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

Excellence Today and Tomorrow Pittsford Schools

Annotation of the *Kytococcus sedentarius* Genome at Locus tags Ksed_07740, Ksed_07780, Ksed_07850 and Ksed_07860

Tristan Bachmann, Zach Rose, Ava Roncone, Courtney Calamel and Caitlin Ullock
Pittsford Mendon High School and the Western New York Genetics in Research Partnership

Abstract

A group of 4 genes from the microorganism *Kytococcus sedentarius* (Ksed_07740, Ksed_07780, Ksed_07850, Ksed_07860) 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, cellular localization data, alternative open reading frame and the structure-based evidence from the amino acid sequence. 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). The genome has been sequenced and contains 2,785,024 bp.

Fig. 1. The 4 locus tags (in yellow) under investigation in this research

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?

Results

Kytococcus sedentarius Ksed_07850:

The initial proposed product by GENI-ACT was a Tryptophan TRNA Ligase. Psort-b predicted this protein is found in the cytoplasm. TMHMM and Phobius did not find transmembrane helixes (meaning not likely to be in cytoplasm membrane). SignalP and Phobius did not find signal peptides on the protein (meaning not likely to be secreted to the outside of the bacterium). Therefore it is most likely found in the cytoplasm. The graph below shows the start codon (red) has a Shine-Dalgarno (SD) sequence upstream. Looking back at the T-COFFEE results suggested there might be some codons missing compared to other proteins in the alignment. I tested an upstream start codon that also had an SD sequence (arrow) and found it gave a BLAST hit with a slighly higher score and lower E-value, suggesting the start codon may have been called incorrectly. I found the protein is an enzyme used as a catalyst in the chemical reaction ATP + L-tryptophan + tRNATrp AMP + diphosphate + L-tryptophyl-tRNATrp

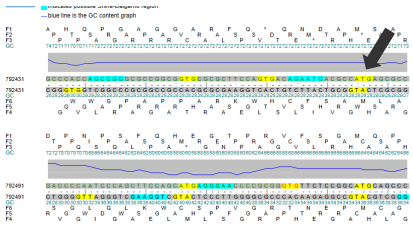

Fig. 2. Alternate Open Reading Frame Module results for Ksed_07850. The start codon indicated by the arrow gave resulted in a protein having a higher score and lower E-value compare to the original start codon.

Kytococcus sedentarius Ksed_07780:

The initial proposed product of this gene by GENI-ACT was a photosystem reaction center subunit H. The top BLAST hit supported this finding (Figure 3). PSORT-B predicted this protein is located inside the cytoplasm, with a final prediction of 7.50, which was further supported by the results of SignalP, TMHMM, as well as Phobius. SignalP predicted that there is no signal peptide. TMHMM predicted that this protein has no transmembrane helixes. The Phobius results supported the results of SignalP and TMHMM, and all three results, when put together, prove that this membrane has to be cytoplasmic. The web logo shows that the beginning and end of the sequence are relatively conserved while the middle is not very

photosystem reaction center subunit H [Arthrobacter sp. M2012083]
Sequence ID: refWP_017198868.1| Length: 276 Number of Matches: 1

Range 1: 1 to 268 GenPept Graphics

Score	Expect	Method	Identities	Positives	Gaps
209 bits(532)	4e-62	Compositional matrix adjust.	137/282(49%)	167/282(59%)	24/282(8%)

```
Query  1    MISKDQLRTLR--DGDLYSVDGDKVGSIEEIYLDDRTNEPEWATVNTGLPGTSQSFVPLS  58
            M++++ +  L    +G++ + +GDKVGSI    Y DD T EP W TV TGLPG+S+SF+PL
Sbjct  1    MLARENIDRLMGLNGNVRTTEGDKVGSIGTFYTDDDTGEPTWVTVKTGLPGSSESFIPLQ  60

Query  59   GATVTEDGLTVQYSKDQIKDAPRVDAGEHLGVEQEQELYRYYGVDGQ----GEGVATDRT  114
             ATV  D + V +SKD IKDAPRV     HL   +E  LY +YG++G+    G G
Sbjct  61   DATVEGDDVVVVPFSKDHIKDAPRVAEDGHLDPGEESRLYNHYGLEGRDSFSGHG------  114

Query  115  DVAAGTAAAGTAGAAGHVGDRDERAVADRDIADRDRGVADRDADATMVRHEEHLNVGTER  174
             D AA T   GT GAAG  G RD          D D G  D DA   M R EE LNVGTER
Sbjct  115  DYAATT---GTVGAAGTAGTRDAVDTDANYTRDTDAGRGD-DA---MTRSEEQLNVGTER  167

Query  175  QETGRARLRKYVDEEEQTVSVPVETESAHVVRRPVD----GEVVSDGKAFQEQEVEMDLE  230
             Q TGRARLRKY+  E   T +VPV+ E    + R P+      G  +S G    E E E+ L
Sbjct  168  QATGRARLRKYITTENVTTTVPVQREEVYLRERPITDANAGNAMS-GPDLTESEHEVTLH  226

Query  231  SERPVVSKETVAVEEVGLEKETHTENVQVSDSVRSERIEVEG  272
             SERPVV KETV VE V LEK+T T+ VQV + VR ERIE EG
Sbjct  227  SERPVVEKETVPVERVRLEKDTVTDEVQVDEEVRKERIETEG  268
```

Figure 3. The top nr BLAST hit for Ksed_07780.

Kytococcus sedentarius Ksed_07860:

The initial proposed product of this gene by GENI-ACT was an RNA Ligase enzyme that is used in the sugar and phosphate bonds that are used in replication. This gene product proposal was supported by the top BLAST hits for the amino acid sequences, the presence of well-curated functional domains within the amino acid sequences, and the cellular location of the amino acid sequences. Based on the results found by the various sources, it can be concluded that Ksed_07860 is found in the cytoplasm of a cell, due to its score being 7.50. This gene is not very conserved in the first half of the sequence, but is well conserved in the last half.

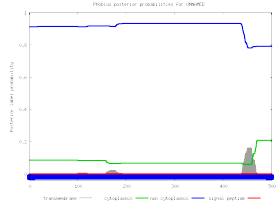

Fig. 4. The Phobius results for Ksed_07860 showing a lack of transmembrane helixes and a signal peptide, consistent with its being found in the cytoplasm.

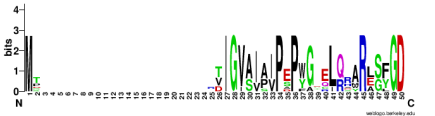

Fig. 5. The amino terminal portion of the WebLogo for Ksed_07860

Kytococcus sedentarius Ksed_07740:

The initial product proposed by GENI-ACT was a CoA Succinate Ligase subunit alpha. PSORT-B predicted that the protein would not be found in the cytoplasm, while TMHMM and Phobius searches did not yield any transmembrane helices, therefore it is unlikely to be found in the cytoplasm membrane. PSORT could not determine the cellular localization. Since the enzyme functions in the citrate cycle, the protein is likely to be found in the cytoplasm. This is indicative of an excretory protein, one which is produced within the bacterium and then released.

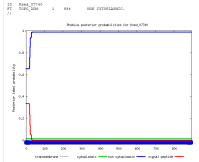

Fig. 6 Phobius results for Ksed_07740 showing a lack of a signal peptide or transmembrane helixes.

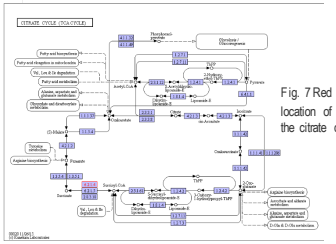

Fig. 7 Red indicates the location of Ksed_07740 in the citrate cycle.

Conclusions

The GENI-ACT proposed gene product did not differ significantly from the proposed gene annotation for three of the genes in the group. Ksed_07850 was originally annotated as an uncharacterized conserved protein, but our data suggest that it might more correctly be annotated as photosystem reaction center protein H. Further work will need to be done to confirm that finding.

Gene Locus	Geni-Act Gene Products	Proposed Annotation
07740	succinyl-CoA synthetase (ADP-forming) alpha subunit	Succinyl-CoA synthetase, alpha subunit
07780	uncharacterized conserved protein	photosystem reaction center protein H
07850	tryptophanyl-tRNA synthetase	tryptophanyl-tRNA synthetase
07860	2'-5' RNA ligase	2'-5' RNA ligase

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