicase RecQ preforms. The Full= Probable ATP-dependent DNA helicase RecQ is a Probable DNA helicase, required for DNA repair and intramolecular recombination. It most likely acts to help generate single strand-DNA from double stranded-DNA breaks. The Gene utilizes ATP + H2O + ADP + phosphate in order to make energy. Ksed_06850 shares genes from the organism Arabidopsis thaliana which play a role in the repair of DNA, are important in genome maintenance, and function through catalyzing the reaction: ATP + H2O → ADP + P and thus driving the unwinding of paired DNA and translocating in the 3' to 5' direction. Conserved Domain Database Search, which investigates Amino acid to determine similarity, revealed that Ksed_06850 is similar to Superfamily II DNA helicase RecQ which is involved in replication, recombination and repair of genetics. The program TIGRFAM looks at protein families known to have similar functions and determined Ksed_06850 is similar to DECH_helichelicase/secretion neighborhood which suggests that this helicase may play a role in conjugal transfer of DNA. Pfam revealed that Ksed_06850 has similar protein domains and families with DEAD/DEAH box helicase, a family of proteins associated with unwinding nucleic acid. TMHMM determined that there is a lack of helixes and the protein is not a transmembrane protein. SignalP data concluded that no signal peptide is present. PSORTb determined that the most probable location of the protein is in the cytoplasm, it may be found with DNA where previous data suggests. Phobius predicted no transmembrane helixes nor a the presence of a signal peptide, supporting the idea that Ksed_06850 is found in the cytoplasm.

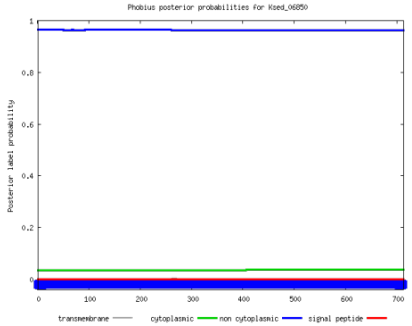


Figure 1. The Phobius prediction for the protein produced by Ksed_06850 showing the lack of transmembrane helixes and no signal peptide supporting the idea that it is a protein that operates the cytoplasm of the cell. This information supports the hypothesis that Ksed_06850 is involved with DNA.

KEGG and Metacyc determined that the protein created by Ksed_06850 is involved with RNA metabolism (Figure 2). No paralogs or pseudogenes were detected. An amino acid blast was performed to locate similar genes in different organisms. From this blast, a Cladogram and Phylogram was produced which indicated that there was no evidence for horizontal gene transfer as *Kytococcus sedentarius* was placed next to closely related species a TreeDyn phylogenetic tree and a Drawtree radial tree output determined the same A Chromosome Viewer Heat Map was analysed and the total genomic percentage was 75% and the genomic percentage of Ksed_06850 was 72% once again determining that no horizontal gene transfer occurred. Ksed_06850 is not involved in the production of RNA as determined by rFam

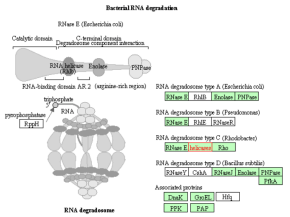
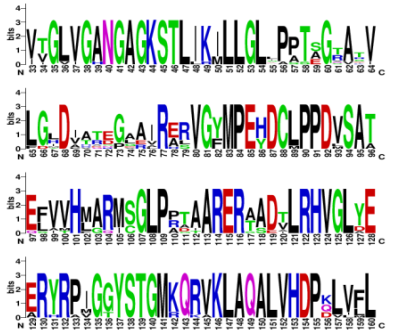


Figure 2. KEGG pathway map for Ksed_06850 suggesting that Ksed_06850 (red text) is part of the RNA degradosome in *Kytococcus sedentarius*.

Ksed_06860:

GENI-ACT initially proposed that Ksed_06860 produced an ATPase component of an ABC-type multidrug transport system. The top BLAST results included Bacitracin transport ATP-binding protein BcrA and Nod factor export ATP-binding protein I, which means that they likely have similar functions to my gene. The CDD result also provided the exact name of the protein produced, and the WEBLOGO demonstrated that its alignment with these similar proteins was well-conserved, further justifying the initial proposal. My TIGRFAM search also resulted in subfamilies of ABC transporters, the first having "daunorubicin resistance," and the second having "lantibiotic protection." While I knew what my gene had in common with other genes, I had no idea what exactly that meant about its function. Within the Structure-based Evidence module, I used the Pfam database to determine what exactly "ABC-type multidrug transport system, ATPase component" meant. The protein produced by this gene is part of the ATP-binding domain of ABC transporters. These are involved in active transport, utilizing ATP to move many compounds across a membrane, which can provide nutrients in prokaryotes or remove toxins in any organism. The ABC protein exists within the cell and links to TMD transmembrane proteins. It, then, made sense that the TMHMM determined that my protein had no transmembrane helices, and that no signal peptide was predicted by the SignalP. KEGG Pathway seemed like it would be useful, as there was a diagram listing ABC transporters, yet my gene was nowhere to be found within it. When using BLAST again and comparing ksed_06860 to *Kytococcus sedentarius* as a whole, many significant paralogs were found. However, this cannot be a pseudogene, as IMG/EDU led to a perfect match between the translated DNA and the expected amino acid sequence and 97.8% of the sequence is covered, meaning that there is not a frameshift mutation or premature stop codon. ScanProsite also showed that every amino acid meets conditions that lead to functionality, meaning that the protein can be used in *Kytococcus sedentarius* cells.

My WEBLOGO shows that many areas of the protein are well-conserved. This can be seen in the wide, tall letters, while smaller letters represent areas in which the protein is more variable. This means that the product of this gene is very related to its orthologs in its amino acid sequence, which causes similar folding and similar function.



Conclusion

Ksed_06850

The GENI-ACT website predicted that Ksed_06850 produced ATP-dependent DNA helicase RecQ and this was supported by my research as indicated in the result area. Ksed_06850 is required for DNA repair and intramolecular recombination, utilizes the reaction of ATP + H2O = ADP + phosphate, regulates transcription of RNA polymerase II-dependent gene, has a key role in repair of DNA, drives the unwinding of paired DNA and translocating in the 3' to 5' direction and is involved with RNA metabolism. Ksed_06850 proteins are most likely to be found inside the nucleus.

Ksed_06860

GENI-ACT suggested that ksed_06860 produces an ATPase component of an ABC-type multidrug transport system. This was supported by my research overall. Ksed_06860 codes for a protein that binds to ATP, breaking it down to use its energy for active transport. It is specifically involved in antibiotic resistance, as it removes drugs from the cell. While this transport system is transmembrane as a whole, ATP is broken down within the cytoplasm, and the protein does this near others that use the resulting energy to move materials.

References

Longshaw, C.m., J.d. Wright, A.m. Farrell, and K.t. Holland. "Kytococcus Sedentarius, the Organism Associated with Pitted Keratolysis, Produces Two Keratin-degrading Enzymes." Journal of Applied Microbiology 93.5 (2002): 810-16. PubMed. Web. 11 May 2016.

Pospísil, S., O. Benada, O. Kofronová, M. Petrícek, L. Janda, and V. Havlícek. "Kytococcus Sedentarius (formerly Micrococcus Sedentarius) and Dermacoccus Nishinomiyaensis (formerly Micrococcus Nishinomiyaensis) Produce Monensins, Typical Streptomyces Cinnamonensis Metabolites." Canadian Journal of Microbiology 44.10 (1998): 1007-011. PubMed. Web. 11 May 2016.

Sims et al. (2009). Complete genome sequence of Kytococcus sedentarius type strain (541T). Standards in Genomic Sciences, 12-20. PubMed. Web. 11 May 2016.

Acknowledgments

Supported by NSF ITEST Strategies Award Number 1311902 and Dr. Stephen Koury and Dr. Rama Dey-Rao