Annotation of the *Kytococcus sedentarius* Genome at Locus Tags KSED_RS14230, KSED_RS14245 and KSED_RS13180

Tessa DeCicco, Iris Izydorczak, Ha Lin Blair Kim, and Dr. Paul A. Weber
Amherst Central High School, Amherst, NY and the Western New York Genetics in Research and Health Care Partnership

Abstract

A group of 3 genes from the microorganism *Kytococcus sedentarius*(KSED_RS14230,KSED_RS14245,KSED_RS13180) 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, enzymatic function, presence or absence of gene duplication and degradation, the possibility of horizontal gene transfer, and the production of an RNA product.

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).
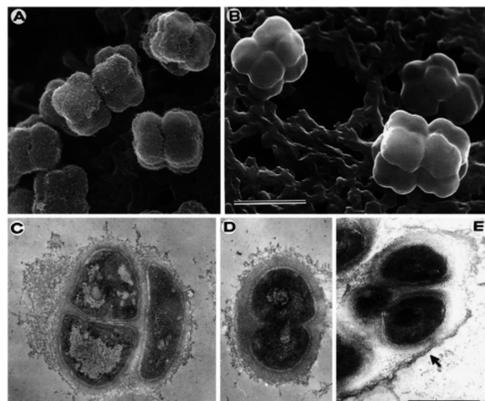

Figure 1. Scanning (A and B) and Transmission (C-E) electron micrographs of *Kytococcus sedentarius*.

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
Basic Information	DNA Coordinates and Sequence, Protein Sequence	What is the sequence of the gene and protein? Where is it located in the genome?
Sequence-Based Similarity	Blast, CDD, T-Coffee, WebLogo	How similar is the protein under investigation to other proteins in GenBank?
Structure-Based Similarity	TIGRfam, Pfam, PDB	What functional domains are present in the protein under investigation?
Cellular Localization	Gram Stain, TMHMM, SignalP, LipoP, Psortb, Phobius	Is the protein under investigation located in the cytoplasm, secreted, in the periplasm or embedded in the cell membrane or cell wall?
Enzymatic Function	KEGG, MetaCyc, E.C. Number	In what process or structure is the protein under investigation involved?
Duplication and Degradation	Paralog, Pseudogene	Are there other forms of the protein under investigation in the same genome? Is it functional?
Horizontal Gene Transfer	Phylogenetic Tree, Gene Neighborhood, GC Content	Has the protein under investigation co-evolved with the rest of the genome or has it been obtained in a different way?
RNA family	Rfam	Does the gene under investigation encode a functional RNA?
Final Annotation	Evaluate data from all modules	Has the gene been correctly called by the pipeline annotation?

Results

KSED_RS13180: The data provided by BLASTp analyses suggests that this gene is a protein, as GENI-ACT predicted. The BLASTp data analysis predicts that this protein is to combine with a large RNA subunit to pair mRNA codons to their tRNA anticodons. Figure 4 displays a Signal P graph indicating that this gene product does not contain a signal peptide.

KSED_RS14230:
The initial proposed product of this gene by GENI-ACT was a protein closely related to integrase or transposase. These proteins are both used to take out DNA from one strand and put it in another. This gene product proposal was supported by the top NR BLAST database (figure 3) hits for the amino acid sequence. The SWISSPROT BLAST database was also searched, but very few sequence comparison alignments were found. Those hits appearing in the SWISSPROT database possessed low E-values and as a result were not considered (data not shown).

KSED_RS14245:
Due to research using BLAST CDD was a protein involved in the metabolism and recombination of DNA. This gene is very similar to other genes with this function such as Hypothetical protein [Brevibacterium aurantiacum] This gene product proposal was supported by the top BLAST hits for the amino acid sequences, the presence of well-curated functional domains within the amino acid sequences, the cellular location of the amino acid sequences, and the enzymatic function of the amino acid sequences.
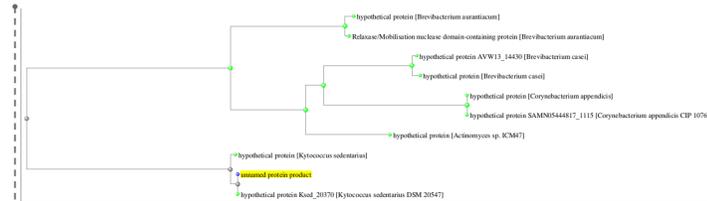

Figure 2 (above) - KSED_RS14245 Distance of tree results, This gene is closely related to the genes above as shown.

Description	Max score	Total score	Query cover	E value	Ident
hypothetical protein [Kytococcus sedentarius]	124	124	100%	3e-36	98%
hypothetical protein [Kytococcus sedentarius]	120	120	100%	1e-34	95%
Transposase InsO and inactivated derivatives [Kytococcus aerolatus]	124	124	100%	4e-33	95%
integrase family protein [Kytococcus sedentarius DSM 20547]	124	124	98%	4e-33	97%
IS481 family transposase [Kytococcus schroeteri]	124	124	100%	5e-33	97%
IS481 family transposase [Kytococcus aerolatus]	124	124	100%	5e-33	95%
IS481 family transposase [Kytococcus sedentarius]	124	124	98%	5e-33	97%
integrase [Micrococcus luteus]	111	111	100%	3e-30	87%
Integrase core domain-containing protein [Micrococcus luteus]	112	112	100%	4e-30	87%
Mobile element protein [Brachybacterium faecium]	115	115	100%	1e-29	85%
transposase [Microbacterium sp. Aq1]	114	114	98%	2e-29	88%

Figure 3 – KSED_RS14230 has top hits with transposase and integrase that transfer and integrate DNA from one strand to another.
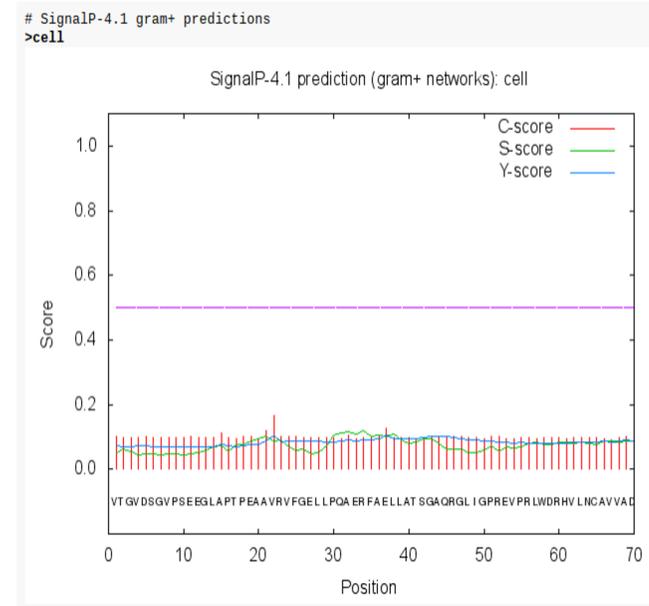

Figure 4 – KSED_13180, results for signal peptide test

Conclusion

KSED_RS13180:
GENI-ACT annotated this gene as a protein. The top BLASTp and Swissport hits further suggest that this gene is a small RNA subunit, involved in pairing mRNA codons with their tRNA anticodons. Figure 4's low C-scores further support this by showing that there is no signal peptide in this gene, implying that activity happens in the cytoplasm, where mRNA and tRNA work. Therefore, this gene is likely involved in protein synthesis.

KSED_RS14230:
The hypothetical protein has top hits with many varieties of both transposase and integrase through conducting the BLAST process. These proteins are responsible for removing, transferring, and integrating DNA from one strand to another. Therefore, the proposed job of the protein may be a protein that works with moving DNA out of a strand and into another. Ultimately, BLAST supported this prediction through the NR database by showing many integrase and transposase hits at the top of the results (figure 3).

KSED_RS14245:
The initial proposed product of this gene was an enzyme involved in the metabolism and recombination of DNA. This gene product proposal 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.

Locus Tag	Pipeline Annotation Product Name	Proposed Annotation	Changes Proposed?
KSED_14230	Hypothetical Protein	Transposase or Integrase Protein	Yes
KSED_14245	Hypothetical Protein	Protein involved in DNA recombination	
KSED_13180	Hypothetical Protein	Small RNA subunit	Yes

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.

References

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