

Manual annotation of Ksed_06890 from *Kytococcus sedentarius* leads to the discovery of a putative membrane protein

Jalese Ragland, Joslynn Scaccia, Megan Burdick, Nyjah Funderburg, Wascar Padilla and Angel Moses

Buffalo Public Schools Career and Technical Education Department, Bennett High School and the Western New York Genetics in Research Partnership

ABSTRACT

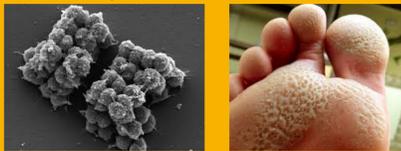
The senior class of Bennett's 2015 Forensic Science Program participated in the gene annotation of *Kytococcus Sedentarius*, using the collaborative genome annotation website GENI-ACT. The Genbank proposed gene product name and function for Ksed_06890 was assessed using basic information, amino acid sequence-based similarity data, structure-based evidence from the amino acid sequence, cellular localization data, potential alternative open reading frames and enzymatic function. Genbank had proposed that this gene coded for a hypothetical protein. Although this gene was not annotated to completion, it is proposed to be a hypothetical membrane protein. Further study of this gene is needed to determine it's function.

INTRODUCTION

For many years Buffalonians have migrated to other cities to seek out a living and pursue their passions. This project allows students to participate in research that can potentially change the look of our future. This is not only important but gives purpose to one's career path. Geneticist study how inherited traits are passed down from one generation to the next generation through genes, and how new traits manifest through mutations. The likelihood of acquiring certain diseases can be passed from parents to children through genes. Although students participating in the Geni-Act project were annotating a bacteria strain named *Kytococcus sedentarius*, they were able to grasp the bigger picture. In order to cure cancer and other diseases, scientist must understand how genes relate and what there functions are. Participating in the annotation of *Kytococcus sedentarius* allowed students to experience this very process.

Kytococcus sedentarius, previously known as *Micrococcus*, is a Gram positive bacteria that has a furry appearance and forms irregularly shaped quadrant clusters. They are found on the skin of humans, and other animals. They are also found in soil, marine and fresh water, plants, dust and air. It is known to be a common cause of foot odor, and causes pitted keratolysis – a skin disorder characterized by crater like indentations of the foot. ¹ It has also been associated with fatal pneumonia, septic arthritis, endocarditis, meningitis and various other diseases.^{2,3} Transmission of this bacteria usually occurs through contaminated surfaces and objects. Person-to-person transmission is not likely.

Our study of *Kytococcus Sedentarius* was carried out via [geni-act.org](http://www.geni-act.org). When the website was not available we used img.jgi.doe.gov/edu. The purpose of our research was to manually annotate the genes assigned to us and determine their function. To initiate this process we determined basic information about the gene such as its coordinates, and protein sequence. We used a protein BLAST search to compare orthologous proteins and draw similarities amongst them. A step by step process was continued using TMHMM, Signal IP, Psort-B and other tools to determine cellular localization as predicted by automated annotation. Based on these resources, we were able to propose that Ksed_06890 which the computer identified as a hypothetical gene as a hypothetical membrane protein.



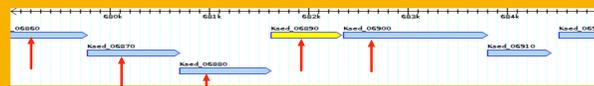
K. sedentarius

Pitted keratolysis

MATERIALS AND METHODS

Modules of the GENI-ACT (<http://www.geni-act.org>) were used to complete *Kytococcus sedentarius* genome annotation. The modules are listed in the table 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?



Five assigned genes: Ksed_06890 is marked in yellow

RESULTS

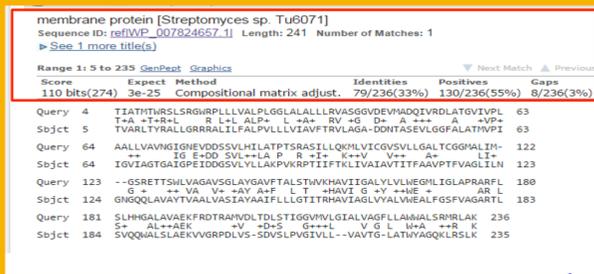


Figure I. The alignment from the first BLAST hit. The search shows 33% identities, and 55% positive alignment of amino acids between the Query Ksed_06890 and a membrane protein from *Streptomyces zinciresistens*.

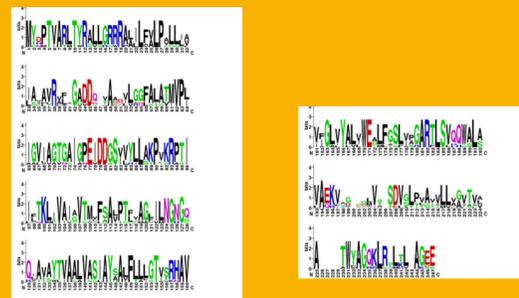


Figure II. The height of the amino acid stacks from N-C terminal of the protein indicates that most amino acids in the protein are well conserved among the ten orthologs chosen for sequence alignment. Specifically, Arginine, Threonine, Leucine and Valine appear with high frequency.

RESULTS

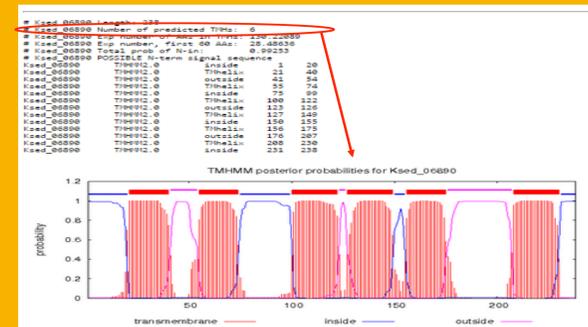


Figure III. There are six trans-membrane helices predicted with high significance by TMHMM that indicates Ksed_06890 to be a membrane protein.

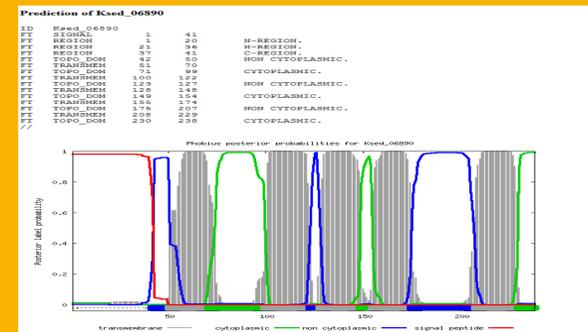


Figure IV. Phobius shows 5 transmembrane helices and a potential signal peptide. This is additional evidence that it is a membrane protein. **Hypothesis** –There are 6 transmembrane helices by TMHMM, and no signal peptide by Signal IP. Phobius shows 5 TMHs and high probability of a signal peptide. Psort-B shows a Score of 10.0 for cytoplasmic membrane. Although there is some ambiguity about the presence of a signal peptide that needs further exploration, this protein is indicated to be a membrane protein with significant probability.

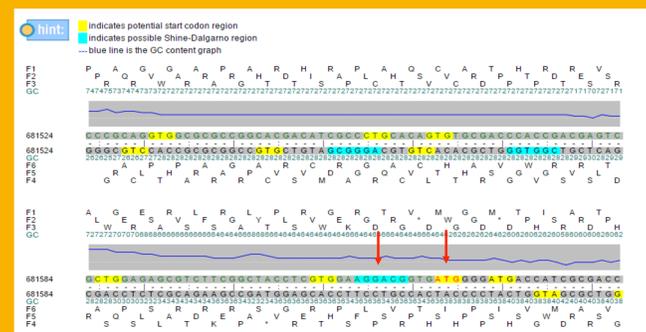


Figure V. We observed a start site ATG as well as the Shine-Delgarno sequence within 5 to 15 base pairs 5 prime to the start site; therefore the DNA coordinates by automated annotation is correct at 681623-682339

CONCLUSIONS

- Kytococcus sedentarius*: Ksed_06890
- While the Genbank proposed gene product for this gene was a hypothetical protein, we propose to change the annotation to a hypothetical membrane protein based on the following evidence:
- BLAST search via the nr database found a **membrane protein** from *Streptomyces zinciresistens* significantly aligned with Ksed_06890.
 - The protein is well conserved.
 - Six trans-membrane helices are observed with high probability by TMHMM and 5 by Phobius with high probability of a signal peptide.
 - The ORF has been correctly called.

Furthermore we propose Ksed_06890 a good candidate gene for continued annotation using bioinformatics tools, to understand structure function relationships.

- Ksed_06860 - Ksed_06900.
- The GENI-ACT proposed gene products for the rest of the genes assigned did not differ significantly from the proposed Genbank annotations.

REFERENCES

- Longshaw, C., Wright, J., Farrell, A., & Holland, K. (n.d.). *Kytococcus sedentarius*, the organism associated with pitted keratolysis, produces two keratin-degrading enzymes. J Appl Microbiol Journal of Applied Microbiology, 810-816.
- Levenga, H., Donnelly, P., Blijlevens, N., Verweij, P., Shirango, H., & Pauw, B. (2003). Fatal hemorrhagic pneumonia caused by infection due to *Kytococcus sedentarius*—a pathogen or passenger? Ann Hematol Annals of Hematology, 447-449.
- Public Health Agency of Canada. (2011). In Best M., Graham M. L., Leitner R., Ouellette M. and Ugwu K. (Eds.), *Laboratory Biosafety Guidelines* (3rd ed.). Canada: Public Health Agency of Canada.

ACKNOWLEDGMENTS

The Career & Technical Education Department and Bennett High School would like to thank Drs. Rama Dey-Rao and Stephen Koury for assisting with the training and implementation of this project.

This work was supported by the National Science Foundation ITEST Strategies Award Number 1311902.