University at Buffalo *The State University of New York* | REACHING OTHERS

Annotation of the *Kytococcus sedentarius* Genome from Ksed_08980 to Ksed_09010

Nicia Bottini, Allyssa Finley, Jenna Papaj, Anna Taylor, Robert O'Connor

Kenmore East High School and the Western New York Genetics in Research Partnership

Abstract

A group of consecutive 4 genes from the microorganism *Kytococcus sedentarius* (Ksed_08980 – Ksed_09010) were annotated using the collaborative genome annotation website GENI-ACT. The Genbank proposed a gene product name for each gene, assessed in terms of the general genomic information, amino acid sequence-based similarity data, structure-based evidence from the amino acid sequence, and cellular localization data. We used bioinformatics databases such as Blast, TIGRfam, PDB and Phobius in order to research our genes and confirm our proposed hypotheses. The annotations of our assigned genes led us to the conclusion that the Genbank's proposed gene product names for each gene were accurate.

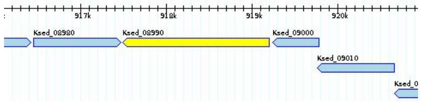
Introduction

Kytococcus sedentarius was first discovered in 1944 and since then still has not been well studied. It is the first member of the Dermacoccaceae family to have its genome completely sequenced.

According to Sims et al. (2009), *Kytococcus sedentarius* is a microorganism of interest for several reasons. This bacterium is a natural source of the oligoketide antibiotics monensin A and monensin B (Sims et al., 2009). *Kytococcus sedentarius* has been implemented as the etiological agent of a number of opportunistic infections including valve endocarditis, hemorrhagic pneumonia, and pitted keratolysis (Sims et al., 2009). Finally, the phylogeny of this microorganism is a source of interest, as it is a member of the family *Dermacoccaceae* within the actinobacterial suborder *Micrococcineae,* which has yet to have been thoroughly studied utilizing bioinformatics (Sims et al., 2009).

As this was our first experience sequencing a genome, we took this opportunity to enrich our knowledge of biology and what the study of genomics really entails. We collaborated on a challenging process by combining our talents to overcome new obstacles.
It was an honor to be a part of this unique experience and have the opportunity to sequence our own assigned genes that have never been manually sequenced previously.

Figure I: Pitted keratolysis, a bacterial infection of the foot. (The Health Science, 2013)



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- Structure-Based Evidence	TIGRfam, Pfam, PDB	Are there functional domains in my protein?
Module 4- Cellular Localization Data	Gram Stain, TMHMM, SignalP, PSORT, Phobius	Is my protein in the cytoplasm, secreted or embedded in the membrane?
Final Annotation	Review data from all modules	Does the student proposed name of the gene agree with that proposed by the automated computer annotation? Are any changes proposed to the pipeline annotation?

Results

*Kytococcus sedentarius*08980:

Genbank proposed that Ksed-08980 was a 6-phosphofructo kinase, which is a kinase enzyme that phosphorylates fructose 6-phosphate in glycolysis. Throughout different annotations, the conclusion was reached that this was an accurate proposal. Localization revealed the gene to be a cytoplasmic protein, Tcoffee showed that it was very conversed throughout, and all the Pfam results lined up with the proposal. Therefore, the original proposal was proved to be correct.

*Kytococcus sedentarius*08990:

Glycosidases assist in the hydrolysis of glycosidic bonds in complex sugars; they are extremely common enzymes with roles in nature including degradation of biomass such as cellulose and hemicellulose, in anti-bacterial defense strategies, in pathogenesis mechanisms and in normal cellular function. According to the data collected, my gene did in fact turn out to be Glcosidase, for it followed all of the previous information.

*Kytococcus sedentarius*09000:

The Genbank proposed that Ksed-09000 was a truncated hemoglobin, which is a shortened or cut-down version of a hemoglobin molecule that is involved in the processes of oxygen transportation and metabolism. In the annotations of this gene, the evidence supports the conclusion that it is in fact a truncated hemoglobin. The cellular localization revealed that it is a cytoplasmic membrane protein. All other results in the modules reinforced the original proposal.

*Kytococcus sedentarius*09010:

The evidence supported the hypothesis that Ksed-09010 is a mechano-sensitive ion channel, which is a protein embedded in the membrane with three trans-membrane helices according to TMHMM. Although SignalP reveals low probability for signal peptide, pSORTB predicts that the protein is in the cytoplasmic membrane and Phobius confirms positive results for TMHs and signal peptide. Despite a slight disagreement, the final conclusion confirms a membrane-bound mechano-sensitive ion channel.

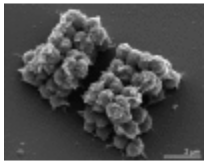


Figure II – *Kytococcus sedentarius as seen under a scanning electron micrograph*

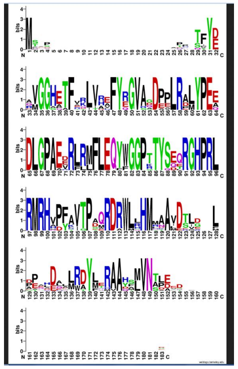
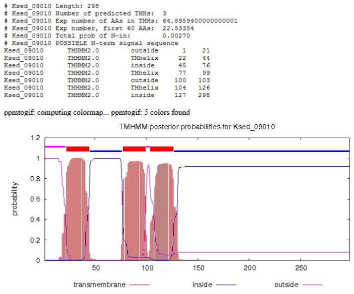


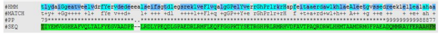
Figure III – Web logo of *Kytococcus sedentarius gene 09000*

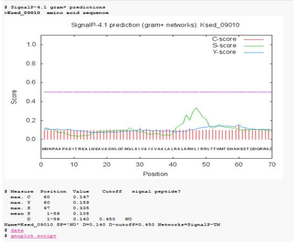


```
# Ksed_09010 Length: 298
# Ksed_09010 Number of predicted TMHs:  3
# Ksed_09010 Exp number of AAs in TMHs: 64.8959400000000001
# Ksed_09010 Exp number, first 60 AAs:  22.53354
# Ksed_09010 Total prob of N-in:        0.00270
# Ksed_09010 POSSIBLE N-term signal sequence
Ksed_09010      TMHMM2.0        outside      1    21
Ksed_09010      TMHMM2.0        TMhelix     22    44
Ksed_09010      TMHMM2.0        inside      45    76
Ksed_09010      TMHMM2.0        TMhelix     77    99
Ksed_09010      TMHMM2.0        outside    100   103
Ksed_09010      TMHMM2.0        TMhelix    104   126
Ksed_09010      TMHMM2.0        inside     127   298

ppmtogif: computing colormap... ppmtogif: 5 colors found
```

Figure IV— Signal Peptide graph of Ksed-09010, showing that the gene contains three trans-membrane helices because it crosses the trans-membrane threshold three times on the graph.





```
# SignalP-4.1 gram+ predictions
>Ksed_09010  amino acid sequence
```

```
# Measure  Position  Value    Cutoff   signal peptide?
  max. C    60       0.147
  max. Y    60       0.159
  max. S    47       0.335
  mean S     1-59    0.108
       D     1-59    0.140   0.450   NO
Name=Ksed_09010 SP='NO' D=0.140 D-cutoff=0.450 Networks=SignalP-TM
# data
# gnuplot script
```

Figure V: T-Coffee Results from Ksed_09000
Figure VI: Pairwise Alignment from Ksed_08990
Figure VII: SignalP Prediction from Ksed_09010

Conclusion

The GENI-ACT proposed gene product did not differ significantly from the proposed gene annotation for each of the genes in the group and as such, the genes appear to be correctly annotated by the computer database.

Ksed_08980 – Gene correctly proposed; 6-phosofructokinase

Ksed_08990 – Gene correctly proposed; Glycosidase

Ksed_09000 – Gene correctly proposed; truncated hemoglobin protein

Ksed_09010 – Gene correctly proposed; Mechano-Sensitive Ion Channel

References

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, Dr. Rama Dey-Rao and Dr. Patricia Masso-Welch for their help in completing the work presented here.