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

stemhs ROCHESTER NY | NSF

Annotation of the *Kytococcus sedentarius* Genome from DNA Coordinates 500258 to 504432

Annie Leach, Nyaiyah White, Christian Mika, Persephone Louissaint, Armani Rhim, Jodi Cook

Rochester City School District, Rochester STEM, Academy of Health Sciences

Abstract

Kytococcus sedentarius is a gram positive species of bacteria that is responsible for causing several diseases including infective endocarditis, hemorrhagic pneumonia , and pitted keratolysis. Although it can be a harmful pathogen it is also a beneficial biotechnology organism known to produce Monensin A and B, important oligoketide antibiotics. The entire genome of *Kytococcus sedentarius* has been annotated using computer software and models; however, manual annotation is necessary to confirm computerized annotation. Manual annotation was performed using the GENI-ACT (Genomics Education National Initiative Annotation Collaboration Toolkit) notebook and basic biotechnology and bioinformatics resources. The computer called coordinates for all five genes annotated were accepted.

Introduction

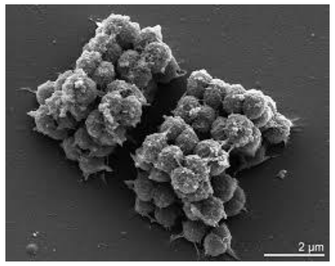

Figure 1. Scanning electron micrograph of *K. sedentarius* (Sims, et al., 2009).

Kytococcus sedentarius is a gram positive species of bacteria that is not well studied. Although it can be potentially harmful, it also has its benefits in biotechnology. As a pathogen it can cause infective endocarditis, one of the many cardiovascular diseases that threatens the main valves and arteries of the heart (Sims, et al., 2009; AHA, 2015). *K. sedentarius* also has the potential to cause hemorrhagic pneumonia (Levenga, et al., 2004), and pitted keratolysis (Longshaw, et al., 2002). While it has the ability to be harmful to humans, it also has the capability of producing Monensin A and B, important oligoketide antibiotics (Pospisil, et al., 1998).

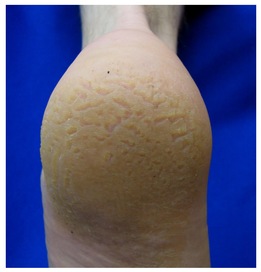

Figure 2. Pitted keratolysis forming on the heel of an individual (The Health Science, 2013).

Even though the entire genome of *Kytococcus sedentarius* has been annotated using computer software and models (Sims et al., 2009), it is important to manually annotate this genome to fully understand the pathways and functions of the different proteins encoded for by the genes and find potential computer called errors. This project brought together several students from Rochester STEM high school with varying backgrounds and interests, including health science, engineering, and information technology. While we all knew very little about annotating genes in the beginning, through this unique experience we were able to collaborate with one another, sharing and gaining valuable knowledge about concepts in biology, informatics, and technology all while contributing to a current and global research project.

The manual annotation of *Kytococcus sedentarius* genes was performed using the GENI-ACT (Genomics Education National Initiative Annotation Collaboration Toolkit) notebook. Our team used different tools, including BLAST (Basic Local Alignment Search Tool), T-Coffee Multiple Sequence Alignment Tools, and WebLogo to find similar sequences in other organisms and possible gene functions. To learn where our proteins were located in the bacterium we used the abilities of TMHMM (prediction of transmembrane helices in proteins), PSORT-b, and Phobius. Some of the group members were also able to analyze if there were functional domains in their protein, if there was any enzymatic function related to their protein, as well as possible alternate open reading frames. One member even explored the possibilities of paralogs and pseudogenes related to their particular gene. We found that all genes had been correctly called by the computer. As a group of genes they varied in cellular localization and function.

Methods and Materials

Modules	Activities	Questions Investigated
Basic Information	DNA Coordinates and Sequence, Protein Sequence	What is the sequence of my gene and protein? Where is it located in the genome?
Sequence-based Similarity Data	Blast, CDD, T-Coffee, WebLogo	Is my sequence similar to other sequences in Genbank?
Structure-based Evidence	TIGRfam, Pfam, PDB	Are there functional domains in my protein?
Cellular Localization Data	Gram Stain, TMHMM, SignalP, PSORT-b, Phobius	Is my protein in the cytoplasm, secreted or embedded in the membrane?
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?
Enzymatic Function	KEGG, MetaCyc, E.C. Number	In what process does my protein take part?
Duplication and Degradation	Paralog, Pseudogene	Are there other forms of my gene in the bacterium? Is my gene functional?

Results

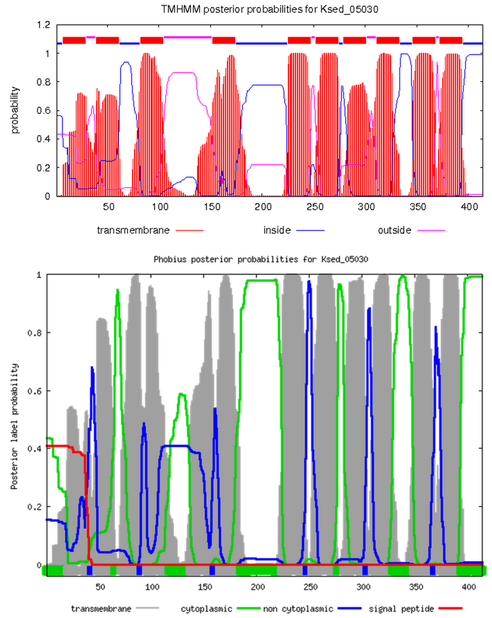

Figure 3. TMHMM (top) results for Ksed_05030 predicting 10 transmembrane helices. Phobius (bottom) for Ksed_05030 confirming the results of the TMHMM test for transmembrane helices. However, the Phobius is predicting a total of 12 helices. These results suggest that the protein is an integral membrane protein.

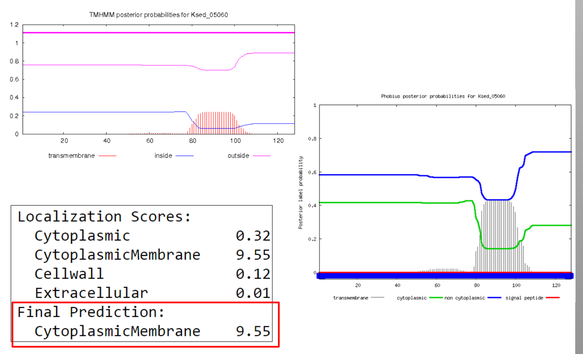

Figure 4. The transmembrane topology graph for Ksed_05060 (top left) shows no evidence of any transmembrane helices. The Phobius probability graph for Ksed_05060 (top right) confirms the results that the protein is not crossing the cytoplasmic membrane. Although, these two analytical tools do not predict a transmembrane protein, the PSORT-b localization prediction tool (bottom left) gives a final prediction of the protein's location in the cytoplasmic membrane with a high score of 9.55 (highlighted by red box). Based on these results, it is possible that the protein may be a peripheral membrane protein, embedded in only the inside part of the membrane and not making it through all the way to the outside portion of the membrane.

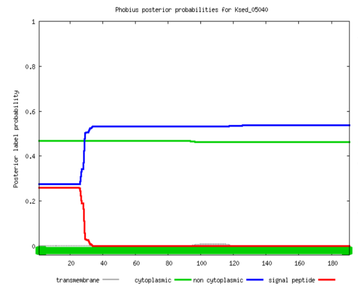

Figure 5. Phobius results for Ksed_05040 showing no transmembrane helices. But the PSORT-b results predict a cytoplasmic membrane with a score of 9.55. This could possibly mean that the protein sits on the edge of the membrane.

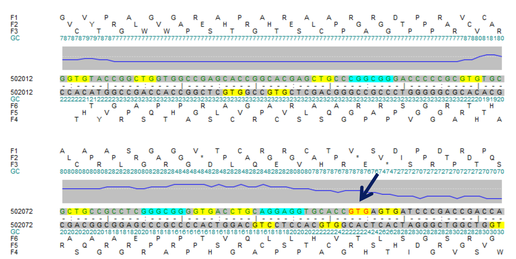

Figure 6. The computer called start codon (in red) for Ksed_05050 has a Shine-Dalgarno region 5 to 15 base pairs upstream, but, this potential start codon is GTG coding for the amino acid Valine (V) rather than the typical ATG sequence codes for the amino acid methionine (M) that signals the start of transcription and translation. Four possible alternate open reading frames were explored , however, none of these gave better scores or e-values in BLAST than the called coordinates.

Conclusion

Ksed_05030	Gene correctly called; transmembrane protein
Ksed_05040	Gene correctly called; peripheral membrane protein; acetyltransferase
Ksed_05050	Gene correctly called; cytoplasmic protein; hypothetical
Ksed_05060	Gene correctly called; peripheral membrane protein; hypothetical
Ksed_05070	Gene correctly called; transmembrane protein

References

American Heart Association. (2015). Heart Valves and Infective Endocarditis. http://www.heart.org/HEARTORG/Conditions/More/HeartValveProblemsandDisease/Heart-Valves-and-Infective-Endocarditis_UCM_450448_Article.jsp

The Health Science. (2013). Pitted Keratolysis. http://www.thehealthscience.com/showthread.php?842200-Pitted-Keratolysis

Levenga, et al. (2004). Fatal hemorrhagic pneumonia caused by infection due to Kytococcus sedentarius—a pathogen or passenger? *Annals of hematology*, 83: 447-449.

Longshaw, et al. (2002). Kytococcus sedentarius, the organism associated with pitted keratolysis, produces two keratin-degrading enzymes. *Journal of Applied Microbiology*, 93: 810-816

Pospisil, et al. (1998). Katococcus sedentarius (formerly Micrococcus sedentarius) and Dermacoccus nishinomiyaensis (formerly Micrococcus nishinomiyaensis) produce monensins, typical Streptomyces cinnamonensis metabolites. *Canadian Journal of Microbiology*, 44: 1007-1011.

Sims, et al. (2009). Complete genome sequence of Kytococcus sedentarius type strain (541T). *Standards in Genomic Sciences*, 1: 12-20.

Acknowledgments

Thank you to Dr. Rama Dey-Rao, Dr. Stephen Koury, and Dr. Patricia Masso-Welch for their guidance, support, and dedication to this project.

Supported by NSF ITEST Strategies Award Number 1311902