Annotation of the *Kytococcus sedentarius* Genome from Locus Tags Ksed_09640 to Ksed_09730

Aaryn Schmitt, Bayana Burnell, Jeffrey Cantwell, Gideon Smith, Bailey DiSanto, Hannah Ferland, Ana Vargas and Dan Sherman Williamson Senior High School, Williamson, New York.





Abstract

A group of consecutive 6 genes from the microorganism Kytococcus sedentarius (Ksed_09640 – Ksed_09730) were annotated using the collaborative genome annotation website GENI-ACT. The Genbank proposed gene productname 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, enzymatic function and the possibility of horizontal gene transfer. 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 in the database. One unique finding was that Ksed_09700 appeared to be horizontally transferred to Kytococcus.

Introduction

Kytococcus sedentarius is a strictly aerobic, non-motile, nonencapsulated, 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/y).

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).

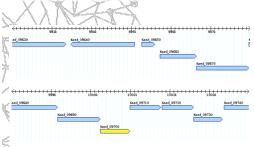


Figure 1. The locus tags and relative position of the genes under investigation in this research.

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?

Results

Ksed 09640:

The initial proposed product of this gene by GENI-ACT was Catalase. By running BLAST and analyzing its top hits, this proposal made by GENI-ACT was supported. The proposal was also supported by the presence of well-curated protein functional domains within the amino acid sequence, the cytoplasmic topography of the amino acid sequence and the cellular location of the amino acid sequence. Therefore, the proposed annotation was confirmed to be Catalase.

Ksed 09650:

The initial proposed product of this gene by GENI-ACT was a DNA-binding transcriptional regulator. This gene product was supported by the top BLAST hits for the amino acid sequence, the presence of well-curated protein functional domains within the amino acid sequence, the cytoplasmic topography of the amino acid sequence, and the cellular location of the amino acid sequence. This gene product proposal is also supported by the identification of Ksed_09700 and Ksed_09710 as Imidazole genes. As such, the proposed annotation is a DNA-binding transcriptional regulator.

(sad 09660·

The initial proposed product of this gene was a ATP phosphoribosyltransferase. This was supported by the top BLAST hits for the amino acid sequence, the presence of well-curated functional domains within the amino acid sequence, the transmembrane topography of the amino acid sequence, the cellular location of the amino acid sequence, and the enzymatic function of the amino acid sequence. As such the proposed amountation is a ATP phosphoribosyltransferase catelyzing a chemical reaction.

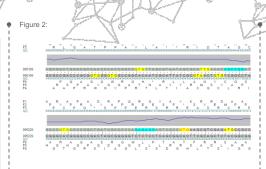


Figure 2 - Ksed_9650 DNA Coordinates. Noting that the start codon is found directly downstream of the Shine Dalgamo region, and using T COFFEE as a reference, it can be assumed that there are no major disruptions or miscalculations in the sequence.

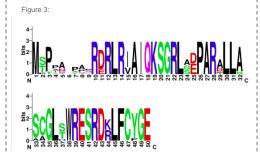


Figure 3 – Ksed_09660 T COFFEE results. The majority of the graph shows a conservation of the sequence when compared with a gene of known function.

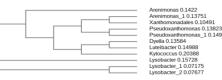


Figure 4 Ksed_09700 phylogenetic tree. Notice Lysobacter and Wooccus, the lineage indicated norizontal transfer.

Ksed 09700:

The initial proposed product of this gene by GENI-ACT was an Imidazole glycerol phosphate synthase. This gene product proposal was supported by the top BLAST hits for the arhino acid sequence, the presence of well-curated protein functional domains within the amino acid sequence, the cytoplasmic topography of the amino acid sequence, and the cellular location of the amino acid sequence. As such, the proposed annotation is an Imidazole glycerol phosphate synthase, unchanged from the initial proposed product. There is strong evidence for horizontal transfer from a Lysobacter bacterium relative (see Figure 4).

Ksed 09710:

The initial proposed product of this gene by GENI-ACT was an imidazole-4-carboxamide isomerase. This gene product proposal was supported by the top BLAST hits for the amino acid sequence, the WebLogo correlation results, the signal peptide prediction results, and the location on the pathway map. As such the proposed annotation is an imidazole-4-carboxamide isomerase, unchanged from the initial proposed product.

Ksed 09730:

The initial proposed product of this gene by GENLACT was a histidine biosynthesis bifunctional protein. By running BLAST and analyzing its top hits, this proposal made by GENLACT was supported. The proposal was also supported by the presence of well-curated protein functional domains within the amino acid sequence, the cytoplasmic topography of the amino acid sequence and the cellular location of the amino acid sequence. Therefore, the proposed annotation was able to be deduced as a histidine biosynthesis bifunctional protein.

Conclusion

The GENI-ACT proposed gene product and not differ for any of the proposed gene annotation of the genes in the group and as such, the genes appear to be correctly annotated by the computer database. Looking at the annotated genes as a whole, however, it was found that Ksed 9640, database related immune-responsive domain, does not seem to associate with the rest of the gene that we annotated.

	/\ /\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\	
Gene Locus	Geni- Act Gene Products	Proposed Annotation
9640	Catalase	Catalase
9650	DNA-binding transcriptional regulator	DNA-binding transcriptional regulator
9660	ATP phosphoribosyltransferase	ATP phosphoribosyltransferase
9700	Imidazole glycerol phosphate synthase	Imidazole glycerol phosphate synthase
9710	Imidazole -4-carboxamide isomerase	Imidazole -4-carboxamide isomerase
9730	histidine biosynthesis bifunctional protein	histidine biosynthesis bifunctional protein

References

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

Acknowledgments

Supported by an NIH Science Education Partnership (SEPA) Award - R25ODO10536

www.buffalo.edu