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

Annotation of the *Kytococcus sedentarius* Genome from Locus Tags Ksed_08380 to Ksed_08410

Adrian Bell, Wenyi Zeng, Jacob LaDue, Samuel Reeder, Brighid Smith and Betsy Vinton

The Harley School Department of Biological Sciences and the Western New York Genetics in Research Partnership

Abstract

A group of consecutive 4 genes from the microorganism *Kytococcus sedentarius* (Ksed_08340 -- Ksed_08350 -- Ksed_08360 -- Ksed_08370) were annotated using the collaborative genome annotation website GENI-ACT. The Genbank proposed gene product name for each gene was evaluated 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. For these 4 genes, the Genbank proposed gene product name did not differ significantly from the proposed gene annotation, suggesting that the genes were correctly annotated by the database.

Introduction

Kytococcus sedentarius is a strictly aerobic, non-motile, non-encapsulated, non-endospore forming, and 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).

In addition, recent research has revealed that microbes play a vital role in humans, as well as many other organisms. Trillions of microbes combine together to create efficient microbiomes that significantly aid and organism's metabolic processes and immune defense. Microbes and organisms have coevolved to benefit each other, creating a mutually beneficial relationship. In fact, the human body hosts more than 100 trillions bacterial and fungal cells, while there are only 30 million human cells. Therefore, it is no surprise that human health is largely attributed to the diversity and well-being of these microbes. (Missing Microbes, 2014)

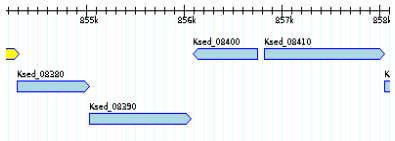

Figure I – *Kytococcus sedentarius* 08380, 08390, 08400, and 08410 vgene neighborhood.

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- Structure-Based Evidence	TIGRfam, Pfam, PDB	Are there functional domains in my protein?
Module 4- Cellular Localization Data	Gram Stain, TMHMM, SignalP, PSORT, Phobius	Is my protein in the cytoplasm, secreted or embedded in the membrane?
Module 5- 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?
Final Annotation	Review data from all modules	Does the student proposed name of the gene agree with that proposed by the automated computer annotation? Are any changes proposed to the pipeline annotation?

Results

***Kytococcus sedentarius* 08380:**

This locus contains DNA coordinates 854299 to 855033 as indicated by the GENI-ACT page. The product sequence is 244 base pairs long. A BLAST query, using the SWISSPRO Database in Genbank, gave a top hit of sequence similarity with Sensory Transduction Protein RegX3 from *Mycobacterium smegmatis*. The CDD Database resulted in two COG's; CheY chemotaxis protein or a CheY-like REC (receiver) domain [Signal transduction mechanisms]. and DNA-binding winged helix-turn-helix domain [Transcription]. The WEBLOGO made by using the multiple sequence alignment from T-Coffee showed that from amino acid 26 to 264, there were non-conserved residues. The TIGRfam analysis resulted in a suggested protein family of a heavy metal response regulator. A search of the PDB indicates that Ksed_08380 is structurally similar to the response regulator RegX3 from *Mycobacterium tuberculosis*. The PSORT-B database indicated that the protein is cytosolic. The TMHMM and Phobius databases suggested that the protein did not have any transmembrane helices and the SignalP database revealed that there were not any signal peptides present in the protein. An alternative open reading frame analysis indicates that the start codon may be at position 854350 instead of 854299.

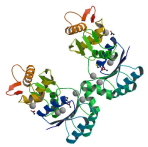

Figure II – Generated biological assembly from Protein Data Bank of the top result for *Kytococcus sedentarius* 08380 (PBD Code: 2OQR).

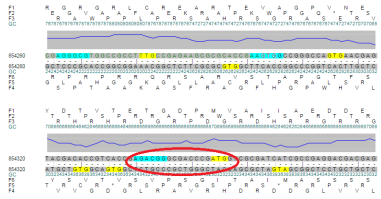

Figure III – An alternate reference frame search on *Kytococcus sedentarius* 08380 revealed another possible start codon, as indicated by the circled (red) Shine-Dalgarno region.

***Kytococcus sedentarius* 08390:**

This locus, Ksed_08390, included DNA coordinates 855034 to 856071, forwards. It consists of 1038 nucleotides, or 345 amino acids. According to BLAST, the top similarity sequence is a histidine kinase from *Phycicoccus* sp. Soil803, with an e-value of 2e-57. The second hit was another histidine kinase from *Phycicoccus* sp. Soil802, with an e value of 3e-57. ACDD search resulted in one COG, COG0642, named BaeS. The e-value is 2.25e-30. The WEGLOGO made from T_Coffee results showed a lack of amino acid residues from 42-108, and 417-425. Residues were sparse from 32-132. TMHMM detected two membrane helices, which was confirmed by Phobius. There was also a signal peptide detected by SignalP, with probability of 0.486. PSORT-B gave a final prediction of cytoplasmic membrane with a probability of 8.78. TIGRFAM gave a TIGRFAM number of TIGR02966, and the name phosphate regulon sensor kinase PhoR. The e-value was very high, 2.3e-05.

***Kytococcus sedentarius* 08400:**

There was no initial proposed product of this gene by GENI-ACT. This gene product proposal was supported by the top BLAST results (methylmalonyl-CoA carboxytransferase) of the NR database using the amino acid sequence. Pfam identified the gene as a GtrA-like protein, which is predicted to be integral membrane protein with three or four transmembrane spans. It is involved in the synthesis of cell surface polysaccharides. TMHMM indicated that there are four transmembrane helices, Signal IP revealed that there is no probability for signal peptide, PSORT-B predicted Ksed_08340 to be a cytoplasmic protein. Phobius shows four transmembrane helixes and no signal peptide is predicted in the SignalIP. Taking all of these results into account, the final prediction for localization is that Ksed_08400 is a transmembrane GtrA-like protein.

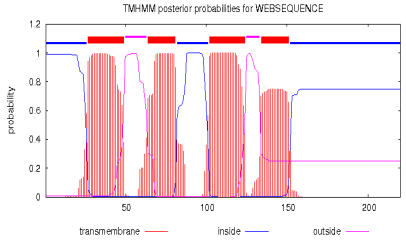

Figure IV – TMHMM determines four transmembrane helix domains exist in *Kytococcus sedentarius 08400*.

***Kytococcus sedentarius* 08410:**

This locus included DNA coordinates 856823 to 858055 on the forward strand as indicated by the GENI-ACT gene page. The gene product is 410 amino acids long. A BLAST query gave a top hit of sequence similarity with Phosphoribosylaminoimidazole carboxylase from Cryptococcus neoformans var. grubii H99 with an e-value of 2e-53. The second hit was also a phosphoribosylaminoimidazole carboxylase from another variant of Cryptococcus. The CDD database search resulted in one COG, (COG0026), a Phosphoribosylaminoimidazole carboxylase (NCAIR synthetase) used in nucleotide transport and metabolism. The WEBLOGO made from the T-Coffee multiple alignment results shows a cluster of non-conserved residues at positions 2-26 and 209-224. A TMHMM analysis did not detect any transmembrane helices. Future analyses could include: SignalP, PSORT-B, Phobius, TIGRFAM, Pfam, PDB, and Alternate Open Reading Frame.

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
08380	response regulator with CheY-like receiver domain and winged-helix DNA-binding domain	heavy metal response regulator
08390	signal transduction histidine kinase	phosphate regulon sensor kinase PhoR
08400	predicted membrane protein	GtrA-like protein
08410	phosphoribosylaminoimidazole carboxylase	Further investigation needed

References

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

Acknowledgments

Supported by NSF ITEST Strategies Award Number 1311902