

Annotation of the *Kytococcus sedentarius* Genome from DNA Coordinates 635832-645403

Hawraa Ahmed, Xionte Donalson, Tazira Allison-Johnson, Gloria Shweka, Michael Long, Josiah Alexander, Jennie Scungio, August Privitera and Deborah Grine

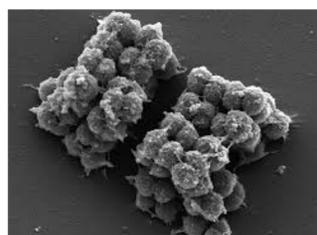
Leonardo DaVinci High School and the Western New York Genetics in Research Partnership

ABSTRACT

Seven genes (Ksed_06430, Ksed_06440, Ksed_06460, Ksed_06470, Ksed_06480, Ksed_06490, Ksed_06520) from the microorganism *Kytococcus sedentarius* were assigned for annotation between the DNA coordinates 635832-645403. These seven genes were annotated using several modules in GENI-ACT (Genomics Educational National Initiative-Annotation Collaboration Toolkit). Similarity to known proteins in the Genbank database was assessed using protein BLAST and CDD. Protein similarity matches suggested by BLAST were evaluated by comparing e-values, that associates a statistical probability to the match results. An e-value of less than E^{-03} was considered significant. Additionally, results of the BLAST were analyzed by multiple sequence alignment tools in T-Coffee and the results were visualized via WebLogo, to determine the portion of the sequences most conserved. We found most proteins to be well conserved through evolution.

INTRODUCTION

Kytococcus sedentarius is a Gram positive bacteria, part of the poorly studied family, Dermacoccaceae. *Kytococcus sedentarius* is found in the ocean. It is harmful to humans because it can cause pitted keratolysis, valve endocarditis and hemorrhagic pneumonia (Sims et al. 1990). However, *Kytococcus sedentarius* is also beneficial because it produces enzymes which could potentially be used to break down callouses (Longshaw et al., 2002) and produces antibiotics monesin A and B (Pospisil, et al., 1998) which are used in animal feeds to promote growth.



Scanning electron micrograph of *Kytococcus sedentarius* Strain 541T (Manfred Rohde, Helmholtz Centre for Infection Biology, Braunschweig)

Gene neighborhoods of two of the six genes annotated

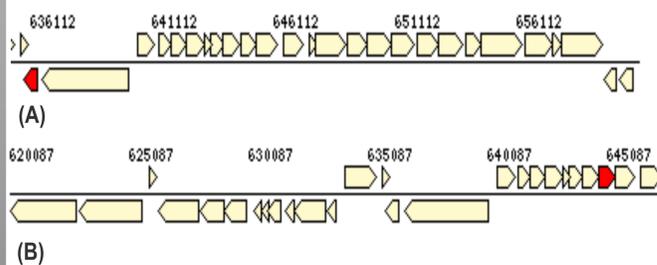


Figure 1

Gene neighborhoods of:

(A) Ksed_06430 (red) DNA coordinates 635832-636389 (-) (558 bp); 185 amino acids.

(B) Ksed_06520 (red) DNA coordinates 644768-645403 (+) (636 bp); 211 amino acids.

METHODS AND MATERIALS

Modules of the GENI-ACT (<http://www.geni-act.org/>) were used to complete *Kytococcus sedentarius* genome annotation. The modules used 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?

Table 1

RESULTS

Ksed_06430:

The proposed product of this gene by Genbank was RNA polymerase sigma factor, sigma-70 family. BLAST search against the Swissprot database showed no matches with high score.

Query	Accession	Score	Expect	Method	Identities	Positives	Gaps	Open	Matches
Query 1	AB011111	321	1e-108	Compositional matrix adjust.	160/181(88%)	171/181(94%)	0/181(0%)		1
Query 2	AB011111	321	1e-108	Compositional matrix adjust.	160/181(88%)	171/181(94%)	0/181(0%)		1
Query 3	AB011111	321	1e-108	Compositional matrix adjust.	160/181(88%)	171/181(94%)	0/181(0%)		1
Query 4	AB011111	321	1e-108	Compositional matrix adjust.	160/181(88%)	171/181(94%)	0/181(0%)		1
Query 5	AB011111	321	1e-108	Compositional matrix adjust.	160/181(88%)	171/181(94%)	0/181(0%)		1
Query 6	AB011111	321	1e-108	Compositional matrix adjust.	160/181(88%)	171/181(94%)	0/181(0%)		1
Query 7	AB011111	321	1e-108	Compositional matrix adjust.	160/181(88%)	171/181(94%)	0/181(0%)		1
Query 8	AB011111	321	1e-108	Compositional matrix adjust.	160/181(88%)	171/181(94%)	0/181(0%)		1
Query 9	AB011111	321	1e-108	Compositional matrix adjust.	160/181(88%)	171/181(94%)	0/181(0%)		1
Query 10	AB011111	321	1e-108	Compositional matrix adjust.	160/181(88%)	171/181(94%)	0/181(0%)		1

Figure 2 BLAST search with the nr database shows a significant match to putative RNA polymerase ECF-type sigma factor from the organism *Gordonia soli* NBRC 108243 with a high score of 321 bits and low e-value of $1e^{-108}$.

Ksed_06440:

BLAST search against nr database returned a match with *TrwC relaxase* in *Microbacterium laevaniformans*. The e-value for this protein is 0 which along with a high score of 2097 bits, 89% identity and 93% positives indicates a highly significant match.



Figure 3: Multiline WebLogo (with 32 symbols per line, in 4 panels) from the multiple alignment by T-coffee of 10 top orthologs from the BLAST search. The height of the amino acid stacks from N-C terminal of the protein indicates very good conservation. Due to the evidence of high conservation, we propose that *TrwC relaxase* also performs important functions in *K. sedentarius*.

RESULTS

Ksed_06460:

The initial proposed product of this gene from Genbank was RNA polymerase. BLAST search with nr database also found significant match with RNA polymerase sigma24 factor with e-value $3x10^{-91}$.

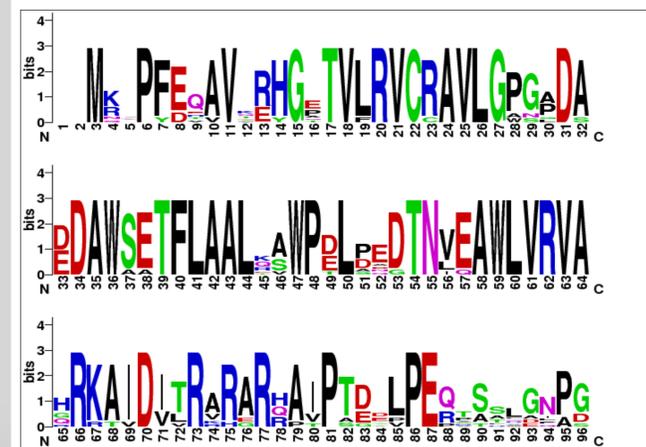


Figure 4 WebLogo generated from multiple alignment using Ksed_06460

Ksed_06470:

BLAST search using the Swissprot database found a significant match with Methylated-DNA--protein-cysteine methyltransferase with a significant e-value of $1x10^{-31}$.

The COG hits, T-Coffee alignments and WebLogos of both *Ksed_06460* and *Ksed_06470* indicate well conserved proteins with similar vital functions in *K. sedentarius*.

Ksed_06480:

Although the Genbank annotation for this locus was- uncharacterized conserved protein, BLAST search against the nr database found a significant match with proline hydroxylase from *Gulosibacter molinivorax*. The e-value was $1x10^{-164}$. However, the top COG hit is an uncharacterized protein with unknown function.

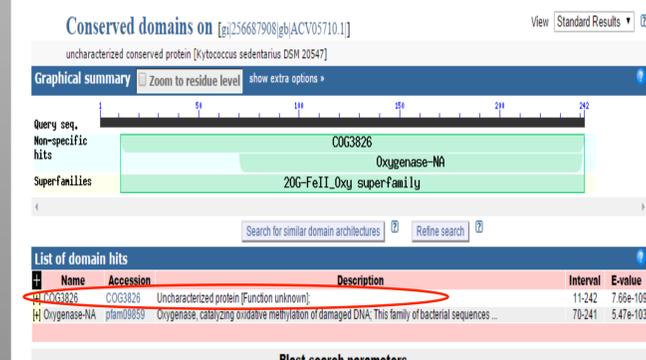


Figure 5: The COG hit indicating an uncharacterized protein for *Ksed_06480*.

Future work: We suggest further annotation to reconcile the significant BLAST matches with the COG hit as well as the original annotation from Genbank.

RESULTS

Ksed_06490:

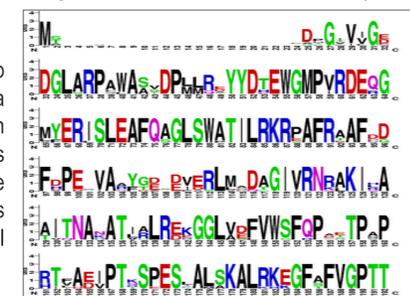
BLAST search found a moderately significant match in nr database with a metal-binding protein with E-value was $2x10^{-45}$ and score 153 bits. The top COG hit was Methylphosphotriester-DNA--protein-cysteine methyltransferase

Ksed_06520:

BLAST search found a significant match with DNA-3-methyladenine glycosylase I in *Gulosibacter molinivorax* against the nr database. The e-value was $3x10^{-126}$ with a high score of 367 bits with 83% identity.

Figure 6:

T-Coffee and WebLogo indicate that this is a well conserved protein among all ten orthologs chosen for the experiment and thus must perform a vital function in *K. sedentarius*



CONCLUSION

Most of the proposed gene products did not differ significantly from the automated Genbank gene annotation except *Ksed_06480*. Initial analysis indicates *Ksed_06480* to be a *proline hydroxylase* instead of the uncharacterized protein predicted by Genbank. Further analyses using other annotation tools and applications may yield substantiating evidence. Although, the only tools used in our study were BLAST, CDD, T-Coffee and WebLogo, we have preliminary evidence to suggest an alternate annotation requiring much further work.

Gene Locus	Geni-Act Product	Proposed Annotation
06430	RNA polymerase	RNA polymerase
06440	TrwC relaxase.	TrwC relaxase.
06460	RNA polymerase	RNA polymerase
06470	Methylated-DNA--protein-cysteine	Methylated-DNA--protein-cysteine
06480	Uncharacterized,	proline hydroxylase
06490	metal-binding protein	metal-binding protein
06520	DNA-3-methyladenine glycosylase I	DNA-3-methyladenine glycosylase I

Table 2

REFERENCES

Longshaw et al.,(2002) *Kytococcus sedentarius*, the organism associated with pitted keratolysis, produces two keratin-degrading enzymes
 Pospisil, et al., (1998) *Kytococcus sedentarius* and *Dermacoccus nishinomiyaensis* produce monesins, typical *Streptomyces cinnamomensis* metabolites
 Sims et al., (1990) Complete genome sequence of *Kytococcus sedentarius* type strain (541T)

ACKNOWLEDGMENTS

We would like to thank Drs. Rama Dey-Rao and Stephen T Koury for assistance with training and teaching of the project. Supported by NSF ITEST Strategies Award Number 1311902