Annotation of the *Kytococcus sedentarius* Genome at Locus Tags Ksed_00335, Ksed_14140, and Ksed_14195

Cayden Piegza, Owen Boccabella, Wyatt Krieger, Ely Personius and Dr. Paul A. Weber
Amherst Central High School, Amherst, NY and the Western New York Genetics in Research and Health Care Partnership

Abstract

A group of 3 genes from the microorganism *Kytococcus sedentarius* (Ksed_00335, Ksed_14140, and Ksed_14195) were annotated using the collaborative genome annotation website GENI-ACT. The GenBank proposed gene product name for each gene was assessed in terms of the general genomic information, amino acid sequence-based similarity data, structure-based evidence from the amino acid sequence, cellular localization data, potential alternative open reading frames, enzymatic function, presence or absence of gene duplication and degradation, the possibility of horizontal gene transfer, and the production of an RNA product. The GenBank proposed gene product name 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 in the r database.

Introduction

Kytococcus sedentarius is a strictly aerobic, non-motile, non-encapsulated, and non-endospore forming gram positive coccoid bacterium, found predominantly in tetrad formation. This organism is classified as a chemoheterotroph, as it requires methionine and several other amino acids for growth. Originally isolated from a microscope slide submerged in sea water in 1944, *Kytococcus sedentarius* grows well in sodium chloride at concentrations less than 10% (w/v).

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).
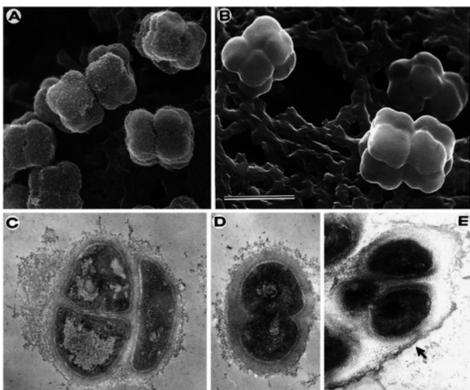
Figure 1. Scanning (A and B) and Transmission (C-E) electron micrographs of *Kytococcus sedentarius*.

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	Questions Investigated
Basic Information	DNA Coordinates and Sequence, Protein Sequence	What is the sequence of the gene and protein? Where is it located in the genome?
Sequence-Based Similarity	Blast, CDD, T-Coffee, WebLogo	How similar is the protein under investigation to other proteins in GenBank?
Structure-Based Similarity	TIGRfam, Pfam, PDB	What functional domains are present in the protein under investigation?
Cellular Localization	Gram Stain, TMHMM, SignalP, LipoP, Psortb, Phobius	Is the protein under investigation located in the cytoplasm, secreted, in the periplasm or embedded in the cell membrane or cell wall?
Enzymatic Function	KEGG, MetaCyc, E.C. Number	In what process or structure is the protein under investigation involved?
Duplication and Degradation	Paralog, Pseudogene	Are there other forms of the protein under investigation in the same genome? Is it functional?
Horizontal Gene Transfer	Phylogenetic Tree, Gene Neighborhood, GC Content	Has the protein under investigation co-evolved with the rest of the genome or has it been obtained in a different way?
RNA family	Rfam	Does the gene under investigation encode a functional RNA?
Final Annotation	Evaluate data from all modules	Has the gene been correctly called by the pipeline annotation?

Results

Ksed_00335

The initial proposed product of this gene by GENI-ACT was a Hemin Transporter. This gene product proposal was supported by the top BLAST hits for the amino acid sequence, as the related proteins all were listed as Hemin Transporters as well. The following two figures (Figure 2 and Figure 3) depict genes from related organisms, as well as the protein products of some such and other organisms; all of these are also Hemin Transporters. This is strong evidence that the product of this gene is therefore also of the same product.

Figure 2 – Ksed_00335: Shows related genes from other bacterium.

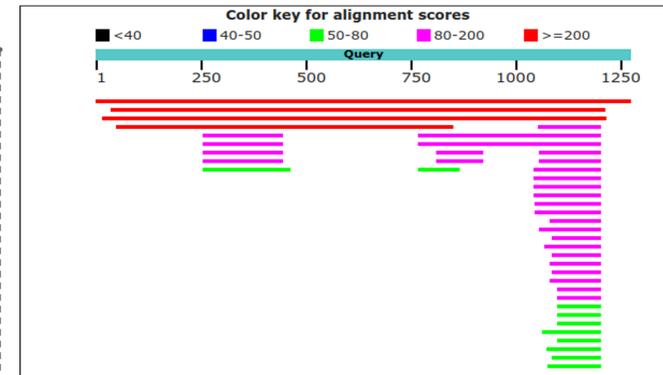**Color key for alignment scores**

Figure 3 – Ksed_00335: Shows related proteins from other bacteria.

Ksed_14140

The initial proposed product of this gene by GENI-ACT was a Disorder Prediction Protein. This gene product proposal was supported by the top BLAST hits for the amino acid sequence, the presence of well-curated protein functional domains within the amino acid sequence, TIGRFAMs, Pfams, and the Protein Data Bank all agree that this protein may have the aforementioned function. It is also indicated that the protein is excreted from the cell. As such, the proposed annotation is a Disorder Prediction Protein. It must be noted that many modules produced null data for this gene, resulting in uncertainty as to the true product of this gene, yet due to the rather limited evidence given, the original proposed product remains.

Figure 4 – Ksed_04140: This chart shows how similar Ksed_04140 is with other genes in different bacterium.
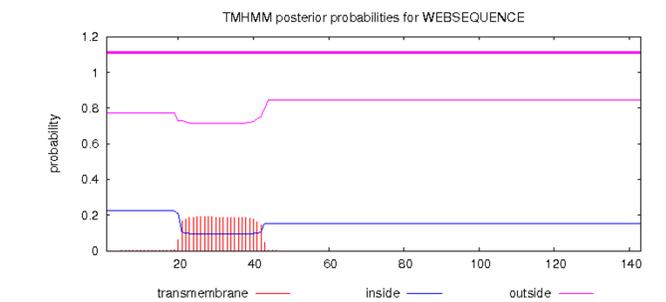
Ksed_14195:

The predicted product of this gene was a protease, supported by the top BLAST hits, the presence of conserved functional domains, and its location outside the membrane. The proposed function is a protease. The following figure (Figure 5) depicts the transmembrane hits of the gene product protein, which provides evidence that the protein may cross the cell membrane and therefore exist outside of the cell. Being that the proposed product is a protease, the implication that this enzyme is extracellular is intriguing because it may thusly take part in the flesh digestion for which *Kytococcus sedentarius* is known.

TMHMM result

HELP with output formats

```
# WEBSEQUENCE Length: 143
# WEBSEQUENCE Number of predicted TMHs: 0
# WEBSEQUENCE Exp number of AAs in TMHs: 4.17921
# WEBSEQUENCE Exp number, first 60 AAs: 4.17871
# WEBSEQUENCE Total prob of N-in: 0.22703
WEBSEQUENCE    TMHMM2.0    outside    1    143
```

TMHMM posterior probabilities for WEBSEQUENCE

plot in postscript, script for making the plot in gnuplot, data for plot

Figure 5 – Ksed_14195: Shows the location of the gene's product to be outside the membrane.

Conclusion

The GENI-ACT proposed gene product did not differ 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.

Locus Tag	Pipeline Annotation Product Name	Proposed Annotation	Changes Proposed?
Ksed_00335	Hemin Transporter	Hemin Transporter	No
Ksed_14140	Disorder Prediction Protein	Disorder Prediction Protein* *based on limited data	No
Ksed_14195	Protease	Protease	No

References

Sims et al. (2009). Complete genome sequence of *Kytococcus sedentarius* type strain (541T). Standards Genomic Sciences,12 - 20.