

Annotation of 3 Consecutive *Kytococcus sedentarius* Genes: *Ksed_05130*, *05140*, and *05150*

Kyla Threats, Mary Stevens, Eli Swing, and Margaret Diamond

The Park School of Buffalo and the Western New York Genetics in Research Partnership

Abstracts

Three consecutive genes from the microorganism *Kytococcus sedentarius* (*Ksed_05130* – *05150*; DNA bases 507695 - 512851) were annotated using the collaborative genome annotation website GENI-ACT. The GENI-ACT proposed gene product identification for each gene was examined in terms of the general genomic information, amino acid sequence-based similarity data (BLAST, CDD, T-Coffee, and WebLogo), structure-based evidence from the amino acid sequence (TIGRFam and Pfam), and cellular localization data (TMHMM, SignalP, and Phobius). The Genbank proposed gene product name did not differ significantly from the proposed gene annotation for each of the three genes studied. Therefore, the genes appear to have been correctly identified by the computer program.

Introduction

Kytococcus sedentarius is a gram positive bacterium that produces certain antibiotics. It is aerobic and requires some amino acids for growth. This bacterium was originally found in a marine environment. Even though it is often overlooked it is known to have caused serious illnesses such as valve endocarditis and pneumonia. It also causes foot odor and pitting of foot calluses (pitted keratolysis). (Sims 2009; James 2012)

Ksed_05130 is a hypothetical protein; as such, it has no known function. BLAST showed that this protein is most similar to a hypothetical protein identified in many other bacterial species. In addition to BLAST, other programs (as listed in Materials and Methods) were used to analyze and annotate the gene *Ksed_05130*. Even though its function is unknown, it was found to be very highly conserved.

Ksed_05140 was identified as N-formylglutamate amidohydrolase (Fgase), a protein involved with amino acid metabolism. It was found to be most similar to a protein sequence from *Ornithinimicrobium pekingense* with a high degree of similarity to many other bacterial protein sequences as shown by the E- values. There is only one identified domain which is in the FGase superfamily.

Using Blast, *Ksed_05150* was found to be most similar to a gene from *Bacillus subtilis subsp. subtilis str. 168*. The protein product is a Na⁽⁺⁾/H⁽⁺⁾ antiporter and a transmembrane protein. Transmembrane proteins function on both sides of a membrane bilayer as they transport molecules across it. Thus, the protein is intertwined in the membrane. There is also another identified domain which is a DUF, or Domain of Unknown Function. Both domains have Pfam hits with significant E-values and the proton antiporter domain had a significant COG hit.

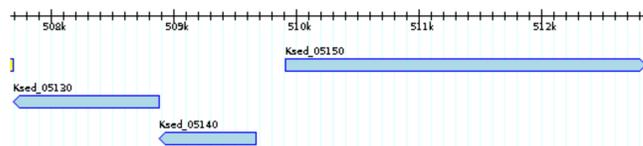


Figure 1. Gene neighborhood of the three genes of interest.

Methods and Materials

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	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

Kytococcus sedentarius 05130:

Comparison to the non-redundant protein database using protein BLAST showed that *Ksed_05130* is most similar to other identified hypothetical proteins and that it has a e value of 3e-172. This is a very high value which means that the compared protein sequences are very similar. WebLogo also illustrated that it was highly conserved throughout. The cellular location of *Ksed_05130* is also uncertain. It most likely is in the cytoplasm since no transmembrane domains or signal sequence were identified, but its cellular location is, as yet, unclear.

Kytococcus sedentarius 05140:

Comparison of the amino acid sequence to the non-redundant protein database using BLAST showed a high degree of similarity to proteins found in many other bacterial species. It contained only a single well-curated functional domain, identified as the enzyme formate dehydrogenase, which is an enzyme involved with the metabolism of the amino acid histidine.

Kytococcus sedentarius 05150:

The product proposed by GENI-ACT for this gene was that it was a (Na⁽⁺⁾)/H⁽⁺⁾ antiporter. This was supported and confirmed by the protein BLAST results. This gene codes for a transmembrane protein and has hydrophilic and hydrophobic domains which gives it the ability to pass in and out of the membrane. It is responsible for moving sodium and hydrogen ions in and out of the cell and plays a role in pH balance and homeostasis. The COG or Clusters of Orthologous number for this protein is high because it has been highly conserved. It has a second domain which is of unknown function

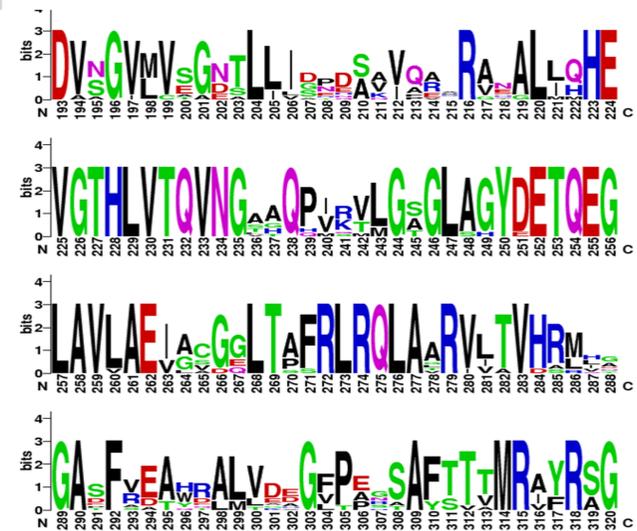


Fig. I This is a section of the WebLogo for *Ksed_05130* taken from the middle of the protein. Each letter represents an amino acid and the larger the letter the greater the conservation between sequences from different organisms. It shows a very high level of conservation which is true for almost the entire sequence.

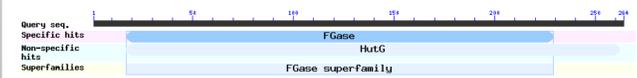


Fig. II This graph shows the conserved domain for the *Ksed_05130* gene product. It is in the FGase (N-formylglutamate amidohydrolase) superfamily

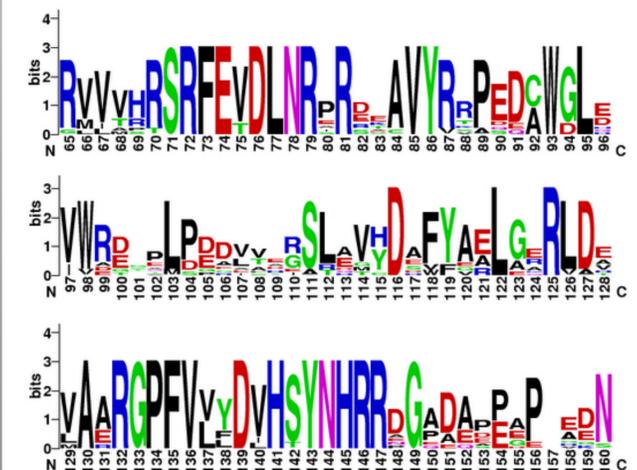


Fig. III This is a representative segment of the WebLogo for the *Ksed_05140* protein sequence showing the high degree of similarity between it and similar identified sequences of other bacterial species.

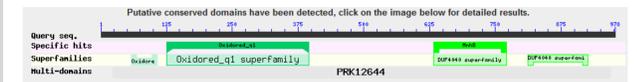


Fig. IV – *Kytococcus sedentarius* gene 05150 has two domains, one of which is well-characterized as a proton antiporter; the other is a DUF or domain of unknown function.

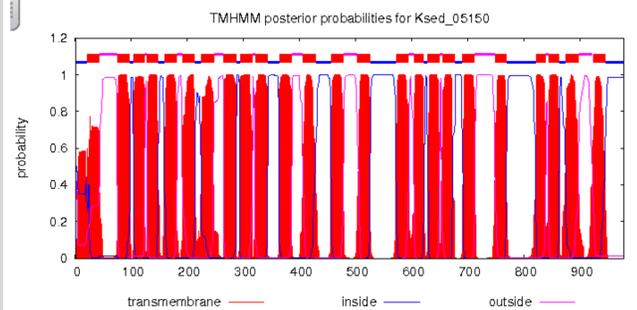


Fig V. This graph shows the TMHMM results for *Ksed_05150* protein. According to the key; one would be able to see that the red shows if the protein is embedded in the transmembrane. There are 23 transmembrane domains shown in this graph. This shows that *Ksed_05150* encodes a transmembrane protein. The blue and pink lines indicate the portions of the protein that loop inside and outside of the membrane.

Conclusion

The GENI-ACT proposed gene products did not differ significantly from the proposed gene annotation for each of the genes analyzed and as such, the genes appear to have been correctly identified by the computer program as follows:

Gene Locus	Geni-Act Product	Proposed Annotation
05130	Hypothetical protein	Hypothetical protein
05140	N-formylglutamate amidohydrolase	N-formylglutamate amidohydrolase
05150	Na ⁽⁺⁾ /H ⁽⁺⁾ antiporter	Na ⁽⁺⁾ /H ⁽⁺⁾ antiporter

References

James et al. (2013). Microbiological and Biochemical origins of foot malodour. *Flavour and Fragrance Journal*, 28: 231-237
 Sims et al. (2009). Complete genome sequence of *Kytococcus sedentarius* type strain (541T). *Standards in Genomic Sciences*, 12 - 20.

Acknowledgments

Supported by NSF ITEST Strategies Award Number 1311902
 Special thanks to Dr. Stephen Koury and Dr. Rama Dey-Rao for their help in completing the work presented here.