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

Annotation of the *Kytococcus sedentarius* Genome from Locus Tags Ksed_08250 to Ksed_08280

Gabrielle Snyder, Megan Cassidy, Zoe Jerome, Austin Feasley, and Michelle Berne

Eden Junior-Senior High School and the Western New York Genetics in Research Partnership

Abstract

A group of four consecutive genes from the *Kytococcus sedentarius* microorganism (Ksed 08250- Ksed 08280) were studied and annotated using the online data base GENI-ACT. This study was done to prove the computer correctly annotated the genome *Kytococcus sedentarius*. The Genbank proposed gene product name for each gene was assessed in terms of the basic genetic information, sequence-based similarity data, structure-based evidence, cellular localization data, enzymatic function, presence or absence of gene duplication and degradation, and the possibility of horizontal gene transfer. The proteins studied were analyzed using BLAST, CDD, T-Coffee, and many other online data bases. We also used WebLogo to create visuals to understand the genome better. After-school STEM (science, technology, engineering, and math) programs such as this are new in Eden High School, but they are very popular and growing in interest. These topics help increase knowledge, get teens excited about them, help build real-life skills, and prepare them for jobs.

Introduction

Kytococcus sedentarius was originally discovered in a marine environment, specifically sea water, on a microscope slide in 1944 but has also been found on human skin (Sims, 2009). Previously classified under the genus Micrococcus, it was later reclassified under the Kytococcus genus. The bacterium also falls under the actinobacterial family name Dermacoccaceae (Sims et al., 2009) and is classified as chemoorgantrophic, which means that it obtains energy from the oxidation of carbon compounds (including amino acids). Ksed is also known to be non-motile, non-endospore forming, nonencapsulated, strictly aerobic, gram-positive, catalase positive, and oxidase positive. It is a spherical bacterium commonly found in tetrads, irregular clusters, and cubical packets of eight. Optimal growth conditions are in temperatures from 25-37° C in conditions rich in amino acids, specifically methionine and sodium chloride at concentrations under 10% (w/v). Though relatively harmless to most humans, a study by Longshaw et al. (2002), suggests *Kytococcus sedentarius* produces two callus digesting enzymes and is responsible for the symptoms of the pitted keratolysis skin disorder. It has also been noted in rare cases that Ksed causes fatal pneumonia in immunosuppressed patients (Chaudary et at. 2010). Species of *Kytococcus* are commonly resistant to penicillin and methicillin as well (Chaudary et at. 2010)

The *Kytococcus* genome is the subject of this study. Produced by DNA, genomes are sets of instructions based on a four letter sequence of adenine, cytosine, guanine, and thymine. The Ksed genome was first sequenced in 2009 by D. Sims et al. and was determined to be 2,758,024 bp long and a single replicon genome that has 2,639 protein codings and 64 RNA genes. Often known as "decoding", genome sequencing is the collecting and finding of the order of nucleotides in a DNA sequence. Each gene within the genome is assigned a locus tag by researches to better identify each. Genomes of *Kytococcus sedentarius* were placed in GENI-ACT for groups to confirm the computer results that were received through a Genome Online Database called GenBank. Students took the results from GenBank and compared them to other amino acid sequences found among other common organisms and bacterium using programs like Blast, CCD, T-Coffee, and WebLogo. Manual annotation of genomes is important to confirm computer results and further understand the biology of organisms. *Kytococcus sedentarius* was chosen specifically for for numerous reasons, the main three being that this bacterium has the potential to be a source of natural oligoketide antibiotics (Sims et al., 2009) such as monensin; has been found to be responsible for varying infections including valve endocarditis, hemorrhagic pneumonia, pitted keratolysis as previously mentioned, and other opportunistic infections (Sime tal., 2009); and has yet to be studied entirely (Sims et al., 2009). Manual annotations of *Kytococcus sedentarius* are crucial in understanding the genome and how it works.

Method 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- Structure-Based Evidence	TIGRfam, Pfam, PDB	Are there functional domains in my protein?
Module 4- Cellular Localization Data	Gram Stain, TMHMM, SignalP, PSORT, Phobius	Is my protein in the cytoplasm, secreted or embedded in the membrane?
Module 5- 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 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?
Final Annotation	Review data from all modules	Does the student proposed name of the gene agree with that proposed by the automated computer annotation? Are any changes proposed to the pipeline annotation?

Results

*Kytococcus sedentarius*08250:

The proposed product of this gene by GENI-ACT was a cupin domain-containing protein. This gene product prediction was supported by the top amino acid sequence hits in BLAST, the PDB search that matched the sequence with *Cupin 2 conserved barrel domain protein from Leptotrichia buccalis* (as pictured), and the CDD search that found the amino acid sequence to relate to *Cupin domain protein related to quercetin dioxygenase*.

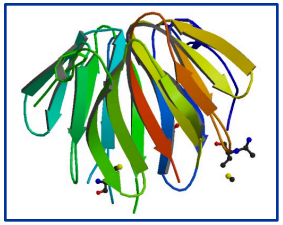

Figure 1: The above image is a computer-generated prediction of the structure of the Cupin 2 conserved barrel domain protein.

*Kytococcus sedentarius*08260:

The gene is showing to be a septum formation inhibitor Maf that is lacks transmembrane helixes, which supports the original prediction which said the Maf would be found in the cellular cytoplasm. All computer called predictions remained very consistent throughout the experimentation. The sequence length was 573 base pairs, and 190 amino acids. The DNA coordinates turned out to be 840769...841341. The gene is believed to be found in the organism *Ornithinimicrobium pekingense*, as shown through BLAST.

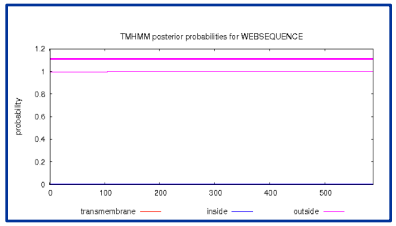

Figure 2: The diagram above shows that the inhibitor Maf lacks transmembrane helixes.

*Kytococcus sedentarius*08270:

The proposed product of this gene by GENI-ACT was an ABC-type multidrug transport system and ATPase and permease component. This gene product proposal was supported by the top BLAST hits for the amino acid sequence of Bacillus subtilis and Putative multidrug export ATP-binding/permease protein YgaD, the presence of well-curated protein functional domains within the amino acid sequence, the transmembrane topography of the amino acid sequence which predicted three transmembrane helices, and the cellular location of the amino acid sequence which was determined to be in the cytoplasmic membrane.

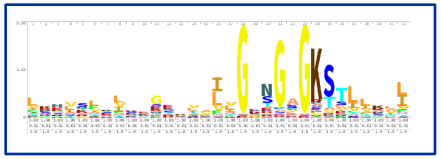

Figure 3: This picture, made using the protein sequence of Ksed_08270, depicts HMM logo resulting from the PFAM tool. The key functional/structural residues of ABC Transporter PF00005, namely G18, G21, G23, and K24.

*Kytococcus sedentarius*08280:

The initial proposed product of this gene by GENI-ACT was an ABC-type bacteriocin/lantibiotic exporter with N-terminal double-glycine peptidase domain. The gene proposal was supported well by the top BLAST hits for the amino acid sequence. The transmembrane topography of the amino acid sequence also showed very similar to the proposal. The T-COFFEE and Weblogo also showed very similar results to the gene. Therefore the proposed annotation of the gene is that it is an ABC-type bacteriocin/lantibiotic exporter with N-terminal double-glycine peptidase domain.

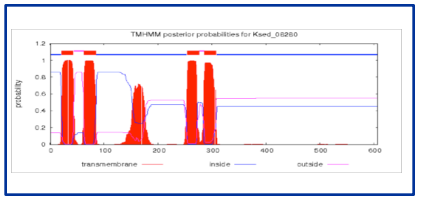

Figure 4: This shows the presence of 4 transmembrane helixes for *Kytococcus sedentarius* 08280.

Conclusion

The proposed gene annotations do not differ from the proposed gene products according to GENI-ACT.

Gene Locus	Geni-Act Products	Proposed Annotations
08250	Cupin Domain-Containing Protein	Cupin Domain-Containing Protein
08260	Septum Formation Inhibitor Maf	Septum Formation Inhibitor Maf
08270	ABC-type multidrug transport system, ATPase and permease component	ABC-type multidrug transport system, ATPase and permease component
08280	ABC-type bacteriocin/lantibiotic exporter with N-terminal double-glycine peptidase domain	ABC-type bacteriocin/lantibiotic exporter with N-terminal double-glycine peptidase domain

References

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

Chaudhary, D., Finkle, S.N. (2010). Peritoneal Dialysis- Associated Peritonitis Due to *Kytococcus sedentarius*. *Peritoneal Dialysis International* 30(2):252-253.

Acknowledgments

Supported by NSF ITEST Strategies Award Number 1311902