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

Excellence Today and Tomorrow
Pittsford Schools

Annotation of *Kytococcus sedentarius* Genome at Locus Tags Ksed_07800, Ksed_07820 and Ksed_07870

Anna Pogharian, Bianka Smeulders, Reith Mantena and Caitlin Ulock
Pittsford Mendon High School and the Western New York Genetics in Research Partnership

Abstract

A group of three genes from the microorganism *Kytococcus sedentarius* (Ksed_07800, Ksed_07820 and Ksed_07870) were annotated using the collaborative genome annotation website GENI-ACT. The Genbankproposed 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, and potential alternative open reading frames. The Genbankproposed 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 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).

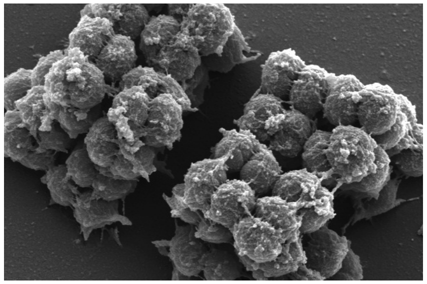


Figure 1. A scanning electron micrograph of *Kytococcus sedentarius* (Taken from Sims et al, 2009)

Methods and Materials

Modules of the GENI-ACT (http://www.geni-act.org/) were used to complete the *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?
Module 5- 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?
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*07820:

The initial proposed product of this gene by GENI-ACT was a bifunctional 5,10-methylene-tetrahydrofolate dehydrogenase/5,10-methylene-tetrahydrofolate cyclohydrolase, an enzyme that functions in the cytoplasm, not the cell membrane. 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, the sequence logo, and the cellular location of the amino acid sequence. Because of this, the proposed annotation is a bifunctional 5,10-methylene-tetrahydrofolate dehydrogenase/5,10-methylene-tetrahydrofolate cyclohydrolase.

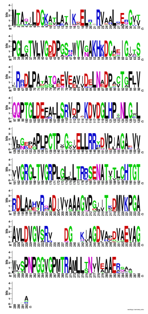
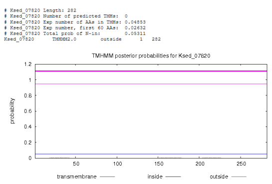


Figure 2. Weblogo of The Ksed_07820 T-COFFEE alignment showing high amino acid conservation among aligned proteins.



```
# Ksed_07820 Length: 282
# Ksed_07820 Number of predicted TMHs:  0
# Ksed_07820 Exp number of AAs in TMHs: 0.04853
# Ksed_07820 Exp number, first 60 AAs:  0.02632
# Ksed_07820 Total prob of N-in:        0.05311
Ksed_07820      TMHMM2.0        outside      1   282
```

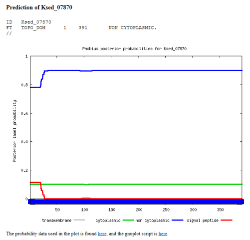
Figure 3. The Ksed_07820 TMHMM result showing the lack of transmembrane helixes in the protein.

*Kytococcus sedentarius*07870:

The initial proposed product of this gene by GENI-ACT was a glycosyl transferase. This gene was predicted to be functional outside of the cell membrane, which was supported by the Phobius Probability Graph. This gene product proposal was supported by the top BLAST hits for the amino acid sequence in the CDD page, the presence of well-curated functional domains within the amino acid sequence, and the cellular location of the amino acid sequence. The enzymatic function of this gene was found from the amino acid sequence, which is to catalyze nucleotide sugars to an acceptor molecule. Supported by this information, the proposed annotation is a glycosyl transferase.

```
glycosyl transferase [Serinicoccus marinus]
Sequence ID: ref|WP_033418860.1|  Length: 384  Number of Matches: 1

Range 1: 1 to 383  GenPept  Graphics                    Next Match  Previous Match
Score          Expect  Method                        Identities     Positives      Gaps
523 bits(1347) 0.0     Compositional matrix adjust.  269/387(70%)   301/387(77%)   4/387(1%)

Query  4    VRVAIVTESFLPSLNGVTTSVCKVLECLADKGHEAAVVCPRSSTLAPEDQVPTTYAGFPV  63
            +RVAIVTESFLP+LNGVTTSVCKVLECL ++GH+A VV P +S  +P    P  YAGFPV
Sbjct  1    MRVAIVTESFLPALNGVTTSVCKVLECLREQGHDALVVAPGTSPWSPV-MTPEHYAGFPV  59

Query  64   HPVASVPVRKFRVGMPGYDTEAVFRGFKPDVAHVASPFVLGVNALSVARQLRVPTVAIYQ  123
            H V SVPVRKFRVG+P Y+ E V   F+PDV HVASPFVLGV  L  AR L +P+VAIYQ
Sbjct  60   HSVTSVPVRKFRVGLPSYEIETVLHRFRPDVMHVASPFVLGVRGLVAARALGIPSVAIYQ  119

Query  124  TDMPSYIRQHSGPAGHLAANATWRWIRRIHAQADLTLAPSQAALTDLREHDIPRTALWGR  183
            TDMPSYIRQH+GPAG L A ATWRWIRRIH QADLTLAPS AALTDL EH +PR ALWGR
Sbjct  120  TDMPSYIRQHAGPAGDLTARATWRWIRRIHEQADLTLAPSTAALTDLAEHQVPRIALWGR  179

Query  184  GVDAKLFHPDRKQTDAVRRLKAELSPRGEVLLGYVGRLAPEKELHRFLELADLPGARLVF  243
            GVDA LFHPDR+        L+  L+PRGE +LGYVGRLAPEKELHR  EL+ LPG RLV
Sbjct  180  GVDADLFHPDRRTDPGALALRDRLAPRGETVLGYVGRLAPEKELHRLTELSSLPGTRLVL  239

Query  244  VGDGPSRDLLGHQFPTAAFLGRMEGEELADAYAAFDLFVHAGTRETFGQTLQEGAAAGLP  303
            VG+GPSR++L  Q P A FLGR EG ELA AYAAFDLFVH GTRETFGQTLQE +AAGLP
Sbjct  240  VGEGPSREILQAQLPEAVFLGRREGAELAQAYAAFDLFVHTGTRETFGQTLQEASAAGLP  299

Query  304  VVAPARGGPIDLIDHGGTGYLFDPDRPGDLRACVEKLVLGPDAAAERERMGRAGRTAVEG  363
            VVAPARGGP+DLI  GGTG LFDPD  G L A V  LV GP AA  RERMG AGR  V+
Sbjct  300  VVAPARGGPLDLIQPGGTGELFDPDVRGALCAAVTPLV-GPTAAEVRERMGAAGRQRVQE  358

Query  364  RSWSAMVDRLIEYYEQVIPLHAAARRV  390
            RSW A+VD+L+++Y  V+  HA +R V
Sbjct  359  RSWPALVDQLVDHYSAVV--HAGSRSV  383
```

Figure 4. Pairwise alignment of Ksed_07870 with its top nr BLAST hit, a glycosyl transferase from *Serinococcus marinus*



Prediction of Ksed_07870

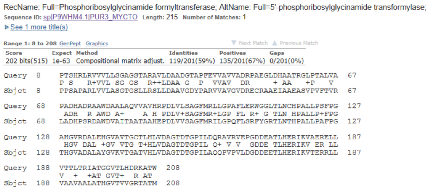
```
ID   Ksed_07870
FT   TOPO_DOM      1    391       NON CYTOPLASMIC.
//
```

The probability data used in the plot is found here, and the gnuplot script is here.

Figure 5. Phobius results for Ksed_07870 showing a lack of both a signal peptide and transmembrane helixes.

***Kytococcus sedentarius* 07800**

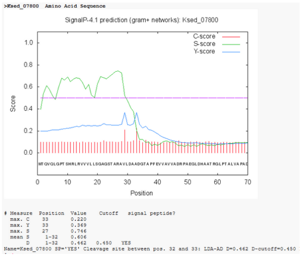
The initial proposed product of this gene by GENI-ACT was a phosphoribosylglycinamide formyltransferase, which is a THF dependent enzyme that is crucial in the production of purine. It catalyzes a nucleophilic acyl substitution of the formyl group, and forms N2 – formyl – N1 – (5 – phospho – 0 – ribosyl) glycinamide. This gene product proposal is supported by the BLAST hits for the amino acid sequence, the presence of well-curated domains within the amino acid sequence, the sequence logo, and the cellular location of the amino acid sequence. This proves that the proposed annotation is a phosphoribosylglycinamide formyltransferase.



```
RecName: Full=Phosphoribosylglycinamide formyltransferase; AltName: Full=5'-phosphoribosylglycinamide transformylase;
Sequence ID: sp|P9WHM4.1|PUR3_MYCTO  Length: 215  Number of Matches: 1
See 1 more title(s)

Range 1: 8 to 208  GenPept  Graphics                    Next Match  Previous Match
Score          Expect  Method                        Identities     Positives      Gaps
202 bits(515)  1e-63   Compositional matrix adjust.  119/201(59%)   135/201(67%)   0/201(0%)

Query  8     PTSHRLRVVVLLSGAGSTARAVLDAADGTAPFEVVAVVADRPAEGLDHAATRGLPTALVA  67
             P S   R+VVL SG GS  R++LDAA G  P  VVAV  DR      + AA   +P   V
Sbjct  8     PPSAPARLVVLASGTGSLLRSLLDAAVGDYPARVVAVGVDRECRAAEIAAEASVPVFTVR  67

Query  68    PADHADRAAWDAALAQVVAVHRPDLVLSAGFMRLLGPAFLERWGGLTLNCHPALLPSFPG  127
              ADH  R AWD A+     A H PDLV+SAGFMR+LGP FL R+ G TLN HPALLP+FPG
Sbjct  68    LADHPSRDAWDVAITAATAAHEPDLVVSAGFMRILGPQFLSRFYGRTLNTHPALLPAFPG  127

Query  128   AHGVRDALEHGVAVTGCTLHLVDAGTDTGPILDQRAVRVEPGDDEATLHERIKVAERELL  187
              HGV DAL +GV VTG T+HLVDAGTDTGPIL Q+ V V  GDDE TLHERIKV ER LL
Sbjct  128   THGVADALAYGVKVTGATVHLVDAGTDTGPILAQQPVPVLDGDDEETLHERIKVTERRLL  187

Query  188   VTTLTRIATGGVTLHDRKATW  208
             V  +  +AT GVT+  R AT
Sbjct  188   VAAVAALATHGVTVVGRTATM  208
```

Figure 6. The pairwise alignment of Ksed_07800 with its top nr BLAST hit, a phosphoribosylglycinamide formyltransferase from *Mycobacterium tuberculosis* CDC1551.

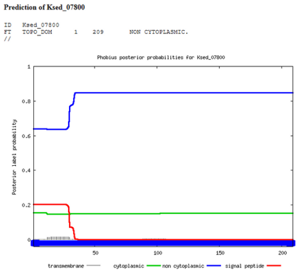
One interesting finding for Ksed_07800 was that there was a discrepancy between the SignalP and Phobius tools in predicting whether or not the protein contained a signal peptide. SignalP did predict the presence of a signal peptide (Figure 7), while Phobius did not (Figure 8). Given that Ksed_07800 most likely is a cytoplasmic enzyme, the Phobius results seems more reliable.



>Ksed_07800 Amino Acid Sequence

```
# Measure  Position  Value    Cutoff   signal peptide?
  max. C    33       0.220
  max. Y    33       0.369
  max. S    27       0.746
  mean S     1-32    0.606
       D     1-32    0.462   0.450   YES
Name=Ksed_07800 SP='YES' Cleavage site between pos. 32 and 33: LDA-AD D=0.462 D-cutoff=0.450
```

Figure 7. SignalP output for Ksed_07800 predicting that the protein contains a signal peptide



Prediction of Ksed_07800

```
ID   Ksed_07800
FT   TOPO_DOM      1    209       NON CYTOPLASMIC.
//
```

Figure 8. Phobius output for Ksed_07800 that fails to predict that the protein contains a signal peptide

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.

Locus	Proposed Annotation	Final Annotation
07800	Phosphoribosylglycinamide Formyltransferase	Phosphoribosylglycinamide Formyltransferase
07820	Bifunctional 5,10-methylene-tetrahydrofolate cyclohydrolase	Bifunctional 5,10-methylene-tetrahydrofolate cyclohydrolase
07870	Glycosyl Transferase	Glycosyl Transferase

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