

Annotation of the *Kytococcus sedentarius* Genome at Locus Tag KSED_RS00555

Sam Erwin, William Michalak, Annika Straight, Niamh Schroeder, Emily Stowell 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 four consecutive genes from the microorganism *Kytococcus sedentarius* (KSED-RS00555) were annotated using the collaborative genome annotation website GENI-ACT. The GenBank proposed gene product-name for each gene. All genes were 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. 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 in the database.

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).

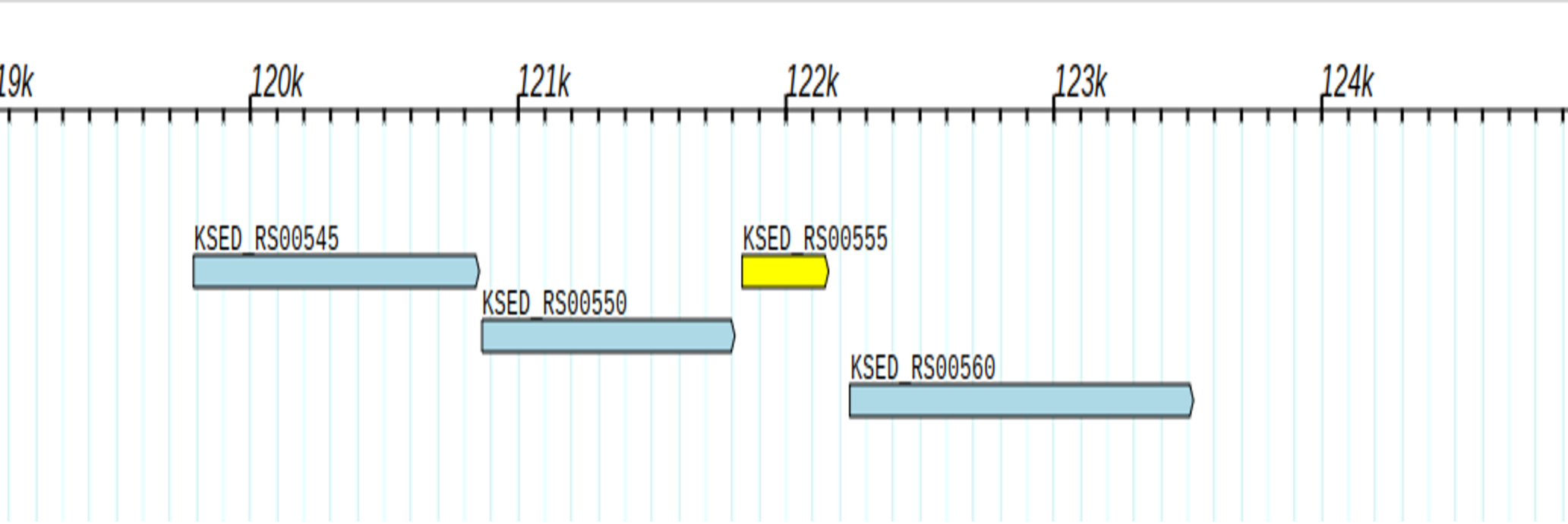


Figure 1. The locus tags and relative position of the genes under investigation in this research

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
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

There is data supporting the idea that KSED_RS00555 is a sulfurtransferase. The information and data we collected strongly linked this to the Rhodanese like domain. The method used to identify KSED_RS00555 as a sulfurtransferase was a compositional matrix adjust using a Basic Local Alignment Search Tool, henceforth referred to as BLAST, which matched our gene to many similar genes, including the protein from the bacteria below, *Kytococcus aerolatus* (see Figure 2). These two genes had an E-value of 2e-50, a 74-88% overall match rate and 164bits, indicating that they are very similar, and likely have the similar (or possibly the same) functions.

Download v GenPept Graphics

rhodanese-like domain-containing protein [Kytococcus aerolatus]

Sequence ID: WP_088818361.1 Length: 108 Number of Matches: 1

Range 1: 5 to 108 GenPept Graphics

Score	Expect	Method	Identities	Positives	Gaps
164 bits (416)	2e-50	Compositional matrix adjust.	77/104(74%)	92/104(88%)	0/104(0%)
Query 4	ITPQDLRDDAVLDVREPEEWAAGHAPNAVHIPLADVPARVDELDPDSTGDLPLVTCRSG	63			
Sbjct 5	ITPQDLRDDA+++DVREP+E+AAGHAPNAVHIPLDVARYELP+PLP+CRSG				
Query 64	GRSSRAVQWLQAQGYEVVNVDDGGMKAWSAAGKQVVSADAGEPQVI	107			
Sbjct 65	GRSRAVQWLQAQGYEVVNVDDGGMKAWSAAGKQVVSADAGEPQVI	108			

Figure 2 - Relationship chart revealing the relationship between *Kytococcus aerolatus* and *Kytococcus sedentarius*.

	Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
1	rhodanese-like domain-containing protein [Kytococcus sedentarius]	214	214	100%	4e-70	100.00%	WP_012801626.1
2	MULTISPECIES rhodanese-like domain-containing protein [Kytococcus]	167	167	100%	6e-52	76.64%	WP_070705276.1
3	Rhodanese-related sulfurtransferase [Kytococcus aerolatus]	165	165	97%	1e-50	74.04%	SW64692.1
4	rhodanese-like domain-containing protein [Kytococcus aerolatus]	164	164	97%	2e-50	74.04%	WP_088818361.1
5	rhodanese-like domain-containing protein [Kytococcus aerolatus]	148	148	96%	4e-44	72.12%	WP_050671451.1
6	rhodanese-like domain-containing protein [Kytococcus aerolatus]	147	147	88%	1e-43	77.89%	WP_052589318.1
7	rhodanese-like domain-containing protein [Kytococcus aerolatus]	144	144	98%	3e-42	65.09%	WP_007180988.1
8	rhodanese-like domain-containing protein [Kytococcus aerolatus]	143	143	98%	4e-42	65.09%	WP_122261089.1
9	rhodanese-like domain-containing protein [Kytococcus aerolatus]	141	141	97%	2e-41	67.92%	WP_130014385.1

Figure 3 – Ksed_00555 with the most closely related amino acid sequences.



Figure 4 - A color coded alignment chart for Ksed_00555.

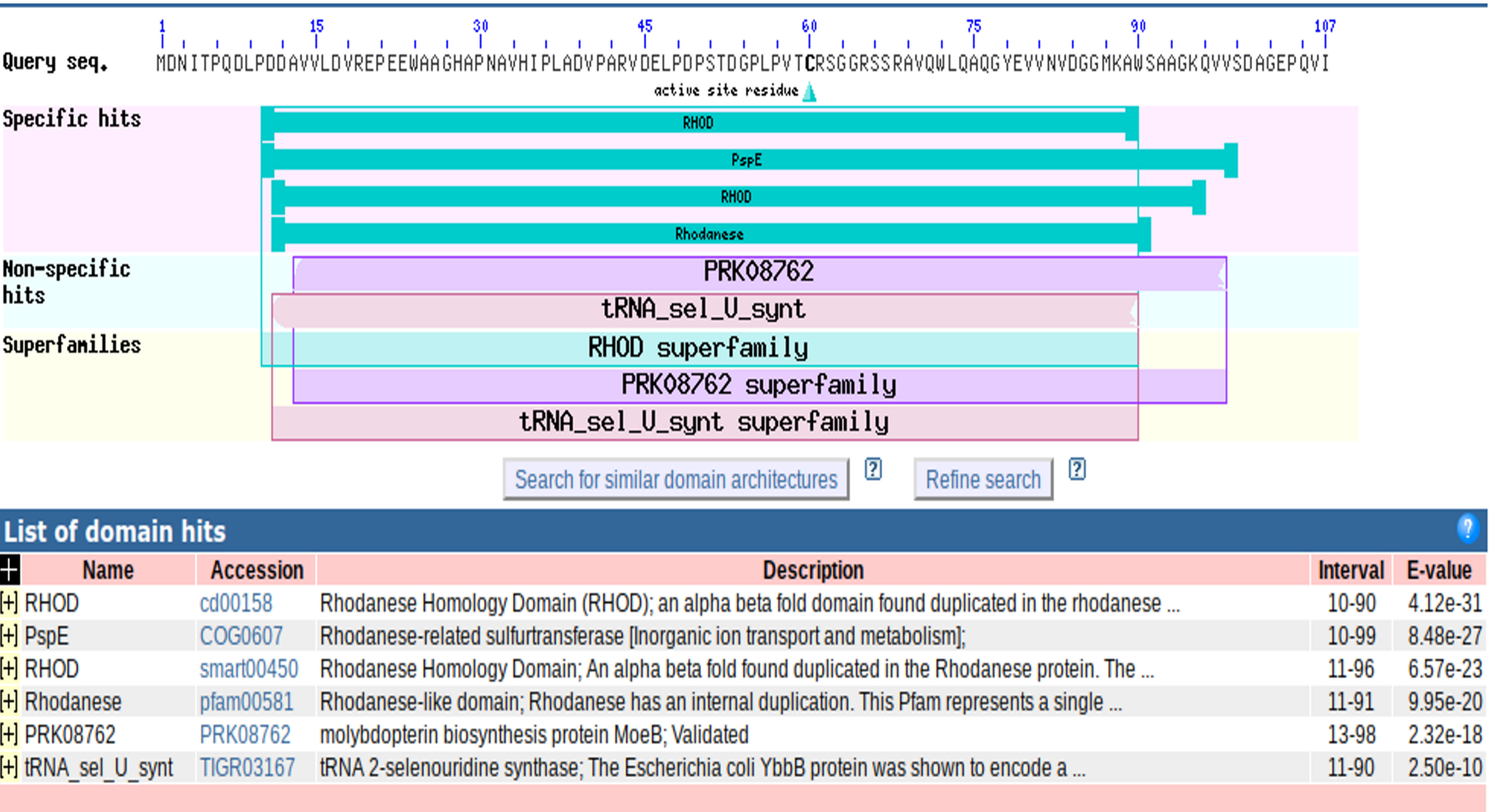


Figure 5 – Ksed_00555 well-curated functional domains which are incomplete at the C-terminal end of the amino acid sequence.

Figure 1: A graph which reveals the Locus tag of our gene and the overall size of KSED_RS00555. In comparison to other genes ours is short in length.

Figure 2: A graphic organizer which works to display the similarities and differences of Kytococcus Sedentarius, KSED_RS00555 and Kytococcus Aerolatus. From this we can further estimate and predict what may have been the function of our gene. Based off of its similarities with K.A. and the Rhodanese like domain it is very apparent that our gene is a sulfurtransferase.

Figure 3: A chart which shows the genes that are most closely related to our gene. As shown the chart the genes most closely related are mostly within the Rhodanese domain or also a Kytococcus gene. This evidence further supports the proposed function

Figure 4: A chart that shows with colors how closely the amino acid sequence of Kytococcus is aligned with other amino acid sequences. From this we can gather (based off of size, length, shape) what might have been the function of KSED_RS00555.

Figure 5: This chart shows the domains of Kytococcus, their sequences, and their E-value. From this information we are able to infer what the whole of the amino acid is used for.

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. The gene KSED_RS00555 is still believed to be a sulfurtransferase and cyanide detoxifier due to the data which supported this proposed function.

References

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

Acknowledgments

Supported by an NIH Science Education Partnership (SEPA) Award - R25GM129209