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

Annotation of the *Kytococcus sedentarius* Genome from DNA Coordinates 702829 to 712207

Sydney Horschel, Stacy Moppert, Erika Switzer, Jessica Olewine, Jacquelyn Maass, Syler Dawson, and Jennifer Borowicz

West Seneca East Senior High School and the Western New York Genetics in Research Partnership

Abstract

An accumulation of 7 consecutive genes from the microorganism *Kytococcus sedentarius* (Ksed_07040 – Ksed_07130) were ascribed, though only 6 were annotated, using the collaborative genome annotation website GENI-ACT. The Genbank proposed a gene product name for each gene that was assessed. The genes were evaluated in terms of the general genomic information, amino acid sequence-based similarity data, structure-based evidence from the amino acid sequence, cellular localization data, and potential alternative open reading frames. The Genbank proposed gene product name did not differ significantly from the proposed gene annotation for each of the genes in the collection shown here. As such, the genes appear to be correctly annotated by the database based on the modules that were completed.

Introduction

We were annotating the organism Kytococcus sedentarius through the GENI-ACT program. To try to finish the modules efficiently and with the ability to comprehend what we were researching, we met once a week for about an hour. Also, to help us increase efficiency we received print copies of the training manuals and worked at our own pace.

When doing general research about the organism we came across the following information on one of the program websites, NCBI PubMed.gov :

"K. sedentarius produces two extracellular enzymes that independently degrade natural, insoluble human callus. Both enzymes are serine proteases and have cleavage preference sites that are present in a range of human keratins.

The identification, in K. sedentarius cultures, of two enzymes which can degrade human callus strengthens the hypothesis that this organism is responsible for the pitting in human epidermis observed in pitted keratolysis. These enzymes may be of commercial use in the biodegradation of a range of keratin polymers, biological washing powders and in the treatment of unwanted callus on human skin." (Longshaw 2002).

Methods and Materials

Modules of GENI-ACT (http://www.geni-act.org/) were used to complete Kytococcus sedentarius genome annotation. The modules are detailed 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?

Results

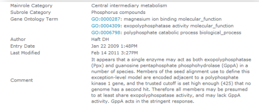
***Kytococcus sedentarius* 07040**:

Kytococcus sedentarius 07040 is related to the protein Enolace. When the amino acid sequence was run in the Blast software. It resulted in another version of the Enolace. It was called phosphopyruvate hydratase, which is a metalloenzyme responsible for the catalysis of the conversion of 2-phosphoglycerate. Enolace is also bacteria which is a key player in metabolism and a probable virulence factor of trypanosomatid parasites

***Kytococcus sedentarius* 07070**:

Based off of the completed modules, one can conclude that the gene is an exopolyphosphatase. Exopolyphosphatase (PPX) is a phosphatase enzyme which catalyzes the hydrolysis of inorganic polyphosphate, a linear molecule composed of up to 1000 or more monomers linked by phosphoanhydride bonds.

This is proven through the BLAST software. The top hit was exopolyphosphatase [Serinicoccus marinus]. In addition to the original search, the Blast website had a new type of blast labeled Smart BLAST. This showed a multitude of results including but not limited to, a phylogenetic tree, a shortened list of the top hits, and a general blast result. The CDD also concurred. In this test the COG name was "Exopolyphosphatase/pppGpp-phosphohydrolase [Nucleotide transport and metabolism, Signal transduction mechanisms, Inorganic ion transport and metabolism]". Lastly, the TIGRFAM results recognizes the gene as an exo_poly_only: exopolyphosphatase.



Mainrole Category	Central intermediary metabolism
Subrole Category	Phosphorus compounds
Gene Ontology Term	GO:0000287: magnesium ion binding molecular_function GO:0004309: exopolyphosphatase activity molecular_function GO:0006798: polyphosphate catabolic process biological_process
Author	Haft DH
Entry Date	Jan 22 2009 1:48PM
Last Modified	Feb 14 2011 3:27PM
Comment	It appears that a single enzyme may act as both exopolyphosphatase (Ppx) and guanosine pentaphosphate phosphohydrolase (GppA) in a number of species. Members of the seed alignment use to define this exception-level model are encoded adjacent to a polyphosphate kinase 1 gene, and the trusted cutoff is set high enough (425) that no genome has a second hit. Therefore all members may be presumed to at least share exopolyphospatase activity, and may lack GppA activity. GppA acts in the stringent response.
Genome Property	GenProp0840: polyphosphate kinase/exopolyphosphatase system (HMM)

Figure I: Explains the job of the Exopolyphosphatase

***Kytococcus sedentarius* 07090:**

In the gram+bacterium, Kytococcus sedentarius, the gene Ksed_07090 occupies the coordinates 706638..707471 and possesses 277 amino acids. During the BLAST test it equated with mostly hypothetical proteins, and one integral membrane protein from the Mobilicoccus organism established in the intestinal tracts of fish. Ksed_07090 does not carry signal peptides, but on the P-SORT B it acquired a 10 for cytoplasmic membrane. Therefore its anticipated location is in the membrane. TMHMM test concurred with these results. This test also exhibited that the protein contains 7 transmembrane helices, symbolizing the protein is constructed inside and outside of the membrane. The tests predominantly alluded towards Ksed_07090 being a gene coding for an integral membrane protein.

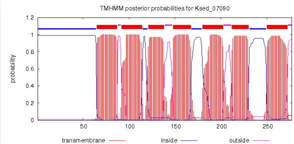


Figure II: Demonstrates the 7 helixes crossing the membrane.

***Kytococcus sedentarius* 07100:**

The results have determined that the protein presides in the cell wall. The protein is a transmembrane, and is non-cytoplasmic. TMHMM predicted a single transmembrane helix at the far right that passes the threshold. The rising helix at the left of the graph doesn't cross the threshold, therefore it's not a transmembrane protein, but is a signal peptide. The protein crosses the membrane once, but is in the extracellular space. The SignalP concurs with the results by predicting there is a signal peptide. In addition, PSORTb agrees due to the cell wall score resulting in 10.0, meaning the predicted position is in the cell wall. Referring to the Phobius graph below, the far right transmembrane is shown and to the far left is the signal peptide. The test results of related proteins came up with multiple types of proteins such as endonuclease, 5' nucleotidase, phosphoesterase. These results demonstrate that the protein's probable function would be to break down bonds by removing a phosphate. Endonucleases cleave phosphodiester bonds in a polynucleotide chain. The 5' nucleotidase cleaves phosphates from 5' nucleotides, and the phosphoesterase breaks phosphodiester bonds. All the suggested protein's functions, and the localization tests show that the protein's probable function is to cleave bonds inside the cell.

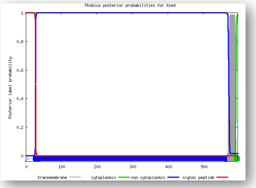


Figure III: The Phobius graph supports protein is inside the cell.

***Kytococcus sedentarius* 07120**:

The gene, Ksed_07120, has two BLAST results that were complementary to it. The genes, 3-ketoacyl-CoA thiolase and Acetyl-CoA acetyltransferase were the top two hits in BLAST. 3-ketoacyl-CoA thiolase has a score of 310 and Acetyl-CoA acetyltransferase scored a 308. The higher the score is the more closely related the gene is to Ksed_07120. TMHMM predicted that Ksed_07120 has no transmembrane helices, leading to the conclusion that the gene codes for a cytoplasmic protein. PSORTb also gave concurring results. The cytoplasmic score is 9.97 [Any score 7.5 or greater is most likely the location of the protein], meaning that it is most likely within the cytoplasm of a cell.

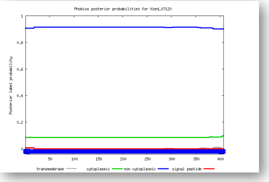
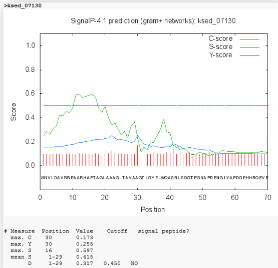


Figure IV: The Phobius shows that the protein is cytoplasmic.

Kytococcus sedentarius 07130:

The protein is located in the cytoplasm of the cell. This is because SignalP, TMHMM, and Phobius show that there are no transmembrane helices predicted. TMHMM and Phobius did not predict the presence of transmembrane helices. For SignalP the signal peptide (S) score is high at the amino terminus for the protein, but without the cleavage (C) score being high at any amino acid position, there is no chance of a signal peptide being present. Thus, the protein is not located in the membrane. The PSORTb shows a cytoplasmic score of 7.5 out of 10. If the score is over 7.5 then that means that it is the final location of the protein. Therefore the protein is located within the cytoplasm.



```
# Measure  Position  Value    Cutoff   signal peptide?
  max. C    30       0.173
  max. Y    30       0.255
  max. S    16       0.597
  mean S     1-29    0.415
       D     1-29    0.317   0.450   NO
Name=ksed_07130 SP='NO' D=0.317 D-cutoff=0.450 Networks=SignalP-TM
```

Figure V: The SignalP graph supports protein is inside the cell.

Conclusion

The table below summarizes the individuals' conclusions based on each their work in GENI-ACT.

Gene Locus Tags	Module Completed *based on Geni-Act Manual	Conclusion
Ksed_07040	Module 2	Predicted to be related to Enolase
Ksed_07070	Module 5	Predicted to be exopolyphosphatase
Ksed_07090	Module 4	Predicted to be coding for integral membrane.
Ksed_07100	Module 5	Predicted to cleave bond inside the cell.
Ksed_07120	Module 4	Predicted to code for a cytoplasmic protein called Thiolase
Ksed_07130	Module 5	Predicted to be located in the cytoplasm of the cell.

References

Longshaw et al. (2002). Kytococcus sedentarius, the organism associated with pitted keratolysis, produces two keratin-degrading enzymes. *Journal of Applied Microbiology,* 93(5):810-6.

Acknowledgments

Special thanks to Dr. Stephen Koury and Dr. Rama Dey-Rao for their assistance with this project. This work was supported by NSF ITEST Strategies Award Number 1311902.