Annotation of the Kytococcus sedentarius Genome DNA Locus Tags Ksed_00740, Ksed_00480, Ksed_00310, Ksed_00960 and Ksed_00510

Summer Grandinetti, Nicholas Dalessandro, Amilleah Rodriguez, Alex Macchia, Michael Jones, Hunter Jurek and Jennifer Borowicz West Seneca East Senior High School, West Seneca, NY

Abstract

A selection of 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?

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 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.

		Full=Glutaminase <u>Q8YSZ5.1</u> Length: 334 Number of Matches:	:1		
Range :	1: 15 t	o 132 GenPept Graphics	,	Next Match 🔺	Previous Match
Score		Expect Method	Identities	Positives	Gaps
108 bi	its(26	9) 3e-28 Compositional matrix adjust.	53/118(45%)	74/118(62%)	0/118(0%)
Query	11	NPLSDLLARVHDRFRTHDEGEVAGYVESTGGVDTR +PL ++ +H ++++ EG VA Y+ V+	DFGLAITTVDGHT F ++I TVDG		70
Sbjct	15	SPLLKVINDLHSKYKSLKEGIVANYIPELAKVNPD			74
Query	71	QSISKAFTYAVALTDAGFDAVDAVIDVEPSGEAFN QSISK F Y +AL D G D V + VEP+GEAFN	EISLQEDTGRPSN	ALINAGAIAS 1 ++NAGAIA+	28
Sbjct	75	ÖSISKVFAYGLALEDHGRDYVLTRVGVEPTGEAFN			32

Figure 2 – 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.

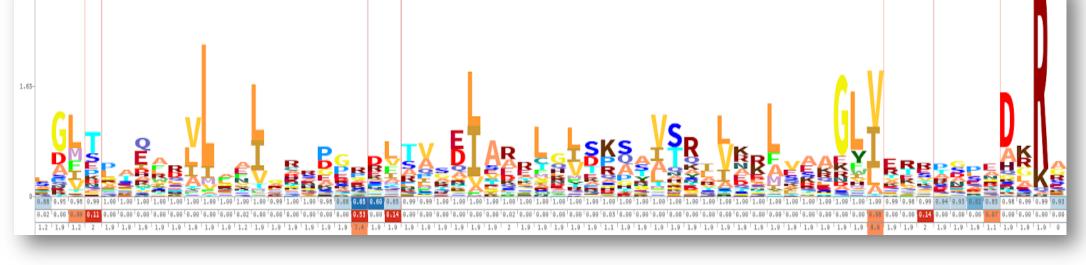
Ksed 00740 (cont): Then when finding a start codon for Ksed 00740 a traditional one cannot be found, so this means there must be an alternative start codon. But when reviewing the data it can be seen that alternate start codons are very far downstream, and therefore probably not a start codon. And when the alternative start codon (and other materials) were put into the BLAST it is seen that it is unrelated to Ksed 00740.

Identities 472 bits(1214) 4e-168 Compositional matrix adjust. 240/240(100%) 240/240(100%) 0/240(0%) MTLGGGGFSETGGKLTPLDHVLLGLAGRRLDAGARGRGERPAVCFVPTASGDAAGYCEKF 60 MTLGGGGFSETGGKLTPLDHVLLGLAGRRLDAGARGRGERPAVCFVPTASGDAAGYCEKF MTLGGGGFSETGGKLTPLDHVLLGLAGRRLDAGARGRGERPAVCFVPTASGDAAGYCEKF 60

Figure 3 – Comparison of BLAST results from original to alternative start codon.

Vext Match 🔺 Previous Match

<u>Ksed 009</u>60: Based on my research Ksed 00960 is most likely a transcriptional regulator. This is based on the fact that my top BLAST hits were both MarR transcriptional regulators. The HMM logo shown is of my top Pfam hit: PF12802.6. The entirety of the section shown was identical to the same section in the logo of the computer's prediction: PF12802. The Pfam proteins were both named MarR 2. This led me to believe that the two proteins were identical or close to being. Since my gene was so close to multiple MarR family transcriptional regulators and PF12802.6 was so close to PF12802 that Ksed 00960 is more than likely a transcriptional regulator.



RecName: Full=Uncharacterized peptidase YgaJ Sequence ID: P71089.2 Length: 230 Number of Matches: 1

Range 1: 5 to 230 GenPept Graphics Next Match A Previous Match Expect Method Identities 167 bits(422) 2e-50 Compositional matrix adjust. 99/241(41%) 136/241(56%) 16/241(6%) SETGGKLTPLDHVLLGLAGRRLDAGARGRGERPAVCFVPTASGDAAGYCEKF + +GGGGFS L+ LD +L + R IAMGGGGFSMEPDNLS-LDQYILNQSKR----E+P +CF+PTASGD+ Y ++F -EOPRICFLPTASGDSONYIORF 5

🖥 Download 🕆 <u>GenPept</u> <u>Graphics</u>

peptidase E [Kytococcus sedentarius DSM 20547] Sequence ID: ACV05169.1 Length: 240 Number of Matches: 1

Range 1: 1 to 240 GenPept Graphics xpect Method

Figure 4 – HMM logo of PF12802.6

Ksed 00510:

The gene Ksed 00510, which comes from the bacteria, Kytococcus sedentarius, codes for an enzyme. The computer predicted protein was "Methyltransferase," which coincided with most of our information. Methyltransferase is a broad category of enzymes, with many classes of them. Class III of the methyltransferases is membrane based. The TMHMM results showed that there was a greater probability that the protein would be excreted rather than residing inside the cell or cell membrane. The Signal-P results showed that there was no signal peptide, but, it is not necessarily going to remain inside the cell because of the lack of a signal peptide.

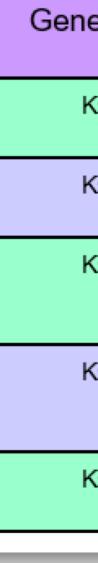
Ksed 00510 (cont): The Phobius results further supported the idea that the protein will be excreted, as it states that it is extracellular. The only outlier was the Psort-B, which stated that the protein was located inside the cell membrane. Our hypothesis as to why this occurred was because there is a class of methyltransferase that is associated with the cell membrane, which may be the reason as to why the results came out as they did.

Ksed 00510 0.6 0.2

Figure 5 – Diagram of the TMHMM results, showing the probability of the protein being inside or outside of the cell. The lack of transmembrane helices in the diagram show that the protein is not in the cell membrane

Conclusion

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



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 for his assistance with this project. Supported by an NIH Science Education Partnership (SEPA) Award - R250D010536



Ksed 00510 Length: 175 # Ksed 00510 Number of predicted TMHs: # Ksed 00510 Exp number of AAs in TMHs: 3.74893 # Ksed 00510 Exp number, first 60 AAs: 1.97744 # Ksed 00510 Total prob of N-in: 1 175 TMHMM2.0

TMHMM posterior probabilities for Ksed_00510

	-		-		-		-	-	
			_						
_									-
									-
	20	40	60	80	100	120	140	160	
	here a second			1	a la la		a stalida		
	transme	mbrane		Ins	side ——		outside –		

e Locus Tags	Module Completed *based on Geni-Act Manual	Conclusion
Ksed_00740	Module 4	Predicted to be a protein called peptidase.
Ksed_00480	Module 5	Predicted to be a protein called catalase.
Ksed_00310	Started Module 2	Initially predicted to be a protein called glutaminase from BLAST.
Ksed_00960	Module 3	Predicted to be a transcriptional regulator protein
Ksed_00510	Module 5	Predicted to be a protein called methyltransferase.

www.buffalo.edu