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

Annotation of the *Kytococcus sedentarius* Genome from DNA Coordinates 807009 to 813566

Nicole Cavanaugh, Ananth Das, *Saloni Chadha,* Grace Costello, Kathleen Kramer

Williamsville East High School and the Western New York Genetics in Research Partnership

Abstract

Four consecutive genes (Ksed_07990, Ksed_08010, Ksed_08000, Ksed_7970) from the gram positive microorganism Kytococcus sedentarius from DNA coordinates 807009-813566 were proposed an annotation. These genes were analyzed using the several GENI-ACT modules (including Cellular Localization Data and Sequence Based Similarity Data) in order to discover similarities in the genes' expressions. Their similarities to other genes were tested using Protein BLAST and CDD to determine their possible respective functions. Their similarities were determined by analyzing their E-Values. Each of these four genes were found to reside and function in the cell's membrane.

Introduction

Kytococcus Sedentarius is a marine dwelling, gram-positive bacteria that can act as both an opportunistic pathogen, causing diseases such as pitted keratolysis, seen in figure 1, as well as a potential source of natural antibiotics, such as oligoketide antibiotics. This bacterium is originally isolated from a marine environment and is strictly aerobic. It's optimum temperature is 25-37 celsius, it is free-living, non motile, and is only able to grow in conditions with many amino acids present.

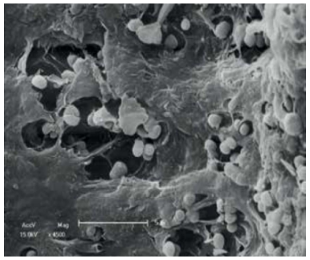


Figure 1: An image of the pitted keratolysis disease caused by Kytococcus sedentarius taken by a scanning electron microscope.

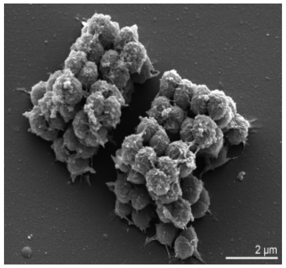


Figure 2: To the right is a photograph of Kytococcus sedentarius under a scanning electron microscope.

Methods and Materials

These modules of GENI-ACT (http://www.geni-act.org/) were used to complete the *Kytococcus sedentarius* gene annotations. 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?
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

Using the modules and databases listed in the results for each individual gene for Kytococcus sedentarius it was determined that all of the genes analyzed reside in or on the cell's membrane. The differences lie in their individual functions.

Kytococcus sedentarius 07970:

The tests supported the conclusion that this protein is located inside the membrane as a Sodium Neurotransmitter Symporter. Information supporting this conclusion was first found with TMHMM, which provided the graph given in Figure 1A, presenting the fact that there is a high probability of the protein coded by this gene touches both inside and outside of the cell membrane, making it a transmembrane protein. Along with this, more supported was gathered using Phobius, in which an output of the graph shown in Figure 1B was produced. This acts as a similar prediction for the localization of Ksed_07970 and it showed that this gene was indeed a transmembrane protein. Finally, PSORT-B provided scores for the different probabilities for different locations of the protein. The highest score, a definite score of 10.0, was provided for the cytoplasmic membrane. This continues to support the fact that the protein is a transmembrane protein.

Both BLAST, TIGRFAM and Pfam concluded that the protein was a Sodium Neurotransmitter Symporter.

Figure1A

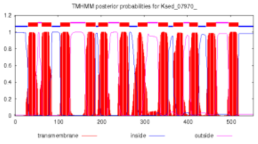
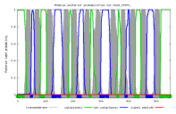


Figure 1B



Kytococcus sedentarius 07990:

In the geni-act module, cellular localization, the TMHMM (Figure 2A) and Phobius tests (Figure 2B) conclude that the protein is a peripheral, non-cytoplasmic protein. The topology graph shows that there are zero predicted transmembrane helices. The signal peptide probability is .242, so it is not n-terminus. The protein is in the revolvase family that has the active site, according to the first hit of the Pfam test in geni-act's structure-based evidence. There were no KEGG pathways found. This supports that the protein is most likely on the outside of the membrane.

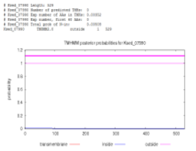


Figure 2A

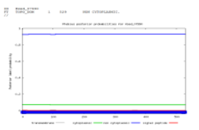


Figure 2B

Kytococcus sedentarius 08000:

Using the module in geni-act, cellular localization, a variety of tests were conducted using PSORT-B, Phobius, and TMHMM in order to determine where in the cell this particular protein, coded by the DNA sequence of Ksed_08000, was located. Each test supported the conclusion that this protein is located outside the membrane as a peripheral protein. Information supporting this conclusion was first found with TMHMM, which provided the graph given in Figure 3A, presenting the probability of the protein being transmembrane as zero, whereas the probability of a peripheral protein located on the outside of the cell was significantly high, approximately 0.98. Along with this, more support was gathered using Phobius, in which an output of the graph shown in Figure 3B was emitted. This acts as another prediction for the localization of Ksed_08000 and it showed that this gene was not cytoplasmic, and its probability of being non cytoplasmic was very high, approximately 0.9, with the probability of it being a transmembrane protein, once again, significantly low. Finally, PSORT-B provided scores for the different probabilities for different locations of the protein. The highest score, 8.16, was provided for the cytoplasmic membrane. This continues to support the fact that the protein is most likely located in the membrane, most likely on the outside.

Figure 3A

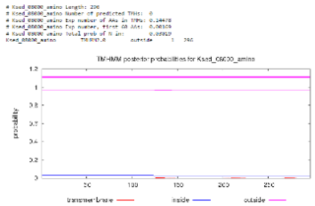
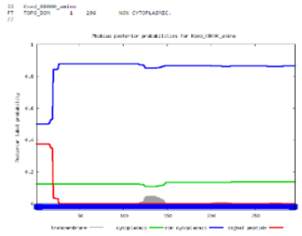


Figure 3B



Kytococus sedentarius 08010:

The initial proposed product of this gene by GENI-ACT is a form of a Signal Peptidase enzyme. This proposal for gene product is supported by using modules such as Cellular Localization Data, as well as Enzymatic Function. As shown in the graph below (from Phobius, Figure 4A), it is most likely that based on the amino acid sequence of Ksed_08010, this gene codes for the production of a transmembrane signal peptidase enzyme. This can be shown by the probability of 1 for signal peptide at the initial lengths of the protein, and how the probability then transfers to 1 numerous times for transmembrane and cytoplasmic regions. The data obtained from SignalP and TMHMM, as well as enzymatic function further provide evidence that this gene sequence may be involved in the coding for a signal peptidase enzyme located on the cellular membrane. Signal peptidases usually function by converting secretory and membrane proteins into their mature structures.

Figure 4A

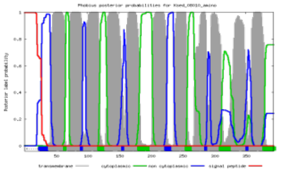
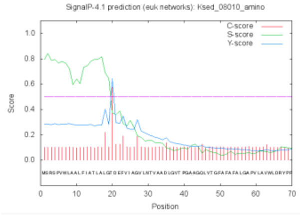


Figure 4B



Conclusion

The proteins that were researched were assigned the same annotations that GENI-ACT recommended. All of the proteins reside on and/or in Kytococcus sedentarius's cell membrane.

Gene	Protein Type	Proposed Annotation
Ksed_07970	Transporter	This protein is a transmembrane sodium neurotransmitter symporter that acts as a channel from one side of the membrane to the other.
Ksed_07990	Peripheral	This protein is a peripheral non-cytoplasmic protein that is part of the revolvase family.
Ksed_08000	Peripheral	This protein is a peripheral protein that is located outside of the membrane.
Ksed_08010	Enzyme	This protein is a signal peptide transmembrane protein.

References

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3035214/

Acknowledgments

The Williamsville East Science and Technology department would like to thank Dr. Rama Dey-Rao and Dr. Stephen Koury for assisting with the formation of this presentation.

Supported by NSF ITEST Strategies Award Number 1311902