Annotation of the *Kytococcus sedentarius* Genome at Locus Tags Ksed_00480, Ksed_00350,Ksed_00460 and Ksed_00780

Macey Chasman, Jenna Connolly, Jack Fiorille, Heather Higgins, Hanna Holland, Asia Hoyte, Timothy Mariani, Claire Martin, Aliantha Palka, Jillian Quevedo, Nathaniel Rumrill and Caitlin Ullock

Pittsford Mendon High School, Pittsford NY and The Western New York Genetics in Research Partnership

University at Buffalo

Abstract

A group of 4 genes from the microorganism *Kytococcus sedentarius* (Ksed_00480, Ksed_00350, Ksed_00460, Ksed_00780) were annotated using the collaborative genome annotation website GENI-ACT. The Genbank proposed gene productname 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, and enzymatic function. 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).

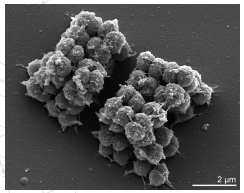

Figure 1. A scanning electron micrograph of Kytococcus sedentarius

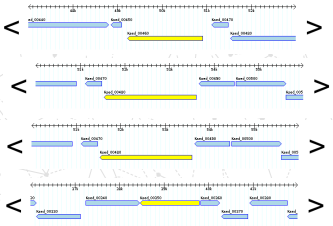

Figure 2. The gene neighborhoods of the genes under investigation.

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
Module 1- Basic Information	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 Sequence	Has the amino acid sequence of my protein been called correctly by the computer?
Module 5- Structure-Based Evidence	TIGRfam, Pfam, PDB	Are there any functional domains in my protein?
Module 6- Enzymatic Activity	Enzymatic Pathway,	What is the function of my gene product?

Results

Ksed_00480

The annotation of the gene under investigation is correct as called by the computer. The name should be Catalase HPII because we found that name in our results for Blast. That was the accepted name of the enzyme as well. The start codon remained the same so the coordinates remained the same. This protein is cytoplasmic and integral to the membrane, so it could be located in the cytoplasm.

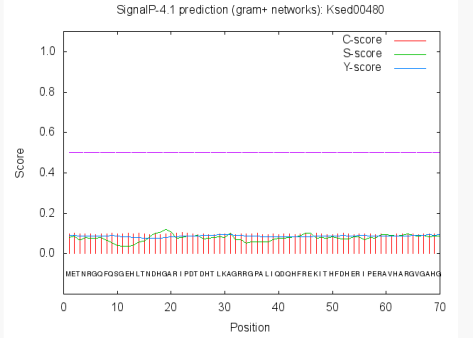

Figure 2. The signalP output for Ksed_00480 demonstrating the lack of a predicted signal peptide

Ksed_00780

The annotation of the gene under investigation is correct as called by the computer. The name should be Dihydrodipicolinate synthase because we found that name in our results for COG, TIGRFAM, PFAM, PDB, and E.C. That was the accepted name of the enzyme as well. The start codon remained the same so the coordinates remained the same. This protein is cytoplasmic and not integral to the membrane, according to PsortB. It is used in Lysine and Monobactam biosynthesis pathway according to KEGG

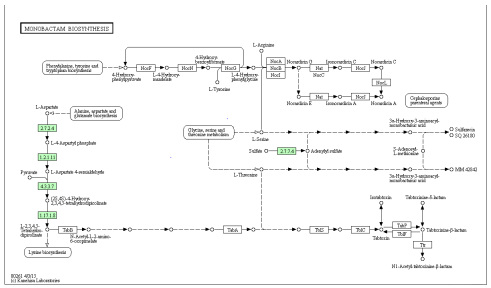

Figure 3. The monobactam biosynthesis pathway in which Ksed_00780 functions. It also is functions in the lysine biosynthesis pathway (not shown)

Ksed_00460

Our gene was a hypothetical protein. This explains why the computer calculated no signal peptide protein, no COG value, and limited hits except for the 93% match BLAST hit of HNH endonuclease [*Intrasporangium calvum* DSM 43043] This protein was gram-positive and non-cytoplasmic because the rest of the tools said unknown except the phobias probability graph indicated this. A Shine-Dalgarno sequence was found at start of the protein sequence. Because this was found 5-10 genes before the start codon, it is likely that this sequence was called correctly. This protein was very limited because it was hypothetical meaning that the function is not completely known yet.

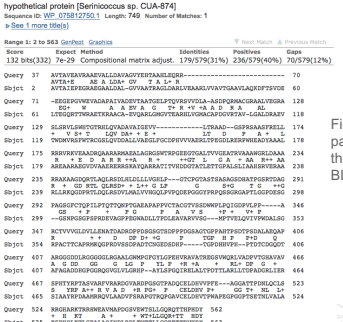

Figure 4. The pairwise alignment of the top Ksed_00460 BLAST hit

Ksed_00350

The annotation of the gene under investigation is correct as called by the computer. The name should be Prolyl Aminopeptidase, as found in analysis results for COG, TIGRFAM, PFAM, PDB, and E.C. That was the accepted name of the enzyme as well. The start codon remained the same so the coordinates remained the same. This gene is cytoplasmic and not integral to the membrane, so its product is most likely found in the cytoplasm of the bacterium. This can be seen in the following PSORT-B results.

```
Localization Scores:
  Cytoplasmic                9.67
  CytoplasmicMembrane        0.01
  Cellwall                   0.15
  Extracellular              0.17
Final Prediction:
  Cytoplasmic                9.67
```

Figure 5. The psortB output for Ksed_00350 isIllustrating the highly significant prediction of a cytoplasmic location.

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.

Gene Locus	Geni-Act Products	Proposed Annotation
Ksed_00480	catalase	Catalase HPII
Ksed_00780	Dihydrodipicolinate synthase	Dihydrodipicolinate synthase
Ksed_00460	HNH endonuclease	Hypothetical/HNH endonuclease
Ksed_00350	prolyl aminopeptidase 2	Prolyl Aminopeptidase

Reference

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

Acknowledgments

Supported by an NSF Innovative Technology Experiences for Students and Teachers (ITEST) Award - 1311902

www.buffalo.edu