Annotation of the *Kytococcus sedentarius* Genome from Locus Tags Ksed_07570 to Ksed_07590

Ehsan Rafique, Fatimah Salman, Hannah Schumacher, and Omer Bitar

Lackawanna High School, and the Western New York Genetics in Research Partnership

Abstract

The investigation of three genes (Ksed_07570, Ksed_07580, and Ksed_07590) from the marine bacteria, Kytococcus sedentarius, have been manually annotated with the assistance of the GENI-ACT (Genomics Educational National Initiative – Annotation Collaboration Toolkit). The Genbank database uses an automatic method of predicting gene presence and function. Manual annotation is used to confirm, reject, or reassign the name and function of the investigated gene. Annotation is achieved by comparing proteins of other organisms by examining similarities in amino acid sequences (sequence based similarity data), structure based evidence, and cellular localization data. In most cases, the annotated genes did not differ from the Genbank predicted function.

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).

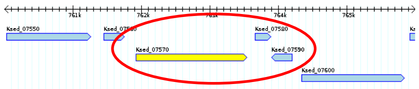


Figure I. Gene neighborhood from IMG-edu. The genes studied are in the red oval

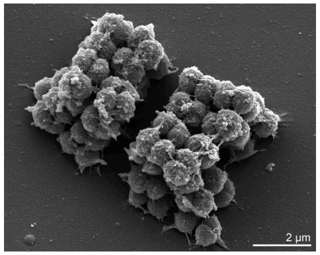


Figure II. Scanning electron Micrograph of *Kytococcus sedentarius* (Sims et al., 2009)

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?

Results

Kytococcus sedentarius 07570:

The initial proposed product of this gene by GENI-ACT was a **molecular chaperone GroEL**. This gene product proposal was supported by the top BLAST hits for the amino acid sequence, the presence of well-curated functional domains within the amino acid sequence, the cellular location of the amino acid sequence, and the enzymatic function of the amino acid sequence. As such, the proposed annotation molecular chaperone GroEL.

Figure III. WebLogo of T-Coffee Multiple Sequences Alignment output for Ksed_07480 showing regions of high conservation of amino acid sequences.

Kytococcus sedentarius 07580:

The initial proposed product of this gene by GENI-ACT was **RNA-binding protein**. This gene product proposal was supported by the top BLAST hits for the amino acid sequence, the presence of well-curated functional domains within the amino acid sequence, the cellular location of the amino acid sequence, and the enzymatic function of the amino acid sequence. As such, the proposed annotation is RNA-binding protein.

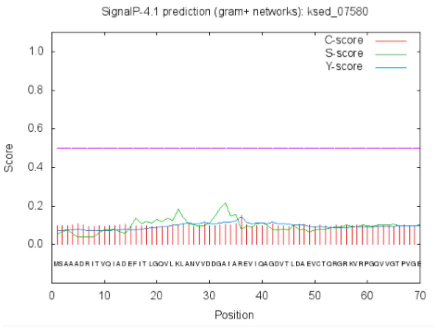
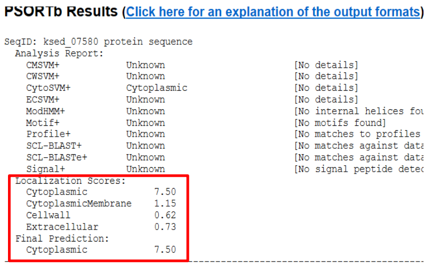


Figure IV. SignalP output displaying the absence of signal peptide in Ksed_07580.



PSORTb Results (Click here for an explanation of the output formats)

```
SeqID: ksed_07580 protein sequence
  Analysis Report:
    CMSVM+              Unknown                          [No details]
    CWSVM+              Unknown                          [No details]
    CytoSVM+            Cytoplasmic                      [No details]
    ECSVM+              Unknown                          [No details]
    ModHMM+             Unknown                          [No internal helices fou
    Motif+              Unknown                          [No motifs found]
    Profile+            Unknown                          [No matches to profiles
    SCL-BLAST+          Unknown                          [No matches against data
    SCL-BLASTe+         Unknown                          [No matches against data
    Signal+             Unknown                          [No signal peptide detec
  Localization Scores:
    Cytoplasmic              7.50
    CytoplasmicMembrane      1.15
    Cellwall                 0.62
    Extracellular            0.73
  Final Prediction:
    Cytoplasmic              7.50
---------------------------------------------------------------------------------
```

Figure V. PSORTb output displaying predicted location of protein

Kytococcus sedentarius 07590:

The initial proposed product of this gene by GENI-ACT was **transcriptional regulator**. This gene product proposal was supported by the top BLAST hits for the amino acid sequence, the presence of well-curated functional domains within the amino acid sequence, the cellular location of the amino acid sequence, and the enzymatic function of the amino acid sequence. As such, the proposed annotation is transcriptional regulator.

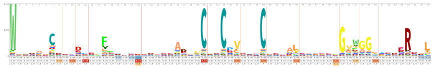


Figure V. HMM Logo output of Ksed_07590.

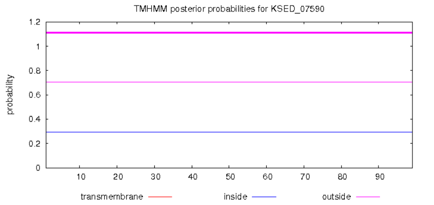


Figure VI. TMHMM output displaying the absence of a trans-membrane protein in Ksed_07590.

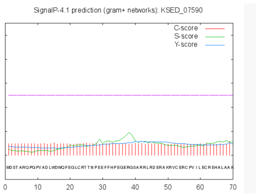


Figure VII. SignalP output displaying the absence of signal peptide in Ksed_07590.

Conclusion

Except for Ksed_07450, 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.

Locus Tag	BLAST Prediction	Proposed Annotation
Ksed_07570	molecular chaperone GroEL	molecular chaperone GroEL
Ksed_07580	RNA-binding protein	RNA-binding protein
Ksed_07590	transcriptional regulator	transcriptional regulator

References

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

Acknowledgments

Many thanks to Dr. Stephen Koury, Dr. Rama Dey-Rao, and Danise Wilson for their guidance, support and dedication to this project. Supported by NSF ITEST Strategies Award Number 1311902