

Annotation of the *Kytococcus sedentarius* Genome from DNA Coordinates 907975 to 911269

Elizabeth Hartley, Joshua Smith, Sierra Tuller and Paula Ferneza
Frewsburg Jr.-Sr. High School and the Western New York Genetics in Research Partnership

Abstract

A group of 3 consecutive genes from the microorganism *Kytococcus sedentarius* (Ksed_08890 – Ksed_08910) 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, and structure-based evidence from the amino acid sequence. The Genbank proposed gene product name did not differ significantly from the proposed gene annotation for 2 of these genes (based on Modules 1-3), but we propose that one (Ksed_08900) should have its annotation changed to energy-dependent translational throtle protein EItA based on the three top hits in the Swiss-Prot database for Ksed_08900.

Introduction

Kytococcus sedentarius is a bacterium that is capable of degrading human callus. *K. sedentarius* is a marine dwelling gram positive, that is found in irregular clusters and pads of eight. *K. sedentarius* performs aerobic respiration, meaning they require free oxygen to create ATP.

According to C. M. Longshaw, this bacterium is of scientific interest because it has the unique ability to breakdown keratin. Keratin is found in callus, hair, nails, feathers, and wool and is highly resistant to being broken down due to its disulphide bonds. It is also believed that *K. sedentarius* can be contributed to pitted keratolysis. Pitted keratolysis is the appearance of circular shallow pits on a person's feet or palms, most often due to excessive sweating. Although pitted keratolysis is normally not harmful to the person afflicted, it is commonly considered unsightly and can cause an unwanted odor of the feet.

Kytococcus sedentarius are bacteria with a genome containing roughly 2,785,024 nucleotides, 2,554 protein genes, and 55 RNA genes. Even with all of this information held within its genome, this species has one of the smaller microbial genomes ever curated. Interest in the genome is partially due to its membership in a poorly studied family located in the actinobacterial suborder Micrococcales and from its ability to produce oligoketide antibiotics.

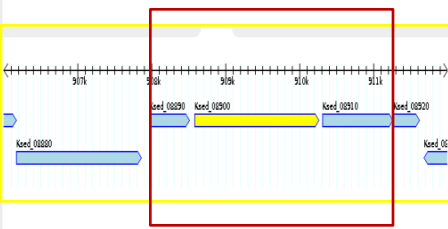


Figure 1: Ksed 08890, Ksed 08900, and Ksed 08910 gene neighborhood of the three genes investigated.

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?

Results

Kytococcus sedentarius 08890:

The initial proposed product of this gene by GENI-ACT was a single stranded DNA-binding protein. This gene product proposal was supported by the top BLAST hit for the amino acid sequence in the Swiss-Prot and nr databases, the presence of well-curated functional domains within the amino acid sequence, and the TGRFAM and Pfam results (Figure 2). As such, the proposed annotation from modules 1-3 is a single stranded DNA-binding protein. The single-stranded DNA-binding protein (SSB) of *Escherichia coli* is involved in all aspects of DNA metabolism: replication, repair, and recombination (4).

Kytococcus sedentarius 08900:

The initial proposed product of this gene by GENI-ACT was an ATPase component of ABC transporters with duplicated ATPase domain. The top two blast hits for the amino acid sequence indicated an energy-dependent translational throtle protein EItA protein (Figures 3 and 4). EItA is an energy-dependent translational throtle that gates ribosome entry into the elongation cycle, sensitive to ATP/ADP ratio because the ATP-bound form of EItA binds to the ribosomal tRNA-exit site and restricts ribosomal and tRNA movements, inhibiting protein synthesis; EItA is a prokaryotic ABC-F family member defined by two ABC transporter-related predicted ATP-binding domains (3) TIGRFAM and Pfam both indicated that this gene codes for ABC binding transporters. This data suggests that Ksed_08890 should be annotated as EItA.

Kytococcus sedentarius 08910:

The initial proposed product of this gene by GENI-ACT was a 23S RNA-specific pseudouridylylate synthase. This gene product proposal was supported by the top nr and Swiss-Prot BLAST hits for the amino acid sequence, the presence of well-curated functional domains within the amino acid sequences, and the TIGRFAM and Pfam results (Figures 4 and 5). As such, the proposed annotation from modules 1-3 is a 23S RNA-specific pseudouridylylate synthase. RluA is a dual-specificity enzyme responsible for pseudouridylylating 23S rRNA and several RNAs (5). Pseudouridine is an isomer of the nucleoside uridine, which is the most prevalent of the over one hundred different modified nucleosides found in RNA (6).

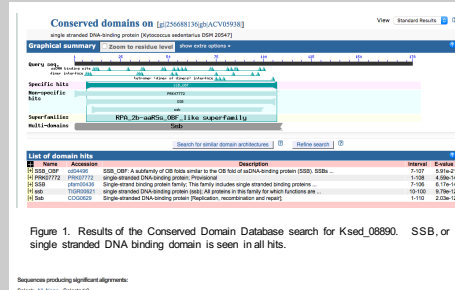


Figure 1. Results of the Conserved Domain Database search for Ksed_08890. SSB, or single stranded DNA binding domain is seen in all hits.

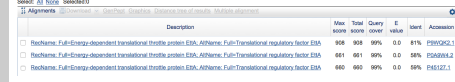


Figure 2. The top Swiss-Prot blast hits for Ksed_08900. All have low E values and high scores, strongly suggesting that Ksed_08900 should be annotated as Energy-dependent translational throtle protein EItA.



Figure 3. The top Swiss-Prot database blast hit for Ksed_08900. This clearly suggests Ksed_08900 should be annotated as Energy-dependent translational throtle protein EItA.

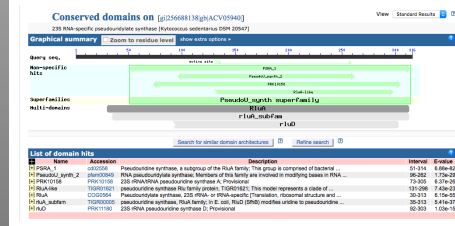


Figure 4. The Conserved Domain Database search results for Ksed_08910. COG, TIGRFAM and Pfam hits all suggest Ksed_08910 is a RluA family RNA-specific pseudouridylylate synthase.

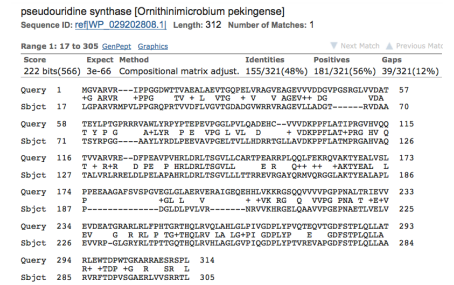


Figure 5. The top nr nrnt hit for Ksed_08910.

Conclusions

The Genbank proposed gene product name did not differ significantly from the proposed gene annotation for 2 of these genes (based on Modules 1-3), but we propose that one (Ksed_08900) should have its annotation changed to energy-dependent translational throtle protein EItA based on the three top hits in the Swiss-Prot database for Ksed_08900.

Gene Locus	Geni-Act Gene Products	Proposed Annotation
08890	Single stranded DNA binding protein	Single stranded DNA binding protein
08900	ATPase component of ABC transporters with duplicated ATPase domain	Energy-dependent translational throtle protein EItA
08910	23S RNA-specific pseudouridylylate synthase	23S RNA-specific pseudouridylylate synthase

References

- Longshaw, C. M., et al. (2002) Journal of Applied Microbiology. 93:5: 810-816.
- Sims et al. (2009). Complete genome sequence of *Kytococcus sedentarius* type strain (5417). *Standards in Genomic Sciences*, 12 - 20.
- <http://eocgene.org/?q=gene/EG12343>
- Microbiol Rev. 1990 Dec;54(4):342-80.
- Mol Cell. 2006 Nov;17(24):535-45.
- <https://en.wikipedia.org/wiki/Pseudouridine>

Acknowledgments

Supported by NSF ITEST Strategies Award Number 1311902