Annotation of the *Kytococcus sedentarius* Genome DNA Locus Tags Ksed_07100, Ksed_07070, Ksed_00740, Ksed_07120, Ksed_00480 and Ksed_00310

Stacy Moppert, Sydney Horschel, Summer Grandinetti, Jason Nowak, Jessica Olewine, Nicholas Dalessandro, Amilleah Rodriguez and Jennifer Borowicz

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

University at Buffalo

Abstract

A selection of 6 genes from the microorganism *Kytococcus sedentarius* 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 group. 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. We met once a week for about an hour. 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 come 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

Modules of 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- 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?
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?

Results

Ksed_00480:

The gene Ksed_00480, in the *Kytococcus sedentarius* organism, codes for a catalase. As seen in the image below, this gene for catalase is very similar to many other catalases of the same kind in varying organisms such as Arthrobacter globiformis or Streptomyces bikiniensis. According to the whole WebLogo, a section of which is shown below, many of the amino acids are conserved between organisms. As seen in the large letters, these genes of the organisms are very similar, with sections being nearly identical at some times. Overall, this gene coding for catalase is close in similarity to other organisms' genes for catalase.

Figure 1 – A section of the WebLogo showing the conservation of amino acids between organisms.

Ksed_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.

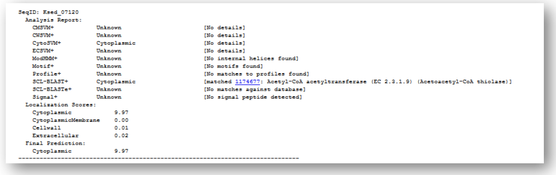

Figure 2 – Ksed_07120 The picture from PSORTb supports the conclusion that the protein thiolase is found in the cytoplasm of the cell

Ksed_00310:

The first hit of the Blast results shows that the most similar DNA segment of another organism codes for the protein, glutaminase, suggesting that this piece of DNA of Kytococcus sedentarius also codes for glutaminase.

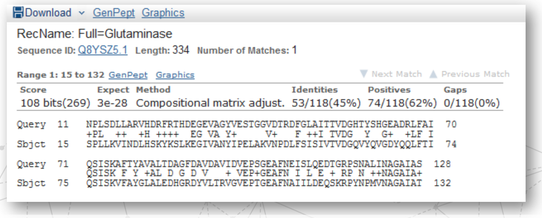

Figure 3 – The BLAST results for the first result with Swiss-Prot.

Ksed_00740:

The gene Ksed_00740 is located at the points 79522..80244. The TMHMM test looks for the locations of the protein that this gene codes for. The protein being specifically looked at here is not a transmembrane protein, which can be seen by the test results of the straight parallel horizontal lines as opposed to clusters of vertical lines. In addition, the specific location cannot be determined which is seen by the test results of the PSORT-B, as they are all under 7.50. From the Signal P test it's shown that the protein produced by gene Ksed_00740 doesn't have any signal peptides suggesting that it is not a neurotransmitter or a hormone/steroid. The BLAST results suggest that this gene sequence codes for peptidase, an enzyme that plays an important role hydrolysis. This is a process that breaks down proteins into smaller parts, which could be the size of peptide chains or as small as amino acids. A peptidase identifies amino acids required for catalysis.

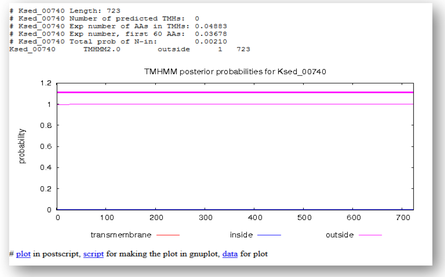

Figure 4 – These TMHMM results show the horizontal lines talked about above. It suggests that it is not a transmembrane protein due to the lack of results.

Ksed_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]". Also, the TIGRFAM results recognizes the gene as an exo_poly_only: exopolyphosphatase. To conclude the computer annotation was correct.

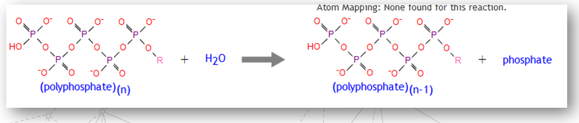

Figure 5 – The chemical pathway shown in MetaCyc.

Ksed_07100

The module results support that the computer annotation was correct. The gene is theorized to code for 5' nucleotidase which catalyzes the cleavage of 5' nucleotides. Nucleotidase works to cleave bonds in 5' nucleotides to allow them to enter the cell. By removing a phosphate a nucleotide becomes a nucleoside which can enter the cell. The protein was said to be transmembrane with a signal peptide based off both the TMHMM results as well as the SignalP which indicated a signal peptide. The PSORT-B indicated that the protein was located within the cell wall and hydrophobic. The localization data indicates that the protein is hydrophobic, based in the cell wall and transmembrane. The EC number for the gene was 3.1.3.5. It corresponded to the meaning of 3. Hydrolases; 3.1 Acting on ester bonds; 3.1.3 Phosphoric-monoester hydrolases; 3.1.3.5 5'-nucleotidase. This would concur with the previous findings of the protein being a 5' nucleotidase which would cleave nucleotides. Based off the evidence it's highly likely that the gene annotation is correct and that the protein was actually a form of 5' nucleotidase that would cleave nucleotides by removing a phosphate. It can be hypothesized that the gene's purpose is to allow nucleotides to enter the cell since they aren't permeable.

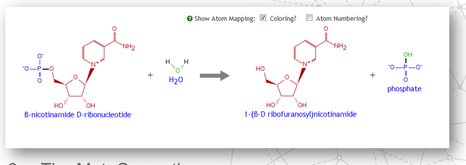

Figure 6 – The MetaCyc pathway map.

Conclusion

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

Gene Locus Tags	Module Completed *based on Geni-Act Manual	Conclusion
Ksed_00480	Module 5	Predicted to be a protein called catalase.
Ksed_07120	Module 4	Predicted to code for a cytoplasmic protein called Thiolase.
Ksed_00310	Started Module 2	Initially predicted to be a protein called glutaminase from BLAST.
Ksed_00740	Module 4	Predicted to be a protein called peptidase.
Ksed_07070	Module 9	Protein is an Exopolyphosphatase, as predicted by the computer annotation.
Ksed_07100	Module 8	Protein is a 5' nucleotidase, as predicted by the computer annotation.

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.

www.buffalo.edu