

Annotation of the *Kytococcus sedentarius* Genome from DNA Coordinates 209 to 1729 (Locus Tags Ksed_00010 or KSED_RS00005)

Joanna Fiorella, Bailee Hauck, Mursal Mohammad Akbar, Emily Schofield, Rolfe Friedenbergl and Robert O'Connor
 Kenmore East Senior High School, Tonawanda, New York and The Western New York Genetics in Research and Health Care Partnership



Abstract

A gene from the microorganism *Kytococcus sedentarius* (Ksed_00010) was annotated using the collaborative genome annotation website GENI-ACT. GenBank proposed the gene's product as a Chromosomal Replication Initiator Protein dnaA in terms of the general genomic information, amino acid sequence-based similarity data, structure-based evidence from the amino acid sequence, cellular localization data, enzymatic function, presence or absence of gene duplication and degradation, and the possibility of horizontal gene transfer. The GenBank proposed gene product name did not change significantly from the research that we conducted, and the gene appears to be correctly annotated by the database.

Introduction

Kytococcus sedentarius is an aerobic, non-motile, gram positive bacterium, found predominantly in tetrad formation. It is of interest as it has the potential to become a natural source of antibiotics. In fact, this bacterium is a natural source of the polyketide antibiotics monensin A and monensin B (Sims et al., 2009). In a majority of studies, *Kytococcus sedentarius* has been shown to be resistant to most forms of antibiotics such as penicillin. While it is relatively isolated from human skin, found primarily in marine environments, it can still cause strains of disease in humans (see Figure 2). *Kytococcus sedentarius* can infect the body through chemotherapy treatment and then go on to cause infections ("Annals of Hematology", 2004). *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). Furthermore, the phylogeny of this microorganism is a source of interest, as it is a member of the family Dermacoccaceae within the actinobacterial suborder Micrococccineae, which has recently been thoroughly studied (Sims et al., 2009).

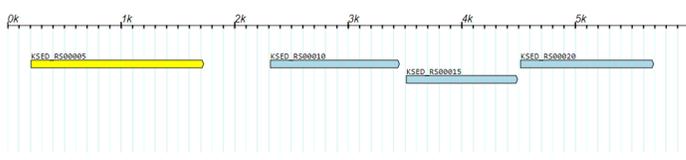


Figure 1 – Location of Ksed_00010

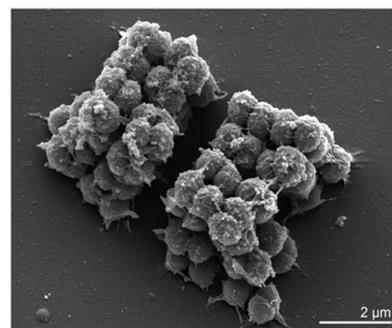


Figure 1 – The image to the left shows a microscopic visual of *Kytococcus sedentarius*.



Figure 2- The image to the right shows *Kytococcus sedentarius*' effect on a foot, resulting in pitted keratolysis.

Methods

Modules of the GENI-ACT (<http://www.geni-act.org>) were used to complete the *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?
Final Annotation	Evaluate data from all modules	Has the gene been correctly called by the pipeline annotation?

Results

Ksed_00010:

The computer pipeline proposed product of this gene was a Chromosomal Replication Initiator Protein dnaA. This gene product proposal was supported by the isoelectric point calculator, the top BLAST hits for the amino acid sequence, T-Coffee, PFAM, InterProScan, SignalP, LipoP, TMHMM, IMG/EDU Gene Finder, KEGG, and the Phylogenetic tree. The amino acid sequence was used to compare proteins, and BLAST assisted us in researching any matches. *Kytococcus sedentarius* shared a common ancestor with *Acidobacteria*, *Ornithinimicrobium*, *Arseicoccus*, and *Serinicoccus*. The results shown beneath *Kytococcus sedentarius* in the tree have no significant evolutionary relationships to our gene. KEGG illustrates the section in the biochemical pathway in which our gene functions.

Results for job tcoffee-I20190430-160540-0132-53472330-p1m

Alignments | Result Summary | Guide Tree | **Phylogenetic Tree** | Submission Details

Phylogenetic Tree

This is a Neighbour-joining tree without distance corrections.

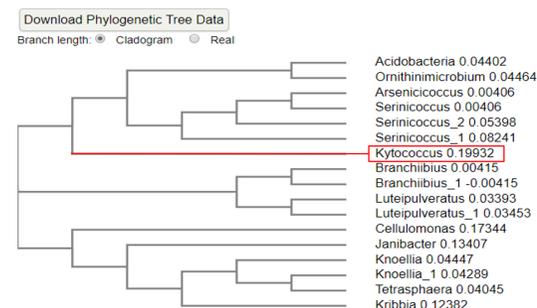


Figure 3 – The image above shows the phylogenetic tree of Ksed

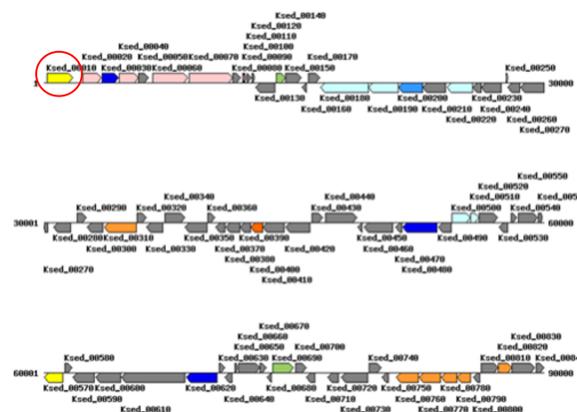


Figure 4 – The pathway in which the gene participates (KEGG)

SSDB Motif Search Result

Motif Id	From	To	Definition	E value	Score
pf:Bac_DnaA	164	381	Bacterial dnaA protein	1.4e-98	-
pf:1stB_1S21	164	306	1stB-like ATP binding protein	1e-13	-
pf:AAA_16	191	224	AAA ATPase domain	0.12	-
pf:AAA_19	194	224	AAA domain	0.33	-
pf:ATPase_2	196	301	ATPase domain predominantly from Archaea	0.0061	-
pf:AAA_14	198	318	AAA domain	0.0013	-
pf:RNA_helicase	200	300	RNA helicase	0.0067	-
pf:AAA_24	200	270	AAA domain	0.023	-
pf:AAA	201	316	ATPase family associated with various cellular activities (AAA)	1.5e-07	-
pf:AAA_21	201	299	AAA domain, putative AbiEII toxin, Type IV TA system	0.13	-
pf:AAA_22	235	296	AAA domain	0.19	-
pf:Bac_DnaA_C	410	477	Bacterial dnaA protein helix-turn-helix	1.4e-29	-

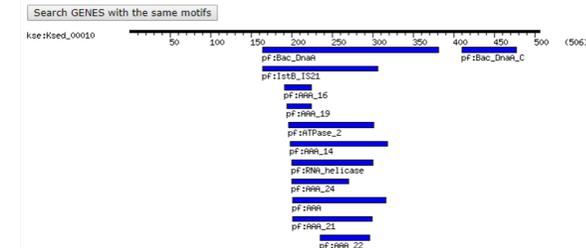


Figure 5 – The figure above is a visualization of differing genes that are similar to Ksed_00010

Conclusion

Based on the series of tests performed on the variety of databases, the conclusion that the online sources predicted was supported. 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. In addition, it did not differ from our hypothesis as well. Our group hypothesized that the gene shared commonalities to a chromosomal replication initiator protein dnaA as it was proposed by the results of GenBank. This specific gene is inferred to initiate DNA replication in bacteria as it promotes the unwinding of DNA at the origin of replication. This annotation is supported by our BLAST hits, as well as the KEGG pathway, which displayed the name and structure of this particular locus of *Kytococcus sedentarius* to a chromosomal replication initiator protein dnaA.

References

- pdfs.semanticscholar.org
- bacmap.wishartlab.com
- ghr.nlm.nih.gov
- link.springer.com,
- standardsingenomics.biomedcentral.com
- science.gov