

Annotation of the *Kytococcus sedentarius* Genome from DNA Coordinates 05470 to 05490

Hailey Bartz*, Daniel Brush*, Kara Dobryznski* and Samantha Evans

Iroquois High School and the Western New York Genetics in Research Partnership *Indicates equal contribution

Abstract

A group of 3 consecutive genes from the microorganism *Kytococcus sedentarius* (Ksed_05470 – Ksed_05490) were annotated using the collaborative genome annotation website GENI-ACT. The Genbank proposed gene product name 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, potential alternative open reading frames, and the possibility of horizontal gene transfer. The Genbank proposed gene product name RNA Polymerase sigma-70 factor is an enzyme that is involved in the transcription of DNA to ultimately produce proteins. The Genbank proposed gene Beta-lactamase class C is a protein that has enzymatic function, helping fight invading antibodies to protect the cell and provide immunity. The Genbank proposed gene product name Chaperone Protein DnaK is a protein that helps with the synthesis and folding of proteins. These three genes all happened to appear as enzymes that have no relatable function, but were all properly annotated according to the modules.

Introduction

Kytococcus sedentarius is a bacterium that is classified under the scarcely populated *Kytococcus* genome, which contains only two species. This organism is gram positive and falls under the actinobacterial family name *Dermacoccaceae* (Sims et al., 2009). Originally classified under the genus *Micrococcus*, the bacterium was later classified under the *Kytococcus* genome after a dissection of the former genus (Sims et al., 2009). The organism is responsible for the symptoms of the pitted keratolysis skin disorder. The bacterium is strictly aerobic and grows best in conditions rich in amino acids, specifically methionine.

Kytococcus sedentarius is important to annotate for several reasons. First, this gene has potential as a source of antibiotics such as monensin. In addition, studies show the ability of this bacterium to be infectious, including the evidence of causing pitted keratolysis, hemorrhagic pneumonia, and other opportunistic infections (Sims et al., 2009). Finally, the annotation of this genome is important because humans have not done extensive bioinformatics research on bacteria closely related to *Kytococcus sedentarius*, meaning that the work done doing the annotation is imperative to the understanding the genome, and to reaching the ultimate goal of having all known genes annotated.

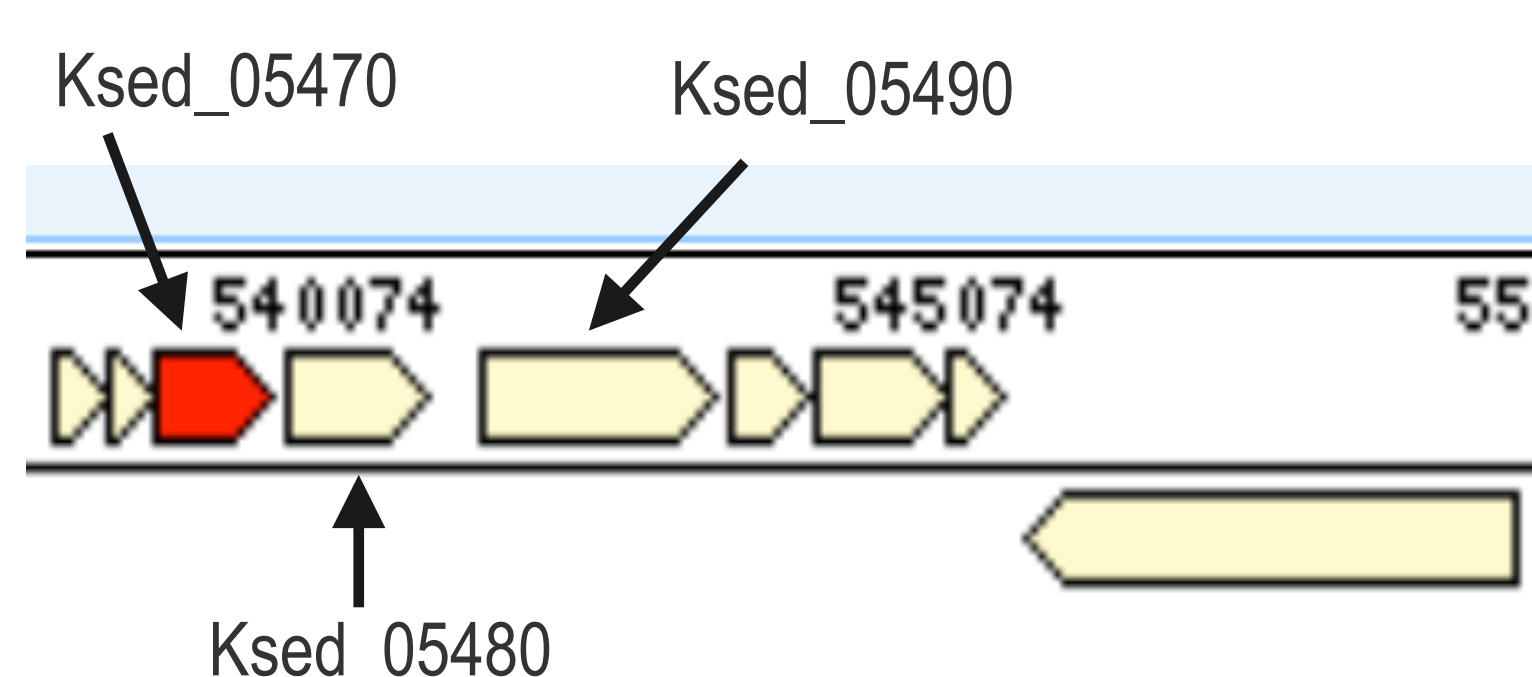


Figure I – Gene Neighborhood for Annotated Genes: Ksed_05470, Ksed_05480 and Ksed_05490

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	In what process does my protein take part?
Module 8- Evidence for Horizontal Gene Transfer	Phylogenetic Tree	Has my gene co-evolved with other genes in the genome?

Results

Kytococcus sedentarius 05470: The initial proposed product of this gene by GENI-ACT was a RNA polymerase sigma-70 factor. Using BLAST, it was found that this protein is most closely related to RNA polymerase sigma factor SigG, found in *Mycobacterium tuberculosis*. WebLogo showed low conservation of amino acids that were compared using CDD and T-Coffee. TMHMM shows that the protein has no transmembrane helices. SignalP shows the existence of signal proteins, and there were none found. Both of these resources show that the protein is found in the cytoplasm. To further support this evidence, PSORT-B stated that 9.97 was the cytoplasmic score, and .01 and .02 were the cell wall and extracellular scores, respectively. There is no Shine Dalgarno located in the sequence. With further investigation into enzymatic function, it was confirmed that the RNA polymerase protein is involved in transcription.

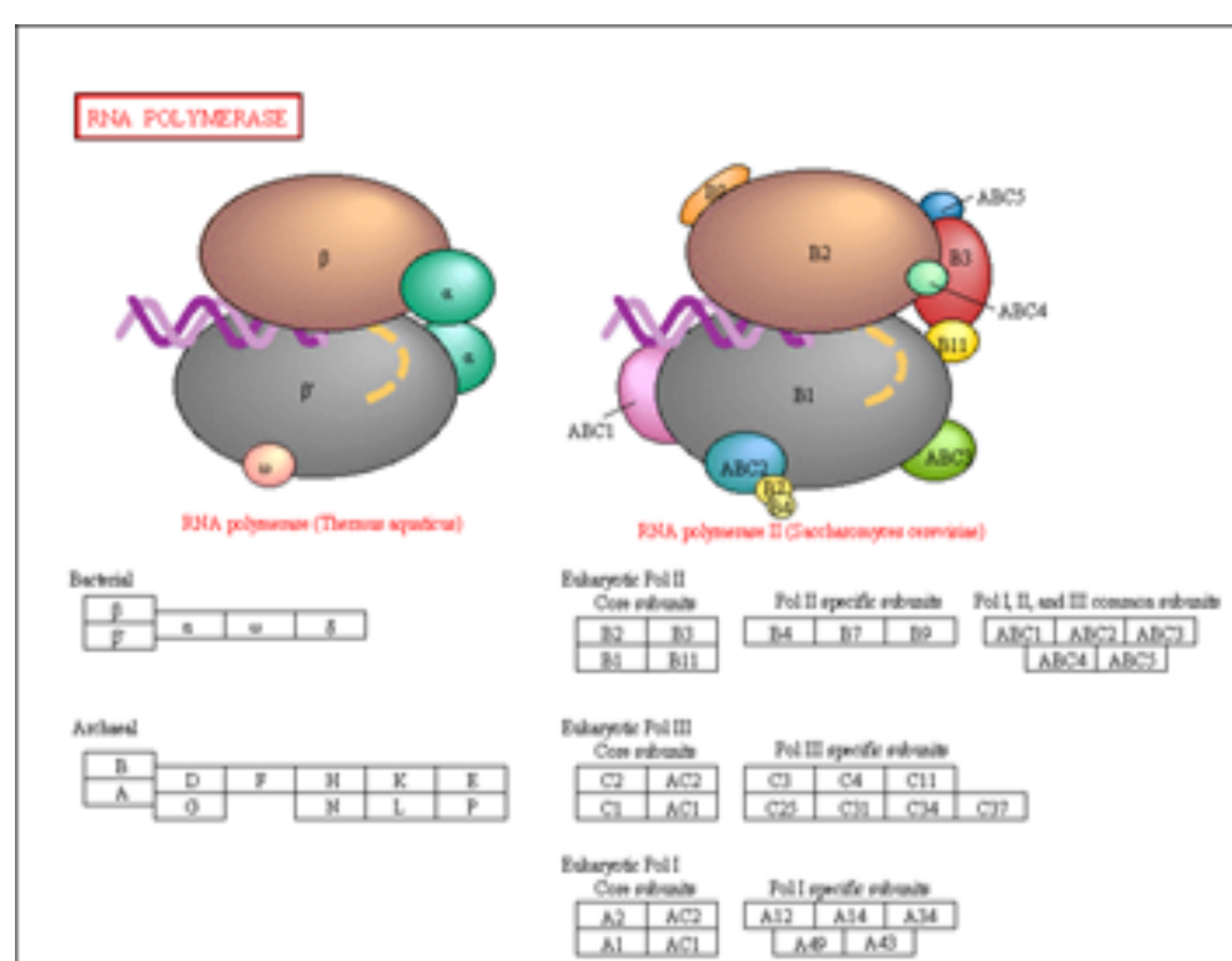


Figure II- KEGG image that shows RNA Polymerase

Kytococcus sedentarius 05480: Penicillin binding protein, beta-lactamase class C is an enzymatic protein used for metabolic processes. The BLAST program showed that this protein is closely related to *Mycobacterium tuberculosis*. In addition, the CDD results showed that the COG name of the protein is Beta-lactamase class C and other penicillin binding proteins, which matches the initial results given by the geni-act program. The next module was intended to locate the protein's position in the cell. The TMHMM program showed no transmembrane helices, indicating that this is not a membrane protein. Both tests from PSORTb and Phobius indicate that the protein is non-cytoplasmic and therefore extracellular, but the results are inconclusive because the SignalP results show that the protein does not show signal peptides. Further research found that this protein is an enzyme that assists metabolic function in the beta-Lactam Resistance pathway. This test and all previous modules conclude that the gene was properly annotated.

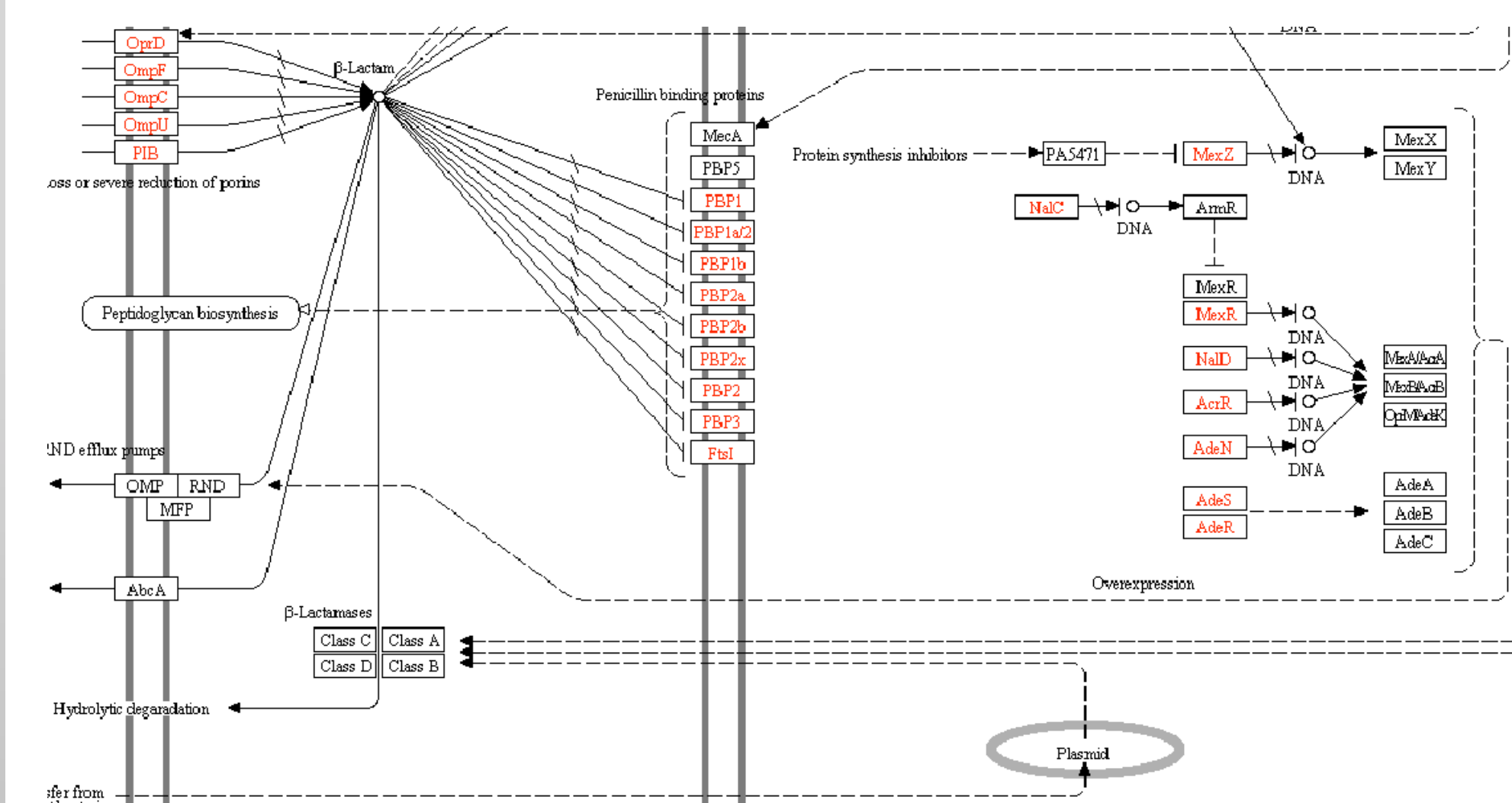


Figure III- The metabolic pathway from KEGG that denotes the enzymatic function of Beta-Lactamase class C

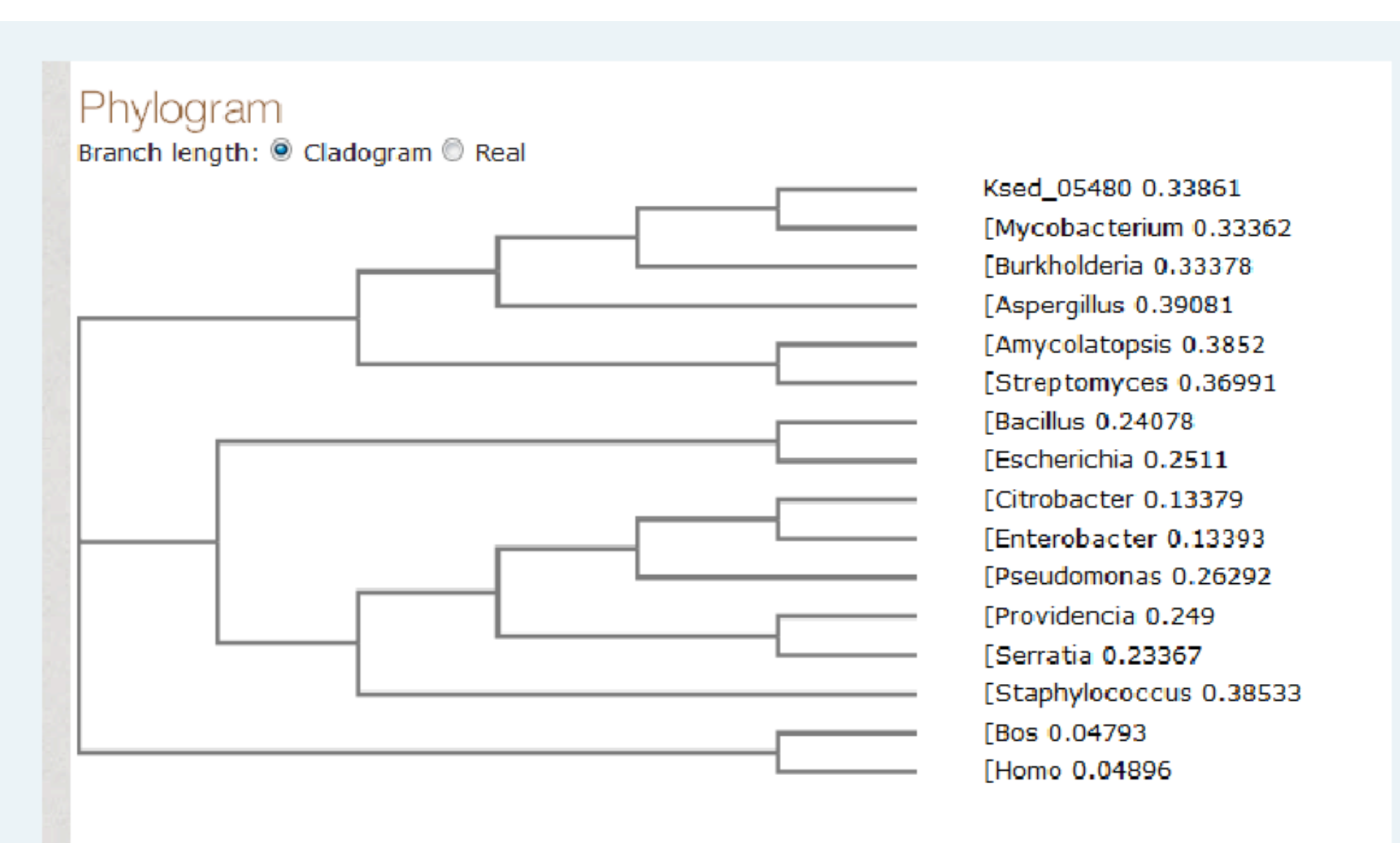


Figure IV- Phylogenetic tree of Ksed_05480 from T-Coffee that shows relationships to other organisms. This gene appears to be most similar to *Mycobacterium tuberculosis*, an infectious bacterium causing tuberculosis and leprosy.

Kytococcus sedentarius 05490:

This is a Heat Shock Chaperone Protein DnaK. After completing BLAST, it was concluded that this protein is closely related to a hypothetical protein found in a marine metagenome. The CDD results showed that the COG was a heat shock protein 70. The COG name matches the results that were given by the program. The TMHMM program indicated that there was no transmembrane helices and the SignalP results showed that there was no signal peptides. The PSORT-B and the Phobius results concluded that the protein was in the cytoplasm. It was concluded that the gene was properly annotated.

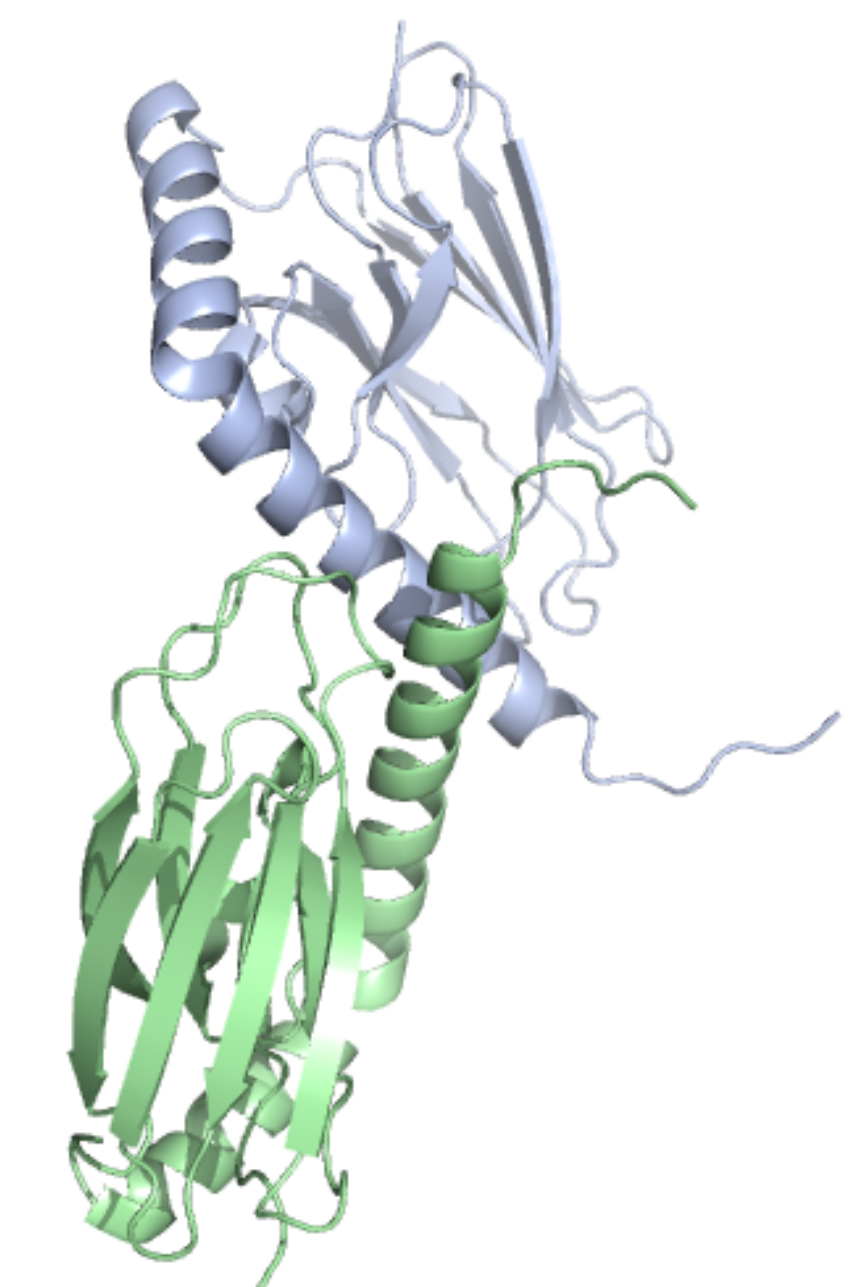


Figure V – The 3-dimensional structure of heat shock protein identified with locus tag Ksed_05490

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.

Gene Locus	Geni-Act Gene Products	Proposed Annotation
05470	RNA Polymerase sigma-70 factor	RNA Polymerase sigma-70 factor
05480	Penicillin binding protein, beta-lactamase class C	Penicillin binding protein, beta-lactamase class C
05490	Chaperone Protein DnaK	Chaperone Protein DnaK

References

Sims et al. (2009). Complete genome sequence of *Kytococcus sedentarius* type strain (541T). *Standards in Genomic Sciences*, 12 - 20.

Acknowledgments

Special thanks to Dr. Stephen Koury and Dr. Rama Dey-Rao for their assistance with this project. This work was supported by the National Science Foundation ITEST Strategies Award Number 1311902.