Annotation of the *Kytococcus sedentarius* Genome DNA Locus Tags Ksed_02660, Ksed_00940, Ksed_03530 and Ksed_00070

Hunter Schoenhardt, Aries Sledge, Mary Cooper, George Foser, Emma Frank and Alexandria LaPorta and Jennifer Borowicz
West Seneca East Senior High School, West Seneca, NY

SEPA SCIENCE EDUCATION PARTNERSHIP AWARD
Supported by the National Institutes of Health

University at Buffalo

Abstract

A selection of 5 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_00070:

We investigated Locus Tag Ksed_00070 that has a gene ID of 644990323. When performing BLAST the protein was identified by the computer as DNA gyrase subunit A. This gene has coordinates of 8710..11337 in the forward direction. While conducting our research, our data has supported that the the computer prediction is accurate. When conducting WebLogo, the picture in Figure 1 supports that the larger letters represent the amino acids similar to the protein that we are coding for. Thus concludes that the proteins must have a function comparing to others.

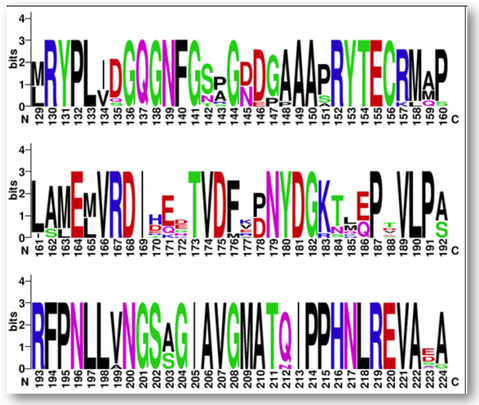

Figure 1 – A portion of the WebLogo showing the similarities between amino acids in different organisms.

Ksed_02660:

Based off our research our protein, Ksed_02660, coded for an enzyme named glutathione peroxidase.The information that the computer called for was supported by our research. The BLAST, TIGRFAM, Pfam, and enzyme numbers matches the computer's original results. Our E.C. number was 1.11.1.9, if it was not an enzyme it would not have had an E.C. number. Our enzyme, is a catalyst to make glutathione disulfide and two water molecules, from two glutathione molecules and a hydrogen peroxide molecule. The main purpose of the our enzyme glutathione peroxidase is to convert hydrogen peroxide into water to neutralize the harmful effects of the peroxide, and protecting an organism from getting any type of oxidative damage. Glutathione is also known as one of the most potent natural antioxidants. As an antioxidant, glutathione gives up an electron and gives it to an unpaired electron known as a free radical to complete the pair. Free radicals are connected to causing cancer, Alzheimer's, Parkinson's disease, the effects of aging, and other illnesses because of the effect they have on the cell when they take the electron away.

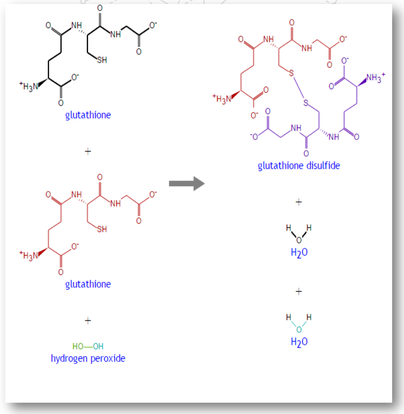

Figure 2 – MetaCyc image of reaction catalyzed by the enzyme glutathione peroxidase.

Ksed_00940:

The computer predicted that the bacterial protein is called cytosine/adenosine deaminase. Though my Blast Generator gene product says that the bacterial protein is called nucleoside deaminase. Even though the names are different, I noticed many similarities that could link the different bacteria names together. For instance, in Module 1, the computer generated the gene's DNA coordinates as 99499..100026(-). When I generated my data it came up as the same. Another example being that when the computer generator caught the Pfam name as Cytidine and deoxycytidylate deaminase zinc-binding region for my gene, while when I generated the Pfam name, I got the same result. Some other results that are the same by comparing my information to the computer's information are: Pfam Number:(PF00383), Helix Presence: None and Signal Peptide Presence: None.

Wondering what Nucleoside meant and what made it different from the name that the computer generated for my bacteria, I looked up the definition; Nucleoside: a compound (e.g., adenosine or cytidine) commonly found in DNA or RNA, consisting of a purine or pyrimidine base linked to a sugar. Purine bases are the chemicals Cytosine and Adenine that make up a nucleotide in DNA and RNA. The same goes for pyrimidine.

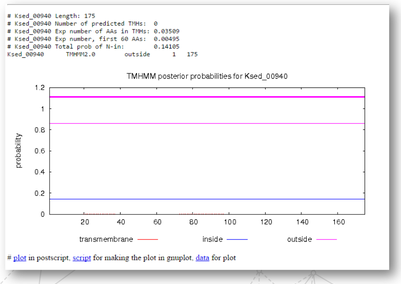

Figure 3 – The TMHMM graph shows that there is no helix present in the bacterial protein.

Ksed_03530:

In the organism *Kytococcus sedentarius*, with the gene Ksed_03530 the computer is predicting the gene codes for a protein called polyphosphate kinase which my research is supporting as shown with the information below. This is shown being supported because the BLAST results are shown below, the top hit's name is similar to the result that the computer gave me.

Figure 4 – Top hit of BLAST results shown above for Ksed_03530.

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_02660	Module 6	Predicted to be a protein called glutathione peroxidase
Ksed_00940	Module 4	Predicted to be a protein called cytosine/adenosine deaminase
Ksed_03530	Module 4	Predicted to be a protein called polyphosphate kinase
Ksed_00070	Module 3	Predicted to be DNA gyrase

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.

Smirnova, G. V., and O. N. Oktyabrsky. "Glutathione in Bacteria." *Biochemistry (Moscow)*, vol. 70, no. 11, 2005, pp. 1-2. *Springer Link*, link.springer.com/article/10.1007/s10541-005-0248-3. Accessed 12 May 2018.

Acknowledgments

Special thanks to Dr. Stephen Koury for his assistance with this project. Supported by an NIH Science Education Partnership (SEPA) Award - R25ODO10536

www.buffalo.edu