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

Annotation of the *Kytococcus sedentarius* Genome from DNA Coordinates 218553 to 218966

Madyson Abbey 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 bacteria. Using the gene annotation website GENI-ACT, a single gene (Ksed_02180) of *Kytococcus sedentarius* was annotated. The website was used to us discover putative gene products, a nucleotide-binding protein. This protein exists in the cytoplasmic membrane and may be involved in transmitting a signal from outside to inside of a cell.

Introduction

Kytococcus sedentarius is a spherical bacteria that occur predominantly in tetrads . *K. sedentarius* is Gram-positive, nonmotile, non-encapsulated, and not endospore-forming. This bacteria is strictly aerobic, requires methionine and grows well in NaCl at concentrations up to 10% [1]

The genome of *Kytococcus sedentarius* 541 2,785,024 base pairs 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.[1]

The gene that we investigated, Ksed_02190, had the DNA coordinates from 218553 to 218966. The gene is 414 base pairs long and codes for 137 amino acids.

The purpose was to look for a verify information found in the gene information page for Ksed_02190 on GENI-ACT. A second purpose is to better understand a poorly studied family within the bacterial suborder *Micrococcineae.* A third purpose is to determine a possible function of the gene product. A final purpose is to provide an opportunity for original research for high schoo students.

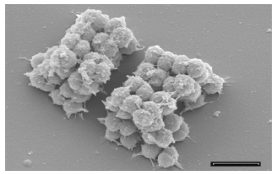

Figure 1. This is an image of Kytococcus sedentarius
Source: Standards in Genomic Science (2009) 1:12-20

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

nucleotide-binding protein [Serinicoccus profundi]
Sequence ID: ref|WP_010146671.1| Length: 136 Number of Matches: 1

Range 1: 1 to 135 GenPept Graphics

Score	Expect	Method	Identities	Positives	Gaps
140 bits(352)	4e-39	Compositional matrix adjust.	72/135(53%)	91/135(67%)	0/135(0%)

```
Query  1    MTAVFIDSTILLHAAGGDHAMRAPSRRWLEAAASGGMALHASVEAGQEFLFHRLRKSTPE  60
            M AVF+DS++LL A G +H  RA   R +E AA+G + LH SVEAGQEFL HRLR+++
Sbjct  1    MRAVFVDSSVLLLAFGDEHPDRAACVRVVELAATGDLRLHQSVEAGQEFLVHRLRRTSTA  60

Query  61   QATAQFGLIDGLITWMPFDEEVLRRSVALAATGALGGRDAMHAATALAAGFTEIVSADRD  120
            +A   F ++D L+ WM FD E LR S  L     GRDA+HAATAL AGF+ IV+ D D
Sbjct  61   EAITDFDVLDRLVWMDFDMETLRESRRLVEREEARGRDAVHAATALLAGFSGIVTTDTD  120

Query  121  FDKVPGLARRDPGDL  135
            FD + GL+R DP DL
Sbjct  121  FDGITGLSRVDPSDL  135
```

Figure 2. This is the top hit for a protein BLAST for the non-redundant database top hit. It shows a nucleotide-binding protein in *Serinicoccus profundi* (e-value 4e-39), Score 140.

```
SeqID: Unknown_1
  Analysis Report:
    CMSVM+              CytoplasmicMembrane
    GWSVM+              Unknown
    CytoSVM+            Unknown
    ECSVM+              Unknown
    HodMM+              Unknown
    Motif+              Unknown
    Profile+            Unknown
    SCL-BLAST+          Unknown
    SCL-BLASTe+         Unknown
    Signal+             Unknown
  Localization Scores:
    Cytoplasmic              0.32
    CytoplasmicMembrane      9.55
    Cellwall                 0.12
    Extracellular            0.01
  Final Prediction:
    CytoplasmicMembrane      9.55
```

Figure 4. These are the results of PSORT B. This shows that Ksed_02190 is found in the cytoplasmic membrane (value 9.55).

Results

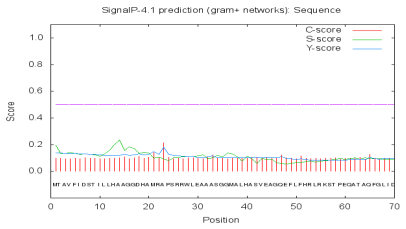

Figure 3. SignalP is a test to determine whether or not Ksed_02190 has is a single peptide.. The test came up negative The results were included because the Psort B predicts the gene product is in the cytoplasmic membrane, but is NOT a signal peptide.

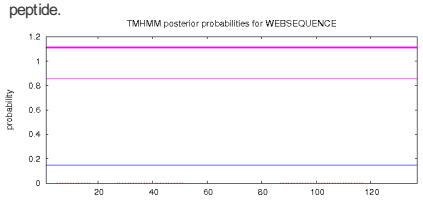

Figure 5. These are the results of TMHMM. This shows that thee are no transmembrane helices present. Again, Ksed_02190 is in the cytoplasmic membrane, but is not a signal peptide or a transmembrane peptide.

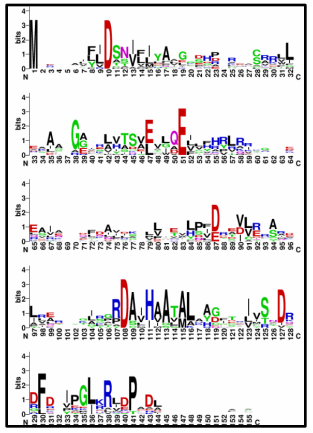

Figure 6.
These are the results of the WebLogo activity. This shows Ksed_02190 is not highly conserved with other orthologs.

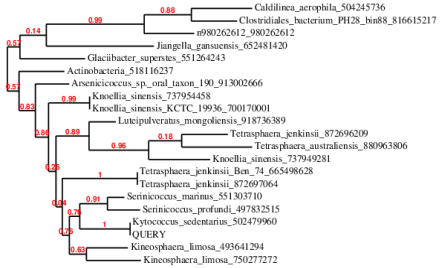

Figure 7. This is the Phylogenetic Tree for Ksed_02190. It is most closely related to *Serinicoccus*; They are in the same order but different family. This is evidence that *Kytococcus* received this gene from heredity.

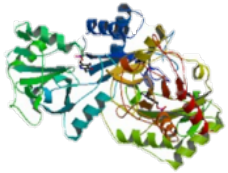

Figure 8. This is a three-dimensional image of Ksed_02190 when blasted in the Protein Data Base. This top hit is a nucleotide-binding protein.

Conclusions

After using the program GENI-ACT, Ksed_02190 is proposed to be a nucleotide-binding protein, found in the Cytoplasmic Membrane.

A Nucleotide binding protein s a protein that binds to nucleotides. For example, kinase proteins bind adenosine triphosphate (ATP) and G proteins bind guanosine triphosphate (GTP). Others bind to cyclic adenosine monophosphate (cAMP), Cyclic guanosine monophosphate (cGMP) and cytidine triphosphate (CTP) [2] It is unclear which nucleotide the gene product bind its.

It is interesting that the gene product is found in the cytoplasmic membrane (PSORTB 9.55, but is not a Signal peptide or a transmembrane helix. G proteins act as switches and transmit signals from outside to inside the cell.

References

1. Sims et al. (2009). Complete genome sequence of *Kytococcus sedentarius* type strain (541T). *Standards in Genomic Sciences*, 12 - 20.
2. http://www.nature.com/subjects/nucleotide- binding proteins

Acknowledgments

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