Gene Annotation of Ksed_04940, a Putative Copper Chaperone Protein in *Kytococcus sedentarius*

Lucas Haskin and Lon Knappenberger

Westfield Academy and Central School, Westfield, New York 14787 and The Western New York Genetics in Research Partnership

Abstract

Kytococcus sedentarius is a Gram-positive cocci. A a gene with the locus tag of Ksed_04940 in this bacteria's genome was annotated to verify computer called information and look for possible function for the gene product. Eight GENI-ACT modules were used to gather this information. The results show the Ksed_04940 protein is likely to be extracellular, but it does not have a signal peptide. The top BLAST hits of Ksed_04940 all are copper binding proteins, whereas the top COG his is for copper chaperone CopZ, These findings support the computer call gene of Ksed_04940 as a copper binding protein. Further the COG hit suggests that Ksed_04940 is a copper binding involved in chaperoning metal ions into intracellular compartments or to copper dependent enzymes.

Introduction

Kytococcus 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 [2]. It is a cocciod or spherical shaped bacterium that can live on the skin (Figure 1) [1].

The genome of *Kytococcus sedentarius* is 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 sequence of Ksed_04940 had the DNA coordinates of 489574...489903, containing 330 base pairs that code for 109 amino acids[1].

The purpose was of this study was to determine if the computer annotation of Ksed_04940 was correct. Annotation of Ksed_04940 also allowed a better understanding a poorly studied family within the bacterial suborder *Micrococcineae*, determined a possible function of the gene product and provided an opportunity for student original research.

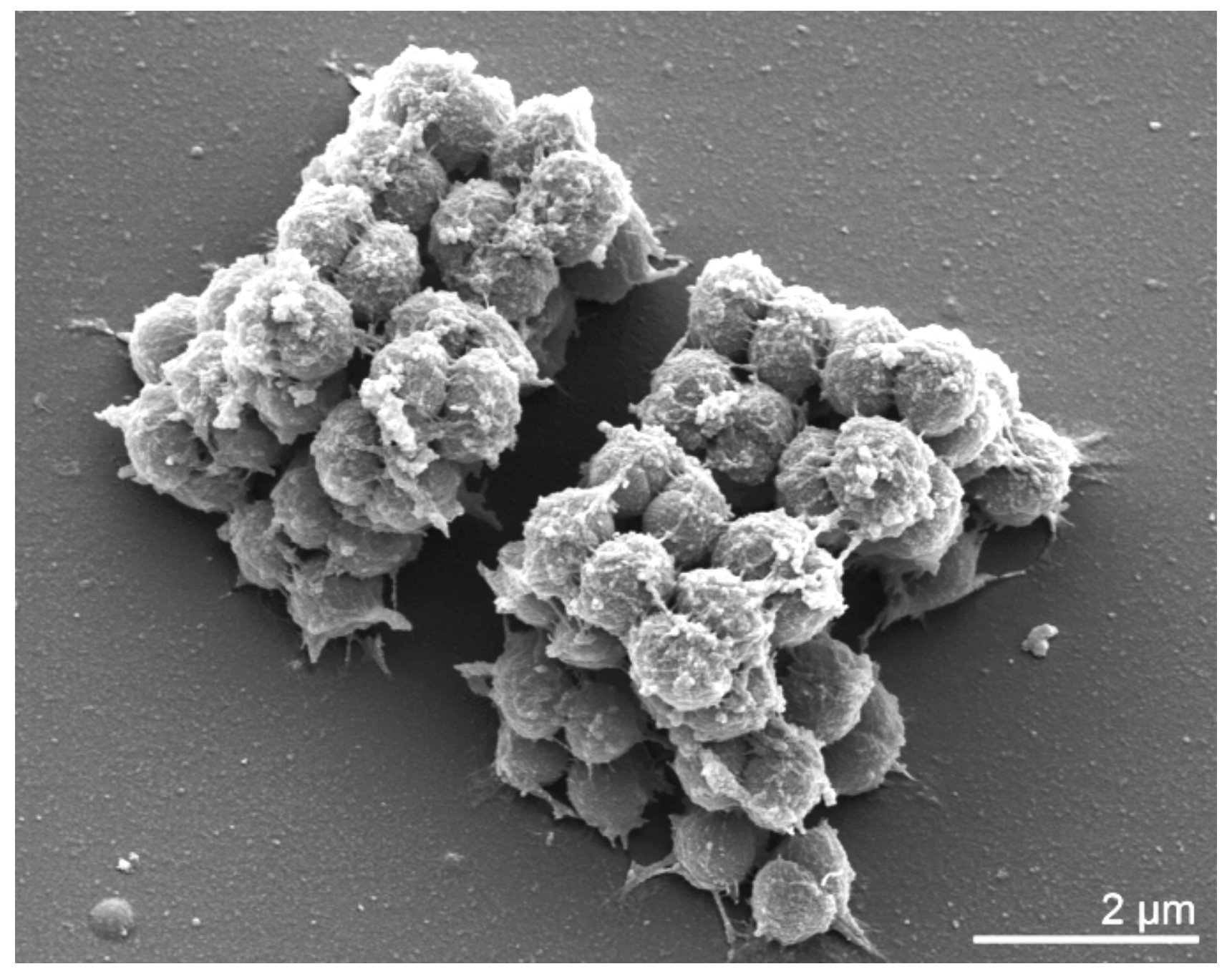


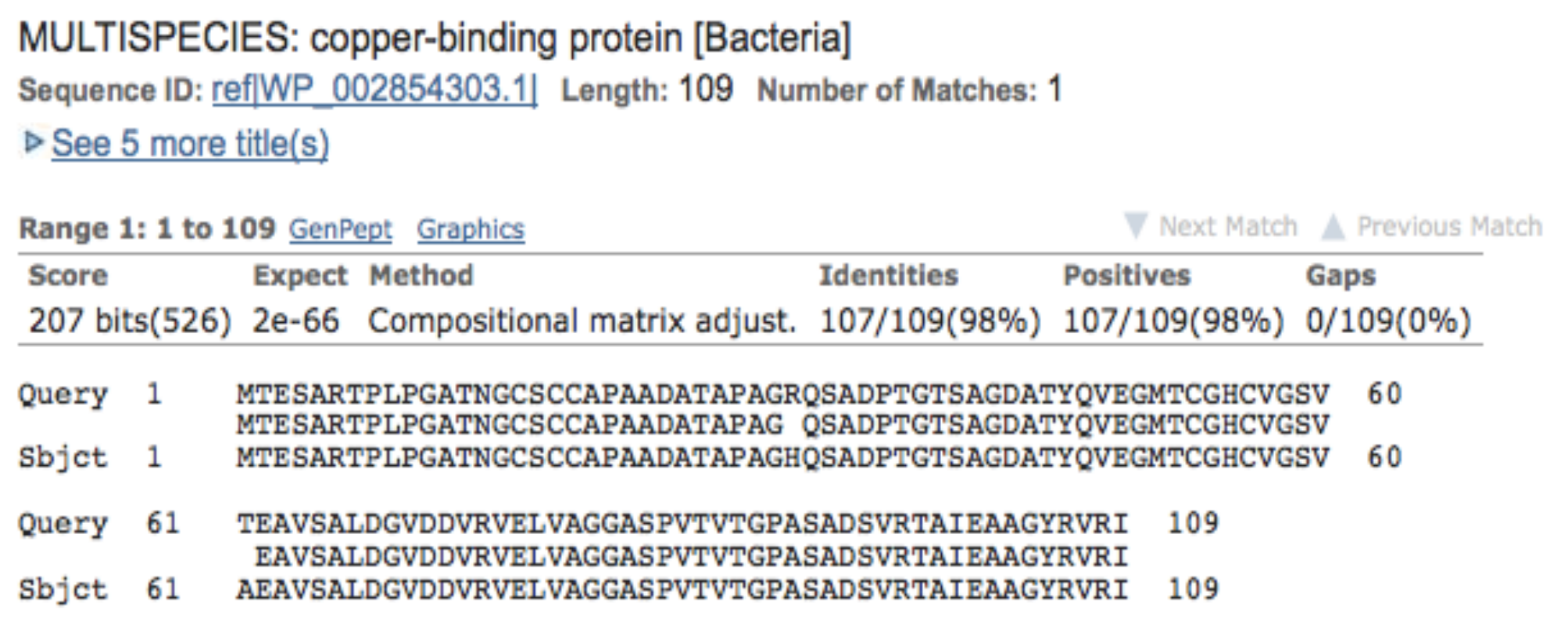
Figure 1. A scanning electron micrograph of *Kytococcus sedentarius* growing in the characteristic cuboidal packet form (Sims et al., 2009)

Materials and 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



```
MULTISPECIES: copper-binding protein [Bacteria]
Sequence ID: ref|WP_002854303.1|  Length: 109  Number of Matches: 1
▶ See 5 more title(s)

Range 1: 1 to 109  GenPept  Graphics                 ▼ Next Match  ▲ Previous Match
 Score          Expect  Method                       Identities     Positives      Gaps
 207 bits(526)  2e-66   Compositional matrix adjust. 107/109(98%)   107/109(98%)   0/109(0%)

Query  1    MTESARTPLPGATNGCSCCAPAADATAPAGRQSADPTGTSAGDATYQVEGMTCGHCVGSV  60
            MTESARTPLPGATNGCSCCAPAADATAPAG QSADPTGTSAGDATYQVEGMTCGHCVGSV
Sbjct  1    MTESARTPLPGATNGCSCCAPAADATAPAGHQSADPTGTSAGDATYQVEGMTCGHCVGSV  60

Query  61   TEAVSALDGVDDVRVELVAGGASPVTVTGPASADSVRTAIEAAGYRVRI  109
             EAVSALDGVDDVRVELVAGGASPVTVTGPASADSVRTAIEAAGYRVRI
Sbjct  61   AEAVSALDGVDDVRVELVAGGASPVTVTGPASADSVRTAIEAAGYRVRI  109
```

Figure 2. The results of a protein BLAST using the Non-redundant database shows a Multi-species copper-binding protein as the top hit (e-value 2e-66).

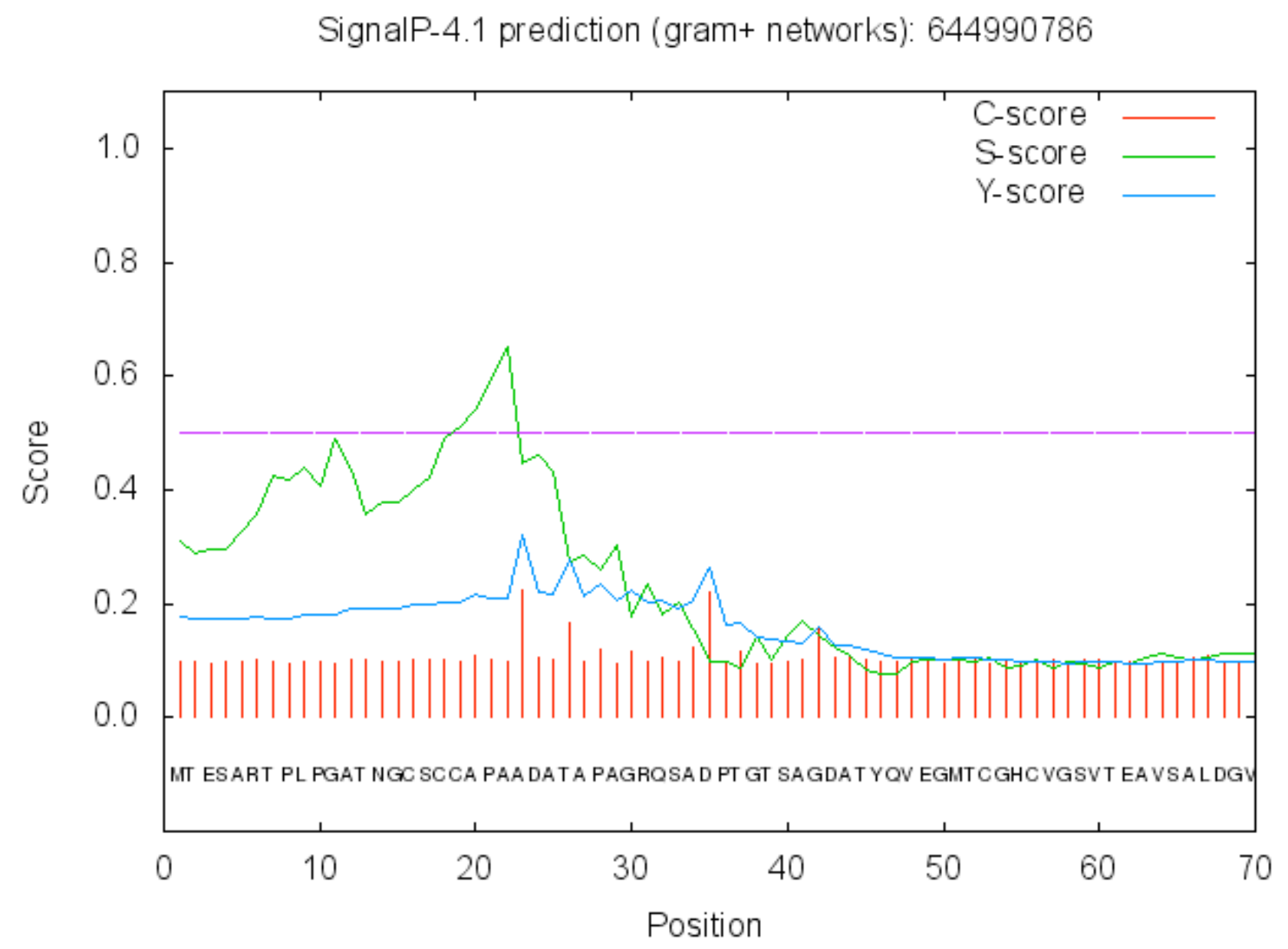


Figure 3. SignalP results that predict whether Ksed_04940 has a signal peptide. No signal peptide is predicted.

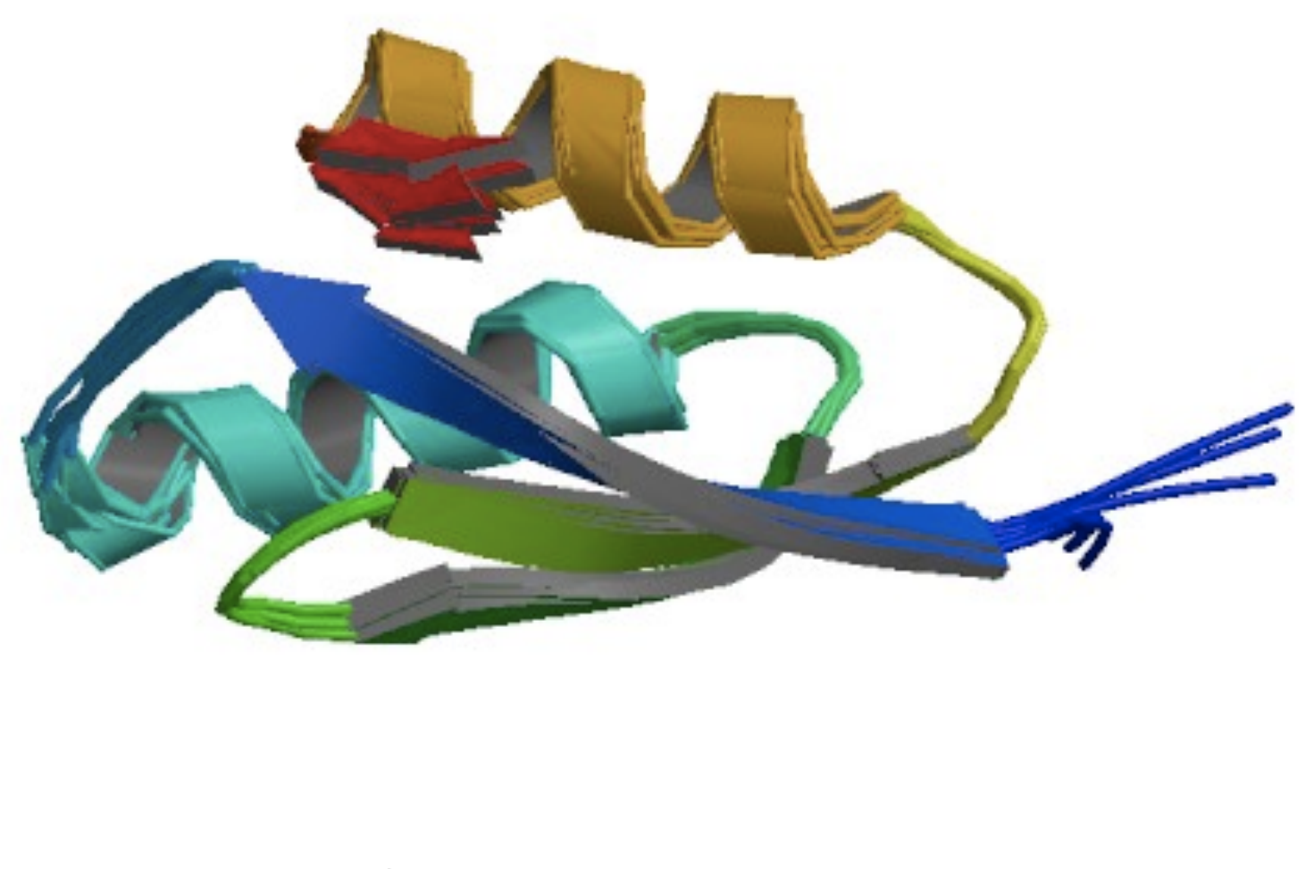


Figure 4. Am image of the three-dimensional structure of the top hit of Ksed_04940 in the Protein Data Base. This is a solution structure of a putative copper-ion-binding protein from *Bacillus anthracis* .

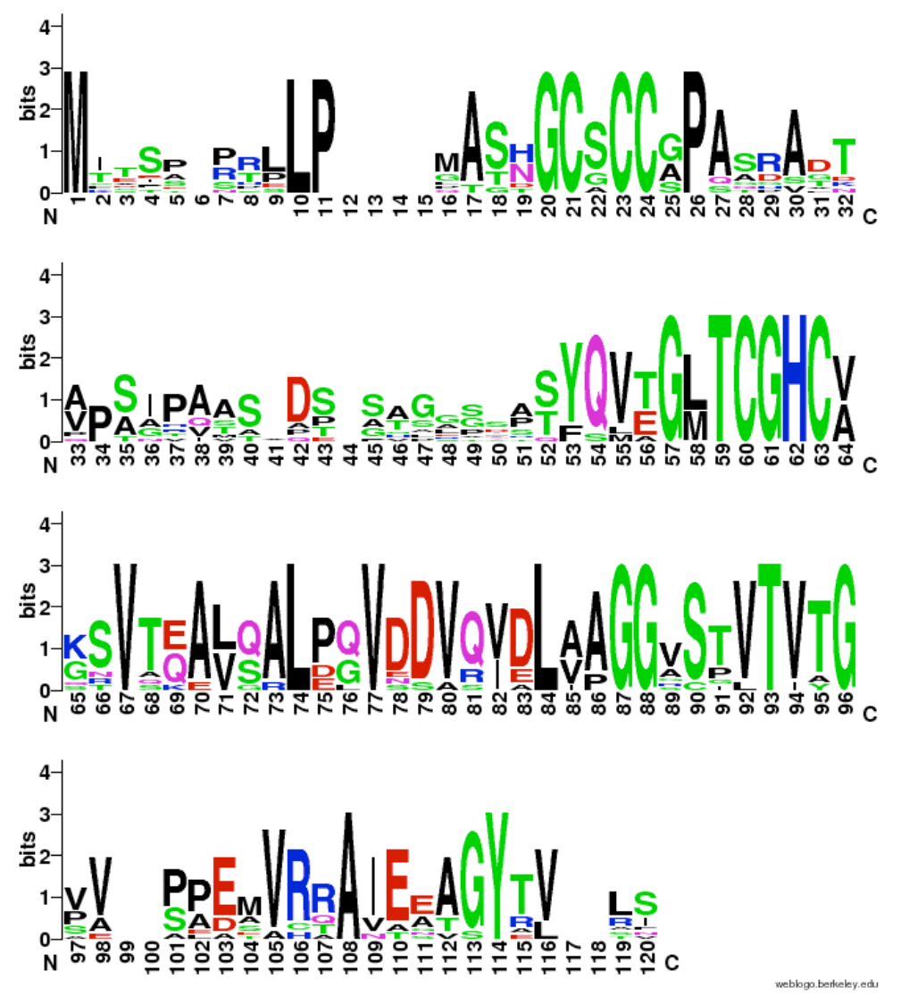
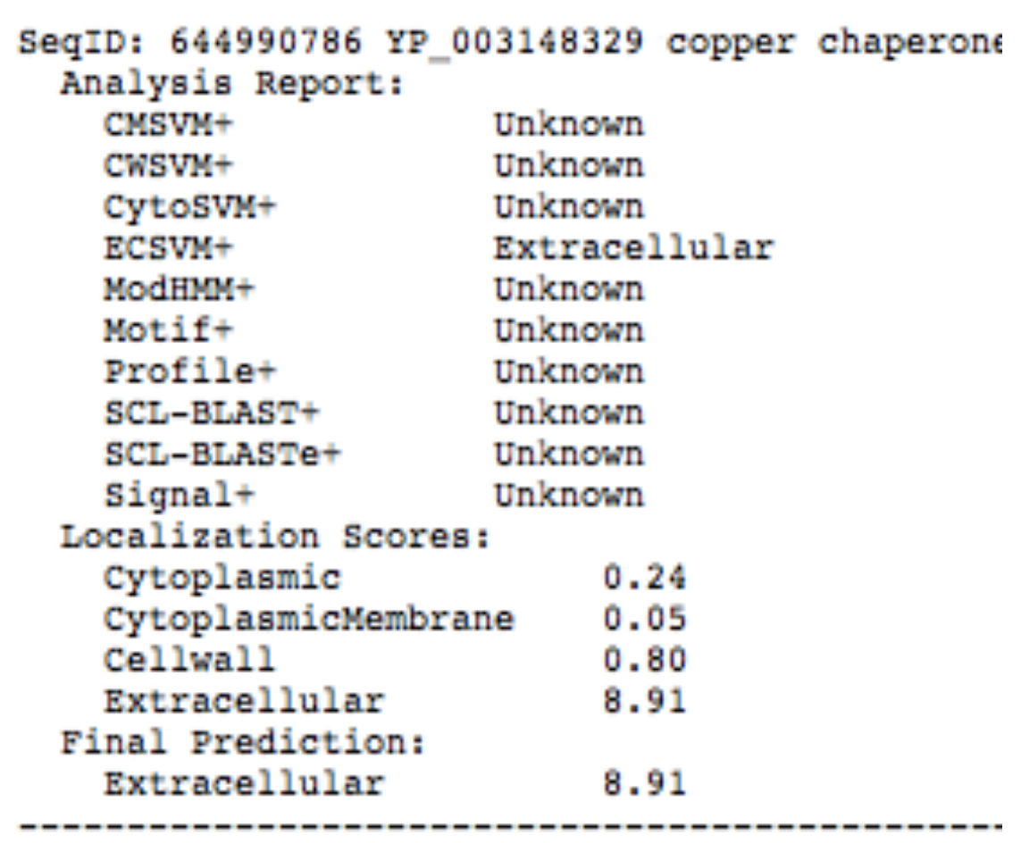
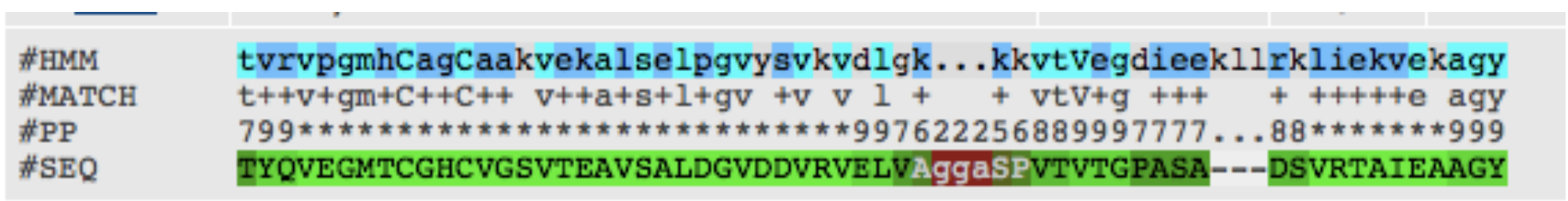


Figure 5. This is the WebLogo of the T-Coffee multisequence alignment for Ksed_04940. It shows that our gene is highly conserved throughout most of the sequence.



```
SeqID: 644990786 YP_003148329 copper chaperone
  Analysis Report:
    CMSVM+              Unknown
    CWSVM+              Unknown
    CytoSVM+            Unknown
    ECSVM+              Extracellular
    ModHMM+             Unknown
    Motif+              Unknown
    Profile+            Unknown
    SCL-BLAST+          Unknown
    SCL-BLASTe+         Unknown
    Signal+             Unknown
  Localization Scores:
    Cytoplasmic                0.24
    CytoplasmicMembrane        0.05
    Cellwall                   0.80
    Extracellular              8.91
  Final Prediction:
    Extracellular              8.91
-------------------------------------------
```

Figure 6. The Psort-B section of the gene notebook, and this shows the location of the gene product as extracellular. This makes sense based on the purpose of chaperone proteins.



```
#HMM     tvrvpgmhCagCaakvekalselpgvysvkvdlgk...kkvtVegdieekllrkliekvekagy
#MATCH   t++v+gm+C++C++ v++a+s+l+gv +v v l +    + vtV+g +++    + +++++e agy
#PP      799***************************99762225688999777...88*******999
#SEQ     TYQVEGMTCGHCVGSVTEAVSALDGVDDVRVELVAggaSPVTVTGPASA---DSVRTAIEAAGY
```

Figure 7. PFAM pairwise alignment shows a heavy metal association domain is conserved. (e value = 8.1e-13)

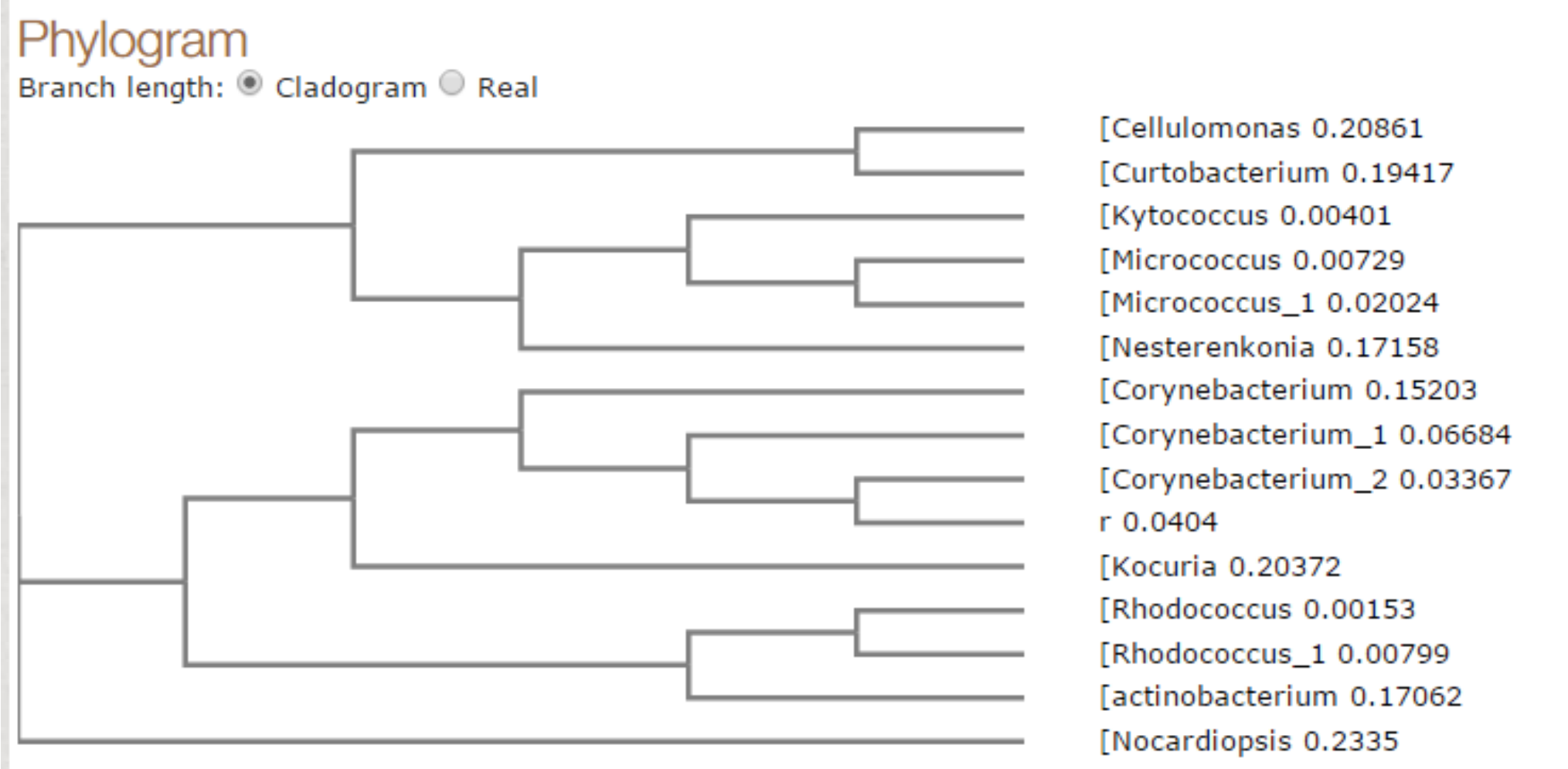


Figure 8. This graph is the phylogenetic tree for the . It shows *Micrococcus* as the closest neighbor. *Kytococcus* and *Micrococcus* are the same down to taxonomic family. This indicates Heredity and not Lateral gene transfer.

Results continued

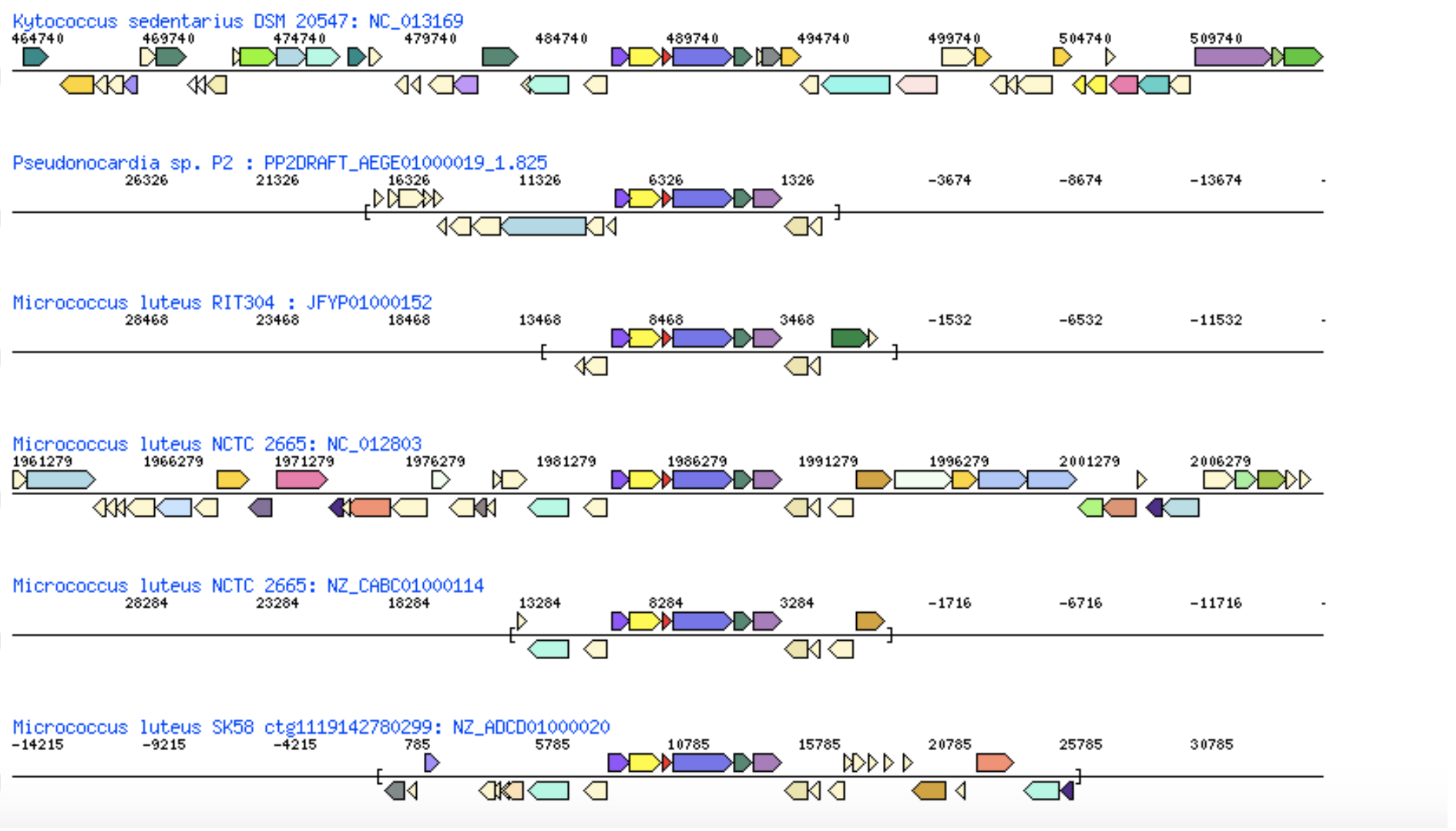
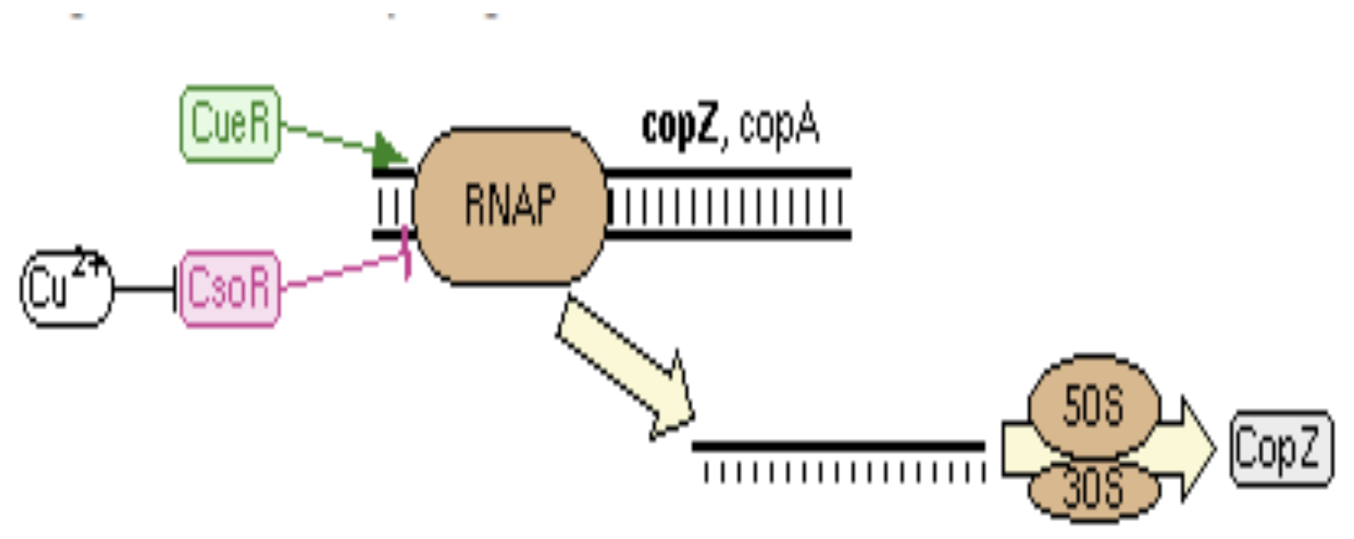


Figure 9.This is the ortholog neighborhood viewer of the gene Ksed_04930. It shows evidence of transfer by heredity. The neighborhood is conserved in *Kytococcus* and its common ancestors *Micrococcus* and *Pseudonocardia.*

Conclusions

The top BLAST hit is a Copper-binding protein, which is extracellular. Copper-binding proteins chaperone copper ions to copper-dependent enzymes and protect the cytoplasm during the transport of metal ions.[3] Mutations to these proteins in Humans have been linked to human diseases such as Menkes disease (growth failure) and Wilson's disease. (liver disease)[4].



http://biocyc.org/BSUB/NEW-IMAGE?object=BSU33510-MONOMER
Bsub cyc pathway.

Figure 10: Illustration of a copper insertion chaperone and transporter component from *Bacillus subtitlis*

Literature Cited

1. https://img-act.jgi-psf.org/notebook/pages/24057/edit
2. Longshaw, et al. Journal of Applied Microbiology 2002, 93, 810-816.
3. Markossian KA1, Kurganov BI. 2003. Biochemistry (Mosc). 2003 Aug;68(8):827-37.
4. http://www.trinity.edu/lhunsick/copper_chaperones.htm
5. Sims et al. (2009). Standards in Genomic Sciences, 12 - 20.

Acknowledgements

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