Annotation of the *Kytococcus sedentarius* Genome from Locus Tags Ksed_12620 to Ksed_12640

Abigail Decicco, Cayden Piegza, Ely Personius and Dr. Paul Weber
Amherst Central High School, Buffalo NY and The Western New York Genetics in Research Partnership

Abstract

A group of 3 genes from the microorganism *Kytococcus sedentarius* (Ksed_12620 – Ksed_12640) were annotated using the genome annotation website GENI-ACT. The Genbank proposed gene function for each gene, and 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, enzymatic function, the possibility of horizontal gene transfer, and the production of an RNA product. 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 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 illustrates the morphology of *Kytococcus sedentarius* by electron microscopy (Pospísil, et al., 1998).

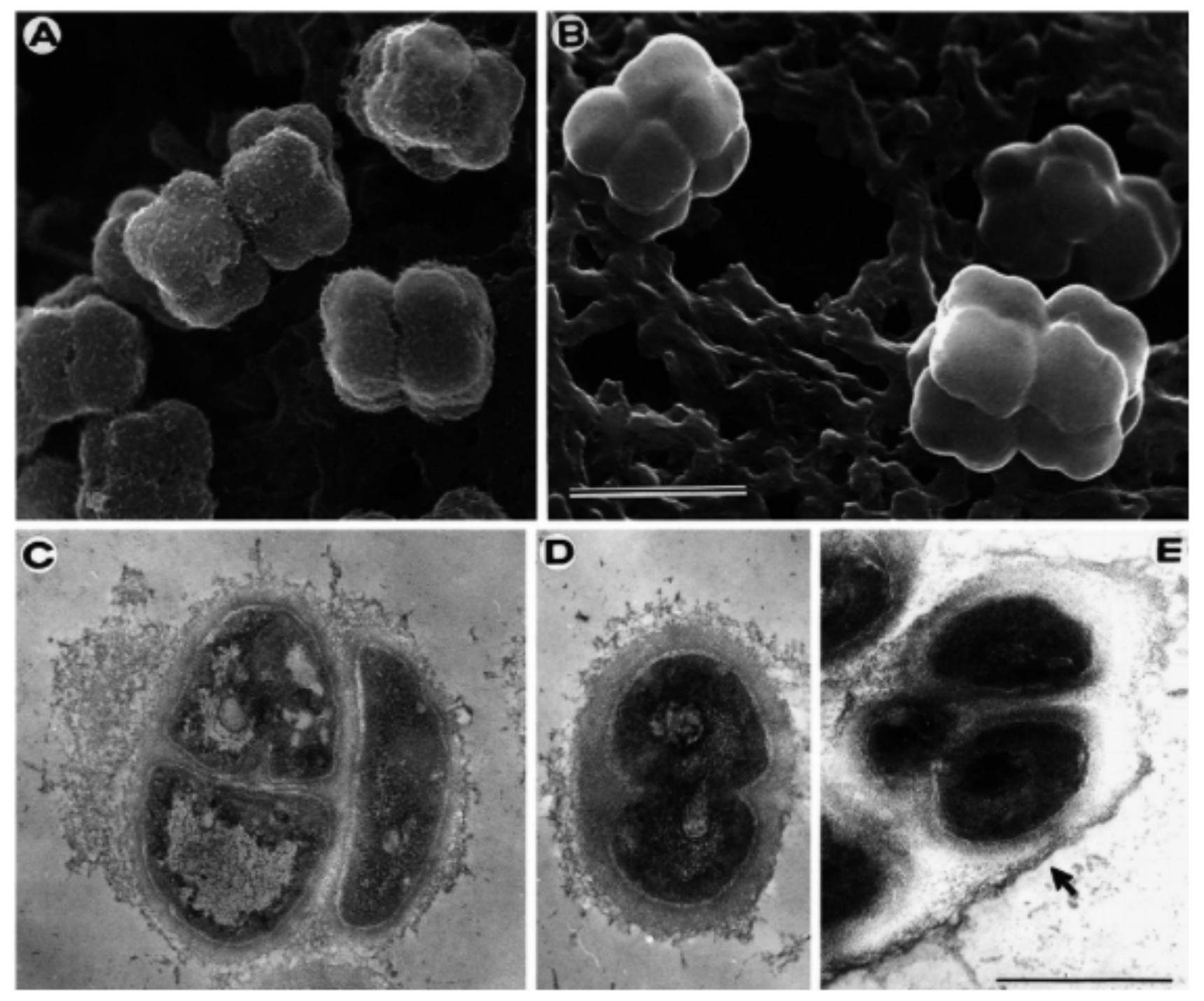


Figure 1. Scanning (A and B) and Transmission (C-E) electron micrographs of *Kytococcus sedentaius.*

Methods

Modules of the GENI-ACT (http://www.geni-act.org/) were used to complete *Kytococcus sedentarius* genome annotation . The modules are described below:

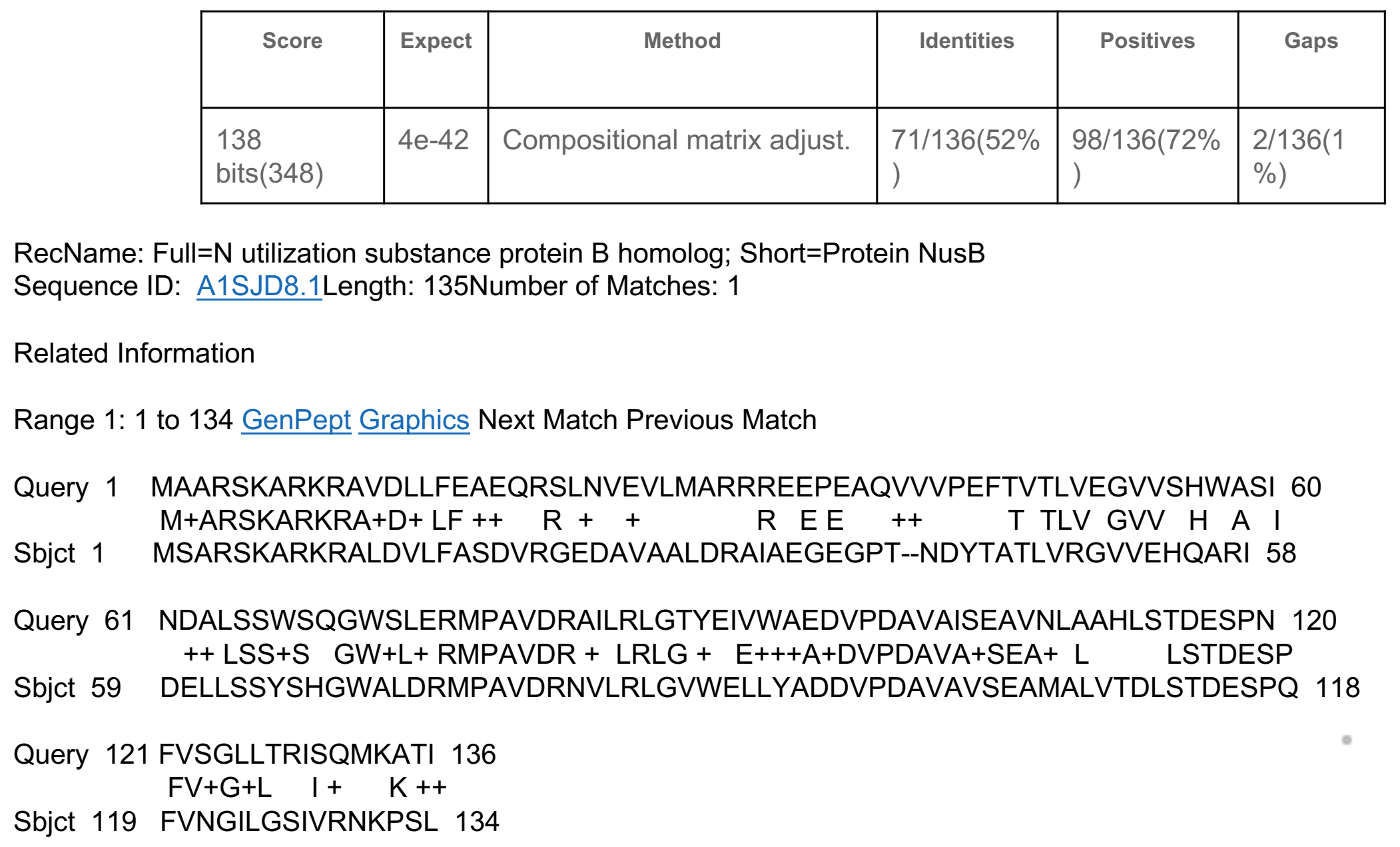
Modules	Activities	Questions Investigated
Module 1- Basic Information Module	DNA Coordinates and Sequence, Protein Sequence	What is the sequence of my gene and protein? Where is it located in the genome?
Module 2- Sequence-Based Similarity Data	Blast, CDD, T-Coffee, WebLogo	Is my sequence similar to 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?
Module 9- RNA	RFAM	Does my gene encode a functional RNA?

Results

Ksed_12620:
With the data that was proposed for Ksed_12620 Indicates that the protein is located in the cell as shown in figures 3 and 4. In figure 3 it can be seen that the protein does not have a signal peptide, while figure 4 shows the lack of transmembrane helixes, suggesting Ksed_12620 is a cytoplasmic protein.

Ksed_12630:
The data that was proposed for Ksed_12630 Indicates that the protein is very similar to other proteins that have a function that is an Elongation factor P. This indicates that my protein could theoretically have an Elongation factor P as its function.

Ksed_12640:
The BLAST results have lead me to the conclusion that my protein is an RNA binding protein. The top two results were all RNA binding proteins, thus, Ksed_12640 is most likely an RNA binding protein



Score	Expect	Method	Identities	Positives	Gaps
138 bits(348)	4e-42	Compositional matrix adjust.	71/136(52%)	98/136(72%)	2/136(1%)

RecName: Full=N utilization substance protein B homolog; Short=Protein NusB
Sequence ID: A1SJD8.1Length: 135Number of Matches: 1

Related Information

Range 1: 1 to 134 GenPept Graphics Next Match Previous Match

```
Query  1    MAARSKARKRAVDLLFEAEQRSLNVEVLMARRREEPEAQVVVPEFTVTLVEGVVSHWASI  60
             M+ARSKARKRA+D+ LF ++     R  +   +            R   E E     ++          T  TLV  GVV   H   A   I
Sbjct  1    MSARSKARKRALDVLFASDVRGEDAVAALDRAIAEGEGPT--NDYTATLVRGVVEHQARI  58

Query  61   NDALSSWSQGWSLERMPAVDRAILRLGTYEIVWAEDVPDAVAISEAVNLAAHLSTDESPN  120
               ++ LSS+S   GW+L+ RMPAVDR +  LRLG +  E+++A+DVPDAVA+SEA+  L          LSTDESP
Sbjct  59   DELLSSYSHGWALDRMPAVDRNVLRLGVWELLYADDVPDAVAVSEAMALVTDLSTDESPQ  118

Query  121 FVSGLLTRISQMKATI  136
             FV+G+L     I +      K ++
Sbjct  119  FVNGILGSIVRNKPSL  134
```

Figure 2. Alignment of the Ksed_12640 with it top SwisProt BLAST hit, NusB.

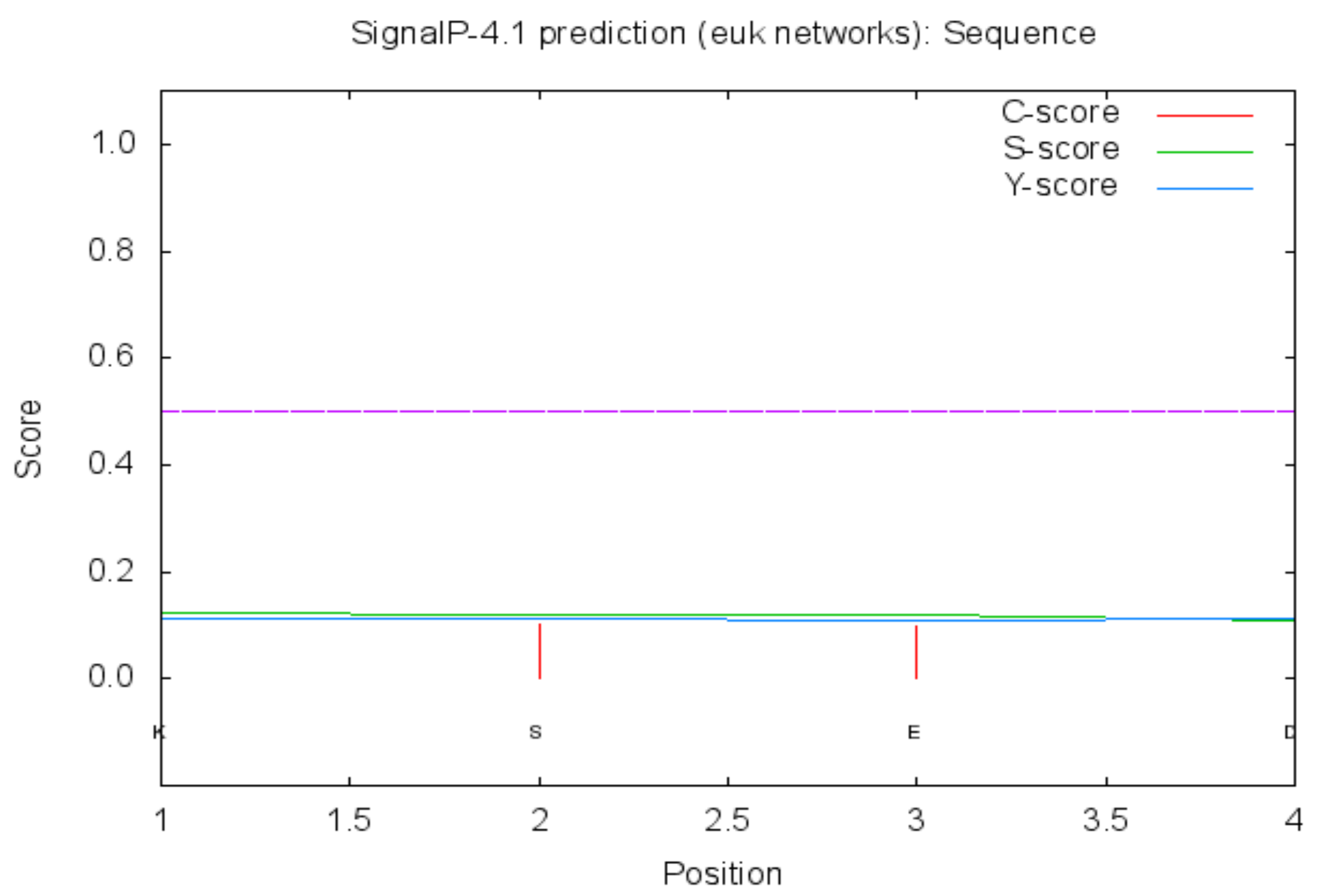


Figure 3 – SignalP output for Ksed_12620 showing it does not have a signal peptide.

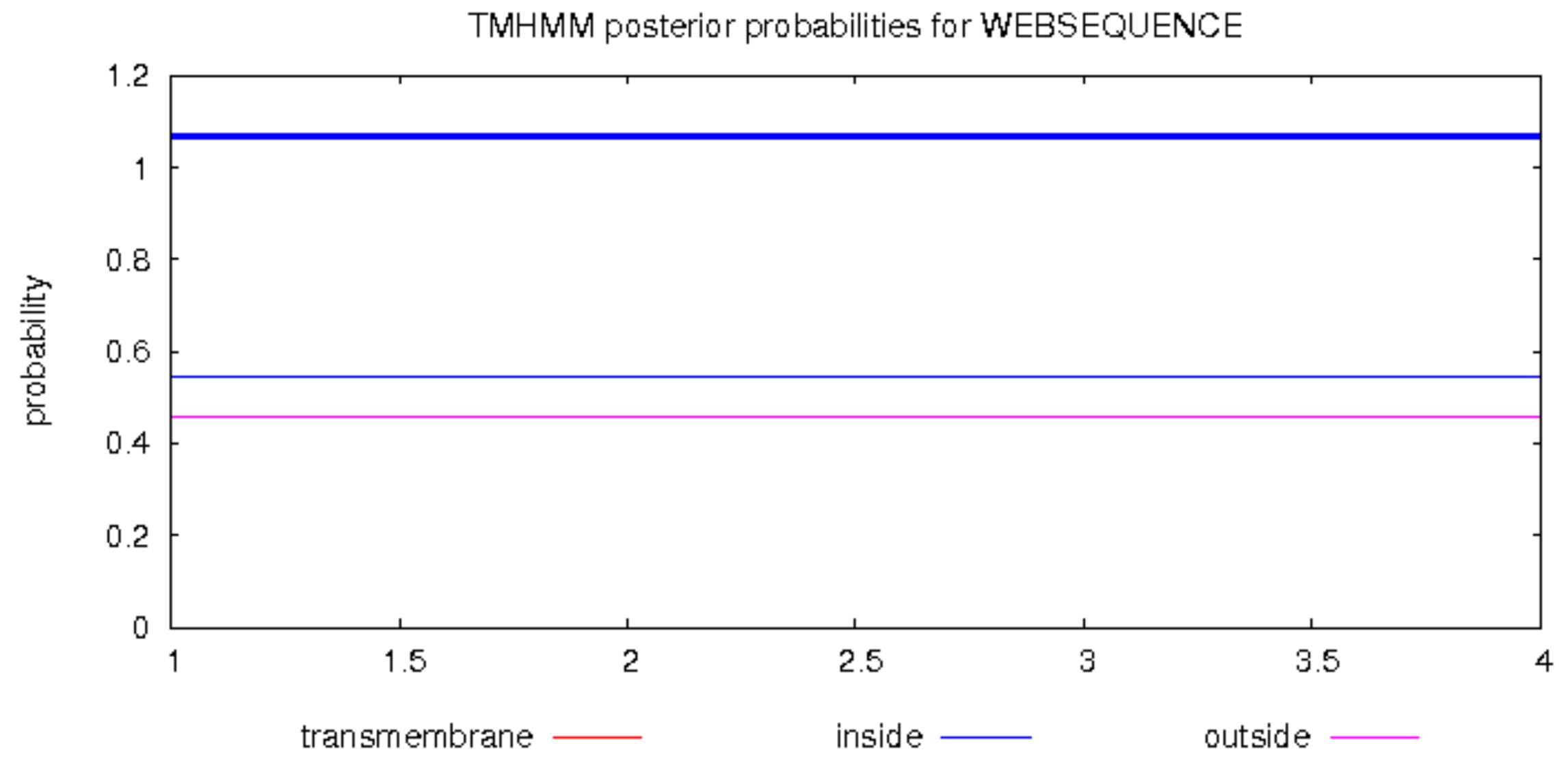


Figure 4. TMHMM output that shows that Ksed_12620 lacks transmembrane helixes and is most likely found in the cytoplasm of *Kytococcus sedentarius.*

Figure 5- Ksed_12630 - shows a phylogenetic tree showing the proteins in other organisms that are closely related to Ksed_12630.

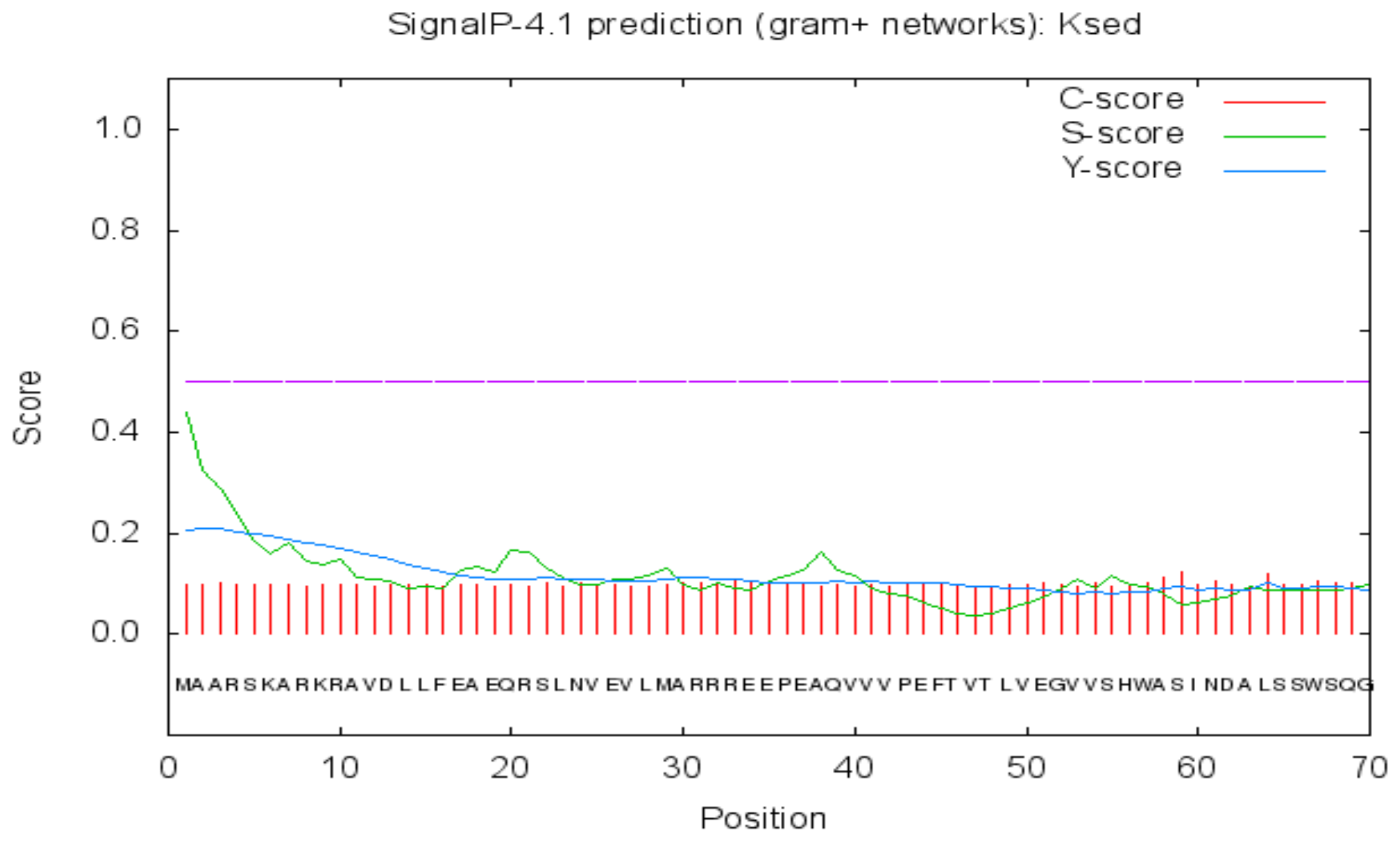


Figure 5 – Ksed_12640 This signalP graph shows that there is no cleavage site, thus no signal peptide in the protein.

Conclusion

Ksed_12620:
The initial proposed product of this gene by GENI-ACT was a glutamate synthase family protein. This was supported by the top BLAST hits for the amino acid sequence, DNA coordinates, and by all my graphs. So my conclusion is that my protein is a cytoplasmic glutamate synthase family protein.

Ksed_12630:
The initial proposed product of this gene by GENI-ACT is elongation factor P. This gene product proposal was supported by the top BLAST hits for the amino acid sequence, The phylogenetic tree, and DNA coordinates. As such, the proposed annotation is possibly an elongation factor P. The elongation factor P helps in translation when a ribosome is making a protein. It is required for efficient peptide bond synthesis on 70S ribosomes.

Ksed_12640:
The initial proposed purpose of this gene by GENI-ACT was RNA binding. This was supported by my top BLAST hits, as well as the signal peptide results. The BLAST results displayed how the protein was similar to other RNA binding proteins and the SignalP results displayed how the protein doesn't have a signal peptide. Therefore, my conclusion is that my protein is an RNA binding protein.

References

Sims et al. (2009). Complete genome sequence of *Kytococcus sedentarius* type strain (541T). Standards Genomic Sciences,12 - 20.
Pospisil, et. al. (1998) Can. J. Microbiol. 44:1007-1011

Acknowledgments

Special thanks to Dr. Brad Goodner, Dr. Stephen Koury and Dr. Rama Dey-Rao for their assistance with this project.
Supported by an NSF Innovative Technology Experiences for Students and Teachers (ITEST) Award - 1311902

www.buffalo.edu