Annotation of the *Kytococcus sedentarius* Genome from DNA Coordinates 403939 to 414240

Maddison Knapp and Lon Knappenberger
Westfield Academy and Central School, 203 East Main Street, Westfield, New York 14787 and the Western New York Genetics in Research Partnership

Abstract

Kytococcus sedentarius is a Gram-positive cocci. A specific gene was annotated from this bacteria to find its function. The database Geni-Act was used to determine what the gene was. Then the data that was found from the gene details page of Geni-Act was used in several different websites. Eight Geni-Act modules were used to gather this information. The results show a protein found in the cytoplasmic. With the help of Geni-Act and several other resources, it was determines that the gene product was a ubiquinone biosynthesis.

Introduction

Kytoccus sedentarius is a Gram positive bacterium that is arranged in packets of eight in a cubical shape (1). Kytococcus sedentarius was found on a partially submerged swing set in San Diego California. It is a cocciod or spherical shaped bacterium that can live on the skin (1).

The genome of *Kytococcus sedentarius* 541 2,785,024 bp long and codes for 2,703 genes. Of the 2,703 genes predicted, 2,639 were protein-coding genes. 72.1% of these genes were assigned with a predicted function while the remaining ones were annotated as hypothetical proteins.

The gene that we investigated, Ksed_02230 had the DNA coordinates of 221238..221912 containing 675 base pairs. The section codes for 224 amino acids.

The purpose was to look and verify information from GENI-ACT, and to better understand a poorly studied family within the tree of life. The other purpose was to determine a possible function of the gene product and provide an opportunity for original research.

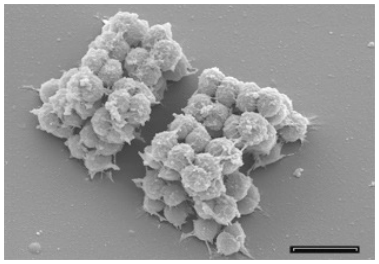


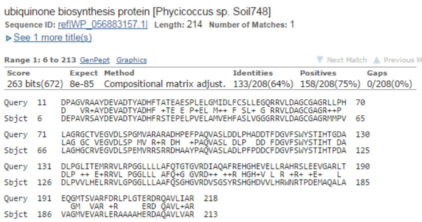
Figure 1. *Kytococcus sedentarius*.
Source: Standards in Genomic Science (2009) 1:12-20

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?
Module 6- Enzymatic Function	KEGG, MetaCyc, E.C. Number,	In what process does my protein take part?
Module 7- Gene Duplication/ Gene Degradation	Paralog, Pseudogene	Are there other forms of my gene in the bacterium? Is my gene functional?
Module 8- Evidence for Horizontal Gene Transfer	Phylogenetic Tree,	Has my gene co-evolved with other genes in the genome?

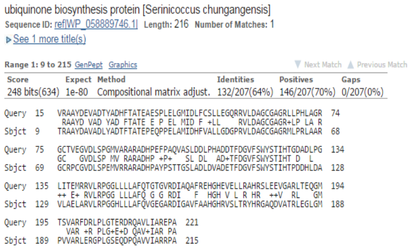
Results



ubiquinone biosynthesis protein [Phycicoccus sp. Soil748]
Sequence ID: ref|WP_056883157.1| Length: 214 Number of Matches: 1
See 1 more title(s)

```
Range 1: 6 to 213 GenPept  Graphics                                  Next Match  Previous Match
Score          Expect  Method                          Identities      Positives       Gaps
263 bits(672)  8e-85   Compositional matrix adjust.    133/208(64%)    158/208(75%)    0/208(0%)

Query  11   DPAGVRAAYDEVADTYADHFTATEAESPLELGMIDLFCSLLEGQRRVLDAGCGAGRLLPH  70
            D   VR+AYDEVADTYADHF +TE E P+EL M++ F SL+ G RRVLDAGCGAGR++P
Sbjct  6    DEPAVRSAYDEVADTYADHFRSTEPELPVELAMVEHFASLVGGGRRVLDAGCGAGRMMPV  65

Query  71   LAGRGCTVEGVDLSPGMVARARADHPEFPAQVASLDDLPHADDTFDGVFSWYSTIHTGDA  130
            LAG GC VEGVDLSP MV R+R DH  +PAQVASL DLP  DD FDGVFSWYSTIHT DA
Sbjct  66   LAGHGCRVEGVDLSPEMVRRSRRDHAAYPAQVASLADLPFPDDCFDGVFSWYSTIHTPDA  125

Query  131  DLPGLITEMRRVLRPGGLLLLAFQTGTGVRDIAQAFREHGHEVELLRAHRSLEEVGARLT  190
            DLP ++ E+RRVL PGGLLL AFQ+G GVRD++ ++R HGH+V L R +R+ +E+   L
Sbjct  126  DLPVVLHELRRVLGPGGLLLAAFQSGHGVRDVSGSYRSHGHDVVLHRWNRTPDEMAQALA  185

Query  191  EQGMTSVARFDRLPLGTERDRQAVLIAR  218
              GM  VAR +R     ERD QAVL+AR
Sbjct  186  VAGMVEVARLERAAAAHERDAQAVLVAR  213
```

Figure 1. This is the top protein BLAST hit for the non-redundant database. Which shows a ubiquinone biosynthesis protein found in *Phycicoccus* (evalue 8e-85).



ubiquinone biosynthesis protein [Serinicoccus chungangensis]
Sequence ID: ref|WP_058889746.1| Length: 216 Number of Matches: 1
See 1 more title(s)

```
Range 1: 9 to 215 GenPept  Graphics                                  Next Match  Previous Match
Score          Expect  Method                          Identities      Positives       Gaps
248 bits(634)  1e-80   Compositional matrix adjust.    132/207(64%)    146/207(70%)    0/207(0%)

Query  15   VRAAYDEVADTYADHFTATEAESPLELGMIDLFCSLLEGQRRVLDAGCGAGRLLPHLAGR  74
             RAAYD VAD YAD FTATE E P EL MID F +LL    RVLDAGCGAGR+LP LA R
Sbjct  9    TRAAYDAVADLYADTFTATEPEQPPELAMIDHFVALLGDGPRVLDAGCGAGRMLPRLAAR  68

Query  75   GCTVEGVDLSPGMVARARADHPEFPAQVASLDDLPHADDTFDGVFSWYSTIHTGDADLPG  134
            GC   GVDLSP MV RARADHP +P+   SL DL  AD+TFDGVFSWYSTIHT D  L
Sbjct  69   GCRPCGVDLSPEMVRRARADHPAYPSTTGSLADLDVADETFDGVFSWYSTIHTPDDHLDA  128

Query  135  LITEMRRVLRPGGLLLLAFQTGTGVRDIAQAFREHGHEVELLRAHRSLEEVGARLTEQGM  194
            ++ E+ RVLRPGG LLLAFQ G G RDI   F  HGH V L R HR  ++V  RL   GM
Sbjct  129  VLAELARVLRPGGHLLLAFQVGEGARDIGAVFAAHGHRVSLTRYHRGAQDVATRLEGLGM  188

Query  195  TSVARFDRLPLGTERDRQAVLIAREPA  221
              VAR +R PLG+E+D QAV+IAR PA
Sbjct  189  PVVARLERGPLGSEQDPQAVVIARRPA  215
```

Figure 2. This is the second protein BLAST hit. It is also a ubiquinone biosynthesis protein , but found in *Serinicoccus chungangensis. (evalue = 1e-80)*

Results

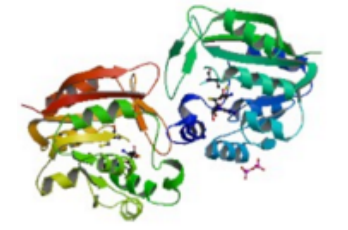
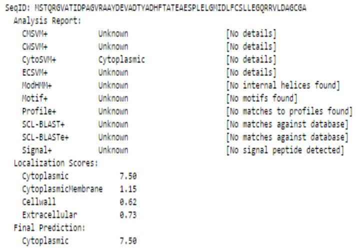
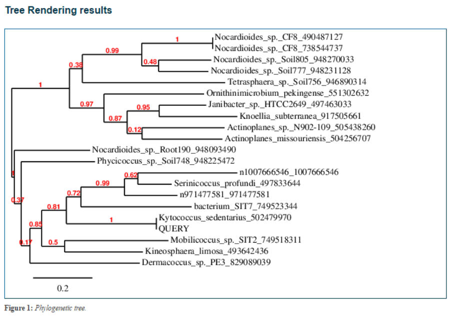


Figure 3. This is the three-dimensional figure of Protein Data Base top hit, SAM-dependent methyltransferase from *Corynebacterium glutamicum*.Methyl transferase are involved in the synthesis of coenzyme Q or ubiquinone.[3]



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

Figure 4. This is the results for the Psort-B for Ksed_02230. This shows that the location of the gene productis in the cytoplasmic.



Tree Rendering results

Figure 1: *Phylogenetic tree.*

Figure 5. Phylogenetic tree shows the Serinicoccus as the closest neighbor to Kytococcus. Both Serinicoccus and Kytococcus are classified in the same family level. This indicates that Ksed_02230 did not arise in Kytococcus by horizontal gene transfer.

Results

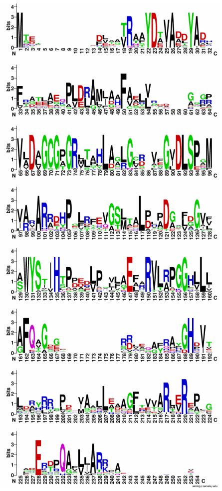


Figure 3. This is a Web Logo for T-COFFEE. Which is fairly highly conserved for Homologs from position 65 to 159.

Conclusion

The GENI-ACT proposed gene product for Ksed_02230 is ubiquinone biosynthesis which is found in the cytoplasmic. Ubiquinone (UQ) is also known as a coenzyme, which is a redox- active lipophilic (combining with or dissolving) molecule that is found in all cells. Also it is in the membranes of many organelles where it participates in a variety of processes. In the inner membrane of the mitochondria, UQ diffuses freely and is necessary for the function of the electron transport chain (ETC), as it enables the transfer of electrons from mitochondrial complexes I and II and mitochondrial complex III. [2]

References

1. http://www.vumicro.com/vumie/help/VUMICRO/Kytococcus_Sedentarious.htm
2. http://www.nature.com/ncomms/2015/150306/ncomms7393/full/ncomms7393.html
3. http://www.biochemistry.ucla.edu/biochem/Faculty/SClarke/pdf2/168.pdf

Acknowledgments

This project was supported by a National Science Foundation Innovative Technology Experiences for Students and Teachers (ITEST) Award (#1311902).