Annotation of the *Kytococcus sedentarius* Genome from DNA Coordinates 764347 to 765843 (Locus Tag Ksed_07600)

Helen Huang, Madison Kaifas-Steiner, Nora Hassan, Jenna Papaj and Robert O'Connor
Kenmore East High School and Genetics in Research Partnership

Abstract

Given the genome Ksed_07600 we annotated this genome relative to the results of certain tests done in order to define the function. These tests include, BLAST, CDD, T-Coffee, WebLogo, NCBI, TMHMM, SignalP, PSORT-B, Phobius, TIGRFAM, Pfam, PDB, KEGG, MetaCyc, ExPASy enzyme, Phylogeny. We predicted that it would share very similar characteristics to a dehydrogenase based on our first blast results. From these sites we obtained information such as the amino acid sequence and nucleotide sequence by blasting our gene and using this to compare results to other known genomes. Using these sequences, we could then gather more information such as the location of our gene in order to confirm that this strain of *Kytococcus sedentarius* primarily shares qualities and characteristics of a dehydrogenase. Dehydrogenase is an enzyme which catalyzes reactions taking place in the electron transport chain during cellular respiration. We can then deduce that Ksed_07600 functions in this was as well. It was a constant cycle of comparing and contrasting information in order to learn more about the qualities that our genome exhibits in order to confirm our original hypothesis that Ksed_07600 is in fact a dehydrogenase.

Introduction

Kytococcus sedentarius is a gram positive bacterium. It is of interest as it has the potential to become a natural source of antibiotics. It also plays a role as an opportunistic pathogen and can cause severe and potentially fatal illness such as severe pneumonia, pitted keratolysis and valve endocarditis (Sigs,2009). It has been found to be resistant to most forms of antibiotics such as penicillin. While it is relatively isolated from human skin, found primarily in marine environments, it can still cause strains of disease in humans.It can infect the body through chemotherapy treatment and then go on to cause such infections ("Annals of Hematology", 2004).

It has also been isolated from such environments as ground water (Sigs,2009). It is aerobic and requires an organic source of carbon and metabolic energy and requires amino acids for growth(BACMAP).

A genome is an organism's complete set of DNA, including all of its genes. Each genome contains all of the information needed to build and maintain that organism (science.gov)). Geni-act allows collaborative gene annotation. Through this website, we were able to go through several tests to suggest information about Ksed_07600. These results can then be reported back to the Gene bank for further investigation. Using the lab notebook and certain given websites, we were able to come to a consensus on its functionality and purpose.

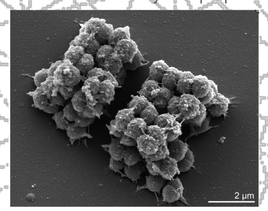


Figure 1. An image of *K. sedentarius* under a scanning electron microscope.

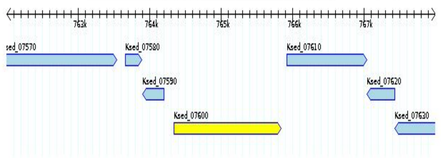


Figure 2. The gene neighborhood of Ksed_076900. Our gene studied is highlighted in yellow.

Methods

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

Modules	Activities	Investigation
Module 1- Basic Information Module	DNA Coordinates and Sequence, Amino Acid Sequence.	Where is this gene located? What is its amino acid sequence and what is the sequence's length?
Module 2- Sequence-Based Similarity Data	Gene Product name, organism, alignment, score, E-value BLAST, CDD, T-Coffee, WebLogo	How does this gene compare to similar genes in Genibank? What does the alignment of our gene look like? What does this tell us?
Module 3- Cellular Localization Data	Gram Strain, TMHMM, SignalP, PSORT-B, Phobius,	Where can we expect to find this protein?
Module 4- Alternative Open Reading Frame	DNA Coordinates, IMG/EDU	Is there evidence of a Shine-Dalgarno Sequence?
Module 5- Structure-Based Evidence	TIGRFAM, Pfam, PDB	Is our protein functionally similar to to other known proteins?
Module 6- Enzymatic Function	KEGG, MetaCyc, E.C. Number	In what pathway does the protein function?
Module 7- Gene Duplication/Gene Degeneration	Pseudogene	Is the gene a pseudogene? Does it have any paralogs?
Module 8- Evidence for Horizontal Gene Transfer	Phylogenetic Tree	Has the gene come from horizontal transfer?

Results

Ksed_07600:

The initial proposed product of the gene by GENI-ACT was an Inosine-5'-monophosphate dehydrogenase. It has 1,497 nucleotides and 498 amino acids in the sequence. This gene product proposal was supported by the results from the BLAST hits of our amino acid sequence. The result of the top hit, Inosine-5'-monophosphate dehydrogenase, catalyzes the conversion of inosine 5'-phosphate (IMP) to xanthosine 5'-phosphate (XMP). Blast of amino acid sequence in the database of non-redundant protein sequences yielded an E-value of 0 suggesting that the top hit was in fact our gene product. The gene comes from the organism Staphylococcus aureus. There were no predicted transmembrane helices. There may be evidence of a shine-dalgarno sequence upstream, but when plugged into IMG for an alternate reading frame, no sequence resulted.

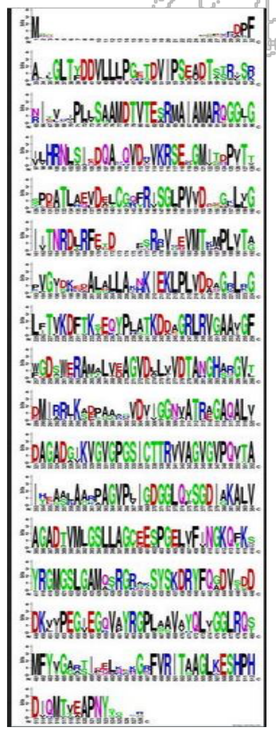


Figure 3. The web logo for Ksed_07600. This allows a much easier interpretation of the alignment. The most common amino acids in each sequence are used in the alignment represented in graphical format. The relative size and length of the letters represents the frequency of that specific amino acid. Taller the letter, more frequently appeared. We can see here that our web logo shows that the alignments are highly conserved because many of the lines have tall wide single letters and there are not a lot of empty spaces. Many of the amino acids are polar amino acids and there are many that are hydrophobic.

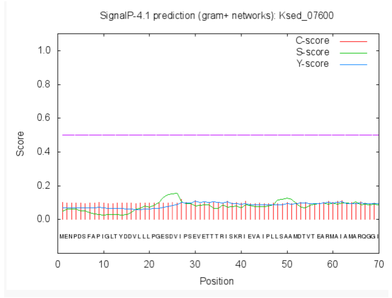


Figure 4. SignalP output. The signal peptide probability must reach a threshold of 0.450 to be significant. The signalP output indicates that no signal peptide is present.

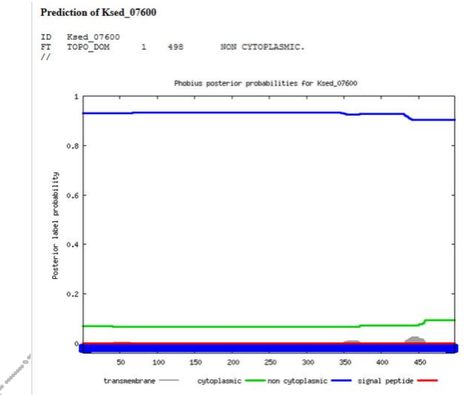


Figure 5. Phobius prediction. We concluded that the location of the protein is most likely cytoplasmic.

Figure 6 is the pathway map found using the KEGG program. The E.C. number found then was # 1.1.1.205. From this we could use ExPASy ENZYME to find the EC Name. Our Results included **Inosine 5'-monophosphate dehydrogenase,** our first hit BLAST hit and our most similar result. This furthers our conclusion and hypothesis.

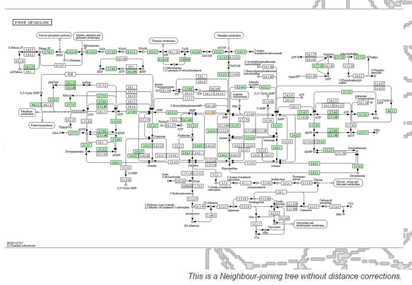
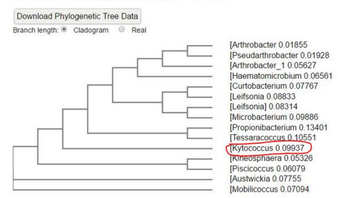


Figure 7 depicts the Phylogenetic Tree. Our gene has large differences in evolutionary time among genera.



Conclusion

TMHMM indicates that there are no transmembrane helices. SignalP reveals low probability of a signal peptide. Many of the amino acids are polar amino acids and there are many that are hydrophobic. PSORT-B indicates that the location of the protein is most likely cytoplasmic and Phobius confirms these results. Our hypothesis is that this protein is located in the cytoplasm. It is likely a type of dehydrogenase, most closely associated to Inosine-5'-monophosphate dehydrogenase. This information is supported by the research done using the different databases and can be seen in the figures on this poster.

References

Website. bacmap.wishartlab.com, bacmap.wishartlab.com/organisms/953.
Website. ghr.nlm.nih.gov, ghr.nlm.nih.gov/primer/hgp/genome.
Website. link.springer.com, link.springer.com/article/10.1007/s00277-003-0831-x.
Website. standardsingenomics.biomedcentral.com, standardsingenomics.biomedcentral.com/articles/10.4056/sigs.761.
Website. www.science.gov, www.science.gov/topicpages/k/kytococcus+sedentarius+type.html.

Acknowledgments

This work was supported by the National Science Foundation ITEST Strategies Award Number 1311902. Imperative contributors also include Dr. Rama Dey-Rao and Dr. Stephen Koury.

www.buffalo.edu