Annotation of the *Kytococcus sedentarius* Genome from Ksed_09090 - Ksed_09130 and Ksed_09100

Kaia Skurzewski, Jeffery Rashko, Cristian Swiatek, and Ashley Vohwinkel
John F Kennedy High School and The Western New York Genetics in Research Partnership

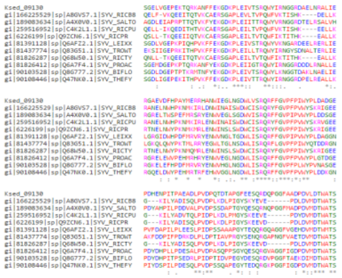
Abstract

A group of consecutive 5 genes from the microorganism *Kytococcus sedentarius* (Ksed_09090 – Ksed_09130) 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. 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 by in the r 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). K. sedentarius is among two cultures that are capable of degrading human callus, and these enzymes have been used in washing powders, keratin polymers, and the expulsion of unwanted calluses on human skin (Public Med, NCBI).

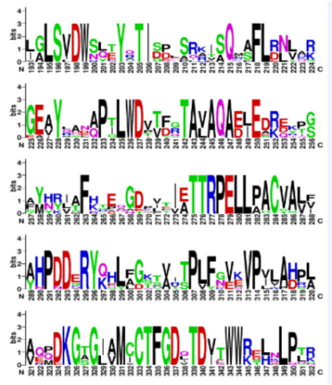
According to Sims et al. (2009), *Kytococcus sedentarius* is a microorganism of interest for several reasons. This bacterium is a natural source of the oligoketide antibiotics monensin A and monensin B (Sims et al., 2009). *Kytococcus sedentarius* has been implemented as the etiological agent of a number 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).



```
Ksed_09130                        SGELVGEPEKTQRKANFFEKGDKPLEIVTSRQWYIRNGGRDAELNRALIE
gi|166225529|sp|A8GVS7.1|SYV_RICB8 QELF-VKQEEITQTVKCAERSGAPLEVLTVPQWFVKTISHK----DELLK
gi|189083634|sp|A4X0V0.1|SYV_SALTO AGDLEIAPRPTTQTVKFYEKGDKPLEIITTRQWYVRNGGRDTELRSALVH
gi|259516952|sp|C4K2L1.1|SYV_RICPU QELL-IKQEDITHTVKCAERSGAPLEILTVPQWFVKTISHK----EELLK
gi|6226199|sp|Q9ZCN6.1|SYV_RICPR   QSLL-TKQEEIIQTVKCAERSGAPLEILTVTQWFIKTITHK----EALLK
gi|81391128|sp|Q6AFZ2.1|SYV_LEIXX  SGDLVGEPKPIQHPVKFFEKGDKPLEIVSTRQWYVKNGARDEELRERLIE
gi|81437774|sp|Q83G51.1|SYV_TROWT  EKSIIGEPRKITHPVKFFEKGDKPLEILLTRQWYIRNGYSDNALTERLIE
gi|81826287|sp|Q68W50.1|SYV_RICTY  QNLL-TKQEEITQTVKCAERSGAPLEILTVTQWFIKTITHK----EALLK
gi|81826412|sp|Q6A7F4.1|SYV_PROAC  SGEMDGEPKPTQRKANFYEKGDKPLEIIGTRQWYIRNGGRDDDLRNALLE
gi|90103528|sp|Q8G777.2|SYV_BIFLO  SGDLDGEPTPTKRMTNFYEKGDKPLEIVTSRQWYLKNGGTDAKLNAELIE
gi|90108446|sp|Q47NK0.1|SYV_THEFY  SGDLIGEPEKITHPVKFFEKGDKPLEIVTTRQWYIRNGGRDPELREALLK
                                      :          : .:  *:.. ***::   **:::.   .      *:.

Ksed_09130                        RGAEVDFHPAYMERRHANWIEGLNGDWLVSRQRFFGVPFPIWYPLDADGE
gi|166225529|sp|A8GVS7.1|SYV_RICB8 RANELNWHPKNMKIRLDNWINAISWDWCISRQRYFGVPFPVWYSKRIGEE
gi|189083634|sp|A4X0V0.1|SYV_SALTO RGRELTWSPEFMRSRYENWVEGLSGDWVISRQRFFGVPIPVWYPLDDSGE
gi|259516952|sp|C4K2L1.1|SYV_RICPU RANELNWHPKNMKIRLENWINAISWDWCISRQRYFGVPFPVWYSKRVGEE
gi|6226199|sp|Q9ZCN6.1|SYV_RICPR   RTNELNWYPKNMKMRLENWINSLSWDWCISRQRYFGVPFPIWYSKRIGEE
gi|81391128|sp|Q6AFZ2.1|SYV_LEIXX  LGRGIDWHPDFMRVRYENWVNGLTGDWLISRQRFFGVPIPVWYPLDAGGN
gi|81437774|sp|Q83G51.1|SYV_TROWT  LGKQLQWYPKTMLRRYEGWLTGLNSDWLISRQRFLGVPFPIWYQTDDRGN
gi|81826287|sp|Q68W50.1|SYV_RICTY  RTNELNWYPKNMQMRLENWINSLSWDWCISRQRYFGVPFPIWYSKRIGEE
gi|81826412|sp|Q6A7F4.1|SYV_PROAC  RGRELEWVPEHMRHRYENWVEGLNGDWLISRQRFFGVPFPVWYPLGTDGE
gi|90103528|sp|Q8G777.2|SYV_BIFLO  RGKELEFHPDFMRVRYENWVHGLNGDWLISRQRFFGVPFPLWYPVNASGE
gi|90108446|sp|Q47NK0.1|SYV_THEFY  RGQELDWYPEHMRTRFEHWVGGLNGDWLISRQRFFGVPFPVWYPLDDKGN
                                      : : *  *  *   *: .:. ** :****::***:*:**        :

Ksed_09130                        PDHENPITPAEADLPVDPQTDTAPGFEESQRDQPGGFAADPDVLDTWATS
gi|166225529|sp|A8GVS7.1|SYV_RICB8 G---KILYADISQLPVDPLKDLPIGYSKYEVE------PDLDVMDTWATS
gi|189083634|sp|A4X0V0.1|SYV_SALTO PDYAHPILPDDVALPVDPSSDAPTGYQESQRNQPGGFMADPDVMDTWATS
gi|259516952|sp|C4K2L1.1|SYV_RICPU G---KILYADVTQLPIDPLKDLPMGYSKEEVE------PDYDVMDTWATS
gi|6226199|sp|Q9ZCN6.1|SYV_RICPR   G---KILYADISQLPVDPLKDLPIGYSKEEVD------PDLDVMDTWATS
gi|81391128|sp|Q6AFZ2.1|SYV_LEIXX  PVFDAPILPLEESLPIDPSSAAAPGYTEQQRGQAGGFVGEHDVMDTWMTS
gi|81437774|sp|Q83G51.1|SYV_TROWT  AKFDDPIFPDRKDLPLDPTIAVPRGYSENQRGAPNGFVAETDVMDTWATS
gi|81826287|sp|Q68W50.1|SYV_RICTY  G---KILYADISQLPVDPLKDLPIGYSKEEVD------PDLDVMDTWATS
gi|81826412|sp|Q6A7F4.1|SYV_PROAC  PDYDHPLLPDESALPVDPASQPPSGYQESQRGVAGGFIGDPDVMDTWATS
gi|90103528|sp|Q8G777.2|SYV_BIFLO  PDYDHPITPSEDRLPIDPTIDVPEGYDESQRDVPGGFTAEKDIMDTWATS
gi|90108446|sp|Q47NK0.1|SYV_THEFY  PIYDSPILPDESQLPVDPSSQAPEGYTEDQRGKPGGFIGDPDVMDTWATS
                                         : .   **:**    . *: : :          : *::*** **
```

A multiple sequence alignment for Ksed_09130 created on *T-Coffee*

Methods and Materials



Using Web Logo, we compared our assigned genes to our top hits. Sequences 3, 4, 8, 11, 16, 17, 18, 20, and 21 resemble Kaia's gene Ksed_09130 the most; 1, 27, 29, 30, and 31 resemble her gene the least.

Results

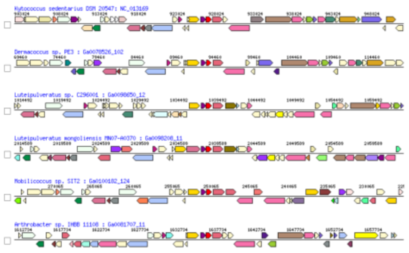
Kytococcus sedentarius 09090:
When put through the BLAST, the results that showed up was ATP-dependent Clp protease proteolytic subunit 1. This was confirmed in the TIGRFAM as well as PFAM in the structure-based evidence, as well as the presence of well-curated domains within the sequence. The proposed annotation is ATP-dependent Clp protease proteolytic subunit 1.

Kytococcus sedentarius 09100:
The initial proposed product of this gene by GENI-ACT was a histidine ammonia-lyase. 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 cellular location of the amino acid sequence, and the enzymatic function of the amino acid sequence. As such, the proposed annotation is a histidine ammonia-lyase.

Kytococcus sedentarius 09130:
Pfam proposed that this gene was related to functional domains like tRNA synthetases class 1 and an anticodon-binding domain of tRNA. 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.

Kytococcus Sedentarius 09110:
Although *Kytococcus sedentarius* 09110 contained about four well-curated functional domains related to the gene. This protein is labeled a caseinolytic protease through ExPASy ENZYME.

These alpha subunits also characteristically contain a molybdopterin dinucleotide binding domain. Interestingly enough, this second domain was found within the consecutive proposed gene: *Kytococcus sedentarius* 09130. These two genes, which are interrupted by a stop codon, appear to encode the alpha subunit of the enzyme formatted dehydrogenase. This anomaly was discovered in both more close and distant phylogenetic relatives of *Kytococcus sedentarius.* The best characterization of this enzyme originates from *Thermobifida fusca YX*, in which the alpha subunit is encoded by a single gene that contains both well-curated functional domains. As such, the proposed annotation for *Kytococcus sedentarius* 09110 and 09130 are respectively the Fe4S4 and molybdopterin dinucleotide binding domains of the alpha subunit of the enzyme formate dehydrogenase.



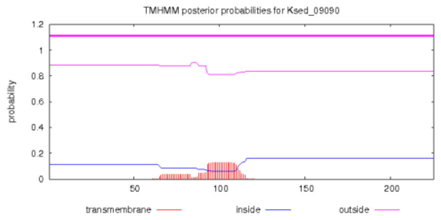
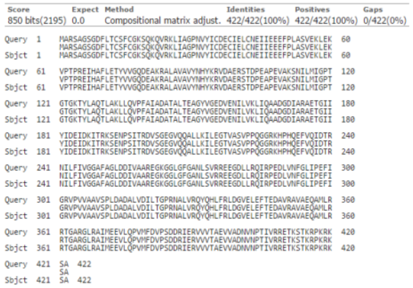


Figure II – *Kytococcus sedentarius* 09090 well-curated functional domains which are incomplete at the C-terminal end of the amino acid sequence.

	Name	Accession	Description	Interval	E-value
[+]	MopB_CT_Formate-Dh-Na-like	cd02792	Formate dehydrogenase N, alpha subunit (Formate-Dh-Na) is a major component of nitrate ...	716-847	3.30e-49
[+]	MopB_Formate-Dh-Na-like	cd02752	Formate dehydrogenase N, alpha subunit (Formate-Dh-Na) is a major component of nitrate ...	1-695	5.66e-143
[+]	Molydop_binding	pfam01568	Molydopterin dinucleotide binding domain; This domain is found in various molybdopterin - ...	721-840	6.19e-16
[+]	MopB_Phenylacetyl-CoA-OR	cd02760	The MopB_Phenylacetyl-CoA-OR CD contains the phenylacetyl-CoA:acceptor oxidoreductase, large ...	4-217	9.48e-10
[+]	Molybdopterin	pfam00384	Molybdopterin oxidoreductase;	2-77	1.71e-06
[+]	formate-DH-alph	TIGR01553	formate dehydrogenase, alpha subunit, proteobacterial-type; This model describes a subset of ...	1-839	0e+00
[+]	BisC	COG0243	Anaerobic dehydrogenases, typically selenocysteine-containing [Energy production and ...	1-847	3.27e-58
[+]	PRK09939	PRK09939	putative oxidoreductase; Provisional	170-272	1.47e-12

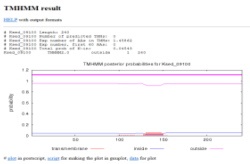
Figure III – Kytococcus sedentarius 09100 well-curated functional domains which are incomplete at the N-terminal end of the amino acid sequence.



Score	Expect	Method	Identities	Positives	Gaps
850 bits(2195)	0.0	Compositional matrix adjust.	422/422(100%)	422/422(100%)	0/422(0%)

```
Query  1    MARSAGSGDFLTCSFCGKSQKQVRKLIAGPNVYICDECIELCNEIIEEEFPLASVEKLEK  60
            MARSAGSGDFLTCSFCGKSQKQVRKLIAGPNVYICDECIELCNEIIEEEFPLASVEKLEK
Sbjct  1    MARSAGSGDFLTCSFCGKSQKQVRKLIAGPNVYICDECIELCNEIIEEEFPLASVEKLEK  60

Query  61   VPTPREIHAFLETYVVGQDEAKRALAVAVYNHYKRVDAERSTDPEAPEVAKSNILMIGPT  120
            VPTPREIHAFLETYVVGQDEAKRALAVAVYNHYKRVDAERSTDPEAPEVAKSNILMIGPT
Sbjct  61   VPTPREIHAFLETYVVGQDEAKRALAVAVYNHYKRVDAERSTDPEAPEVAKSNILMIGPT  120

Query  121  GTGKTYLAQTLAKLLQVPFAIADATALTEAGYVGEDVENILVKLIQAADGDIARAETGII  180
            GTGKTYLAQTLAKLLQVPFAIADATALTEAGYVGEDVENILVKLIQAADGDIARAETGII
Sbjct  121  GTGKTYLAQTLAKLLQVPFAIADATALTEAGYVGEDVENILVKLIQAADGDIARAETGII  180

Query  181  YIDEIDKITRKSENPSITRDVSGEGVQQALLKILEGTVASVPPQGGRKHPHQEFVQIDTR  240
            YIDEIDKITRKSENPSITRDVSGEGVQQALLKILEGTVASVPPQGGRKHPHQEFVQIDTR
Sbjct  181  YIDEIDKITRKSENPSITRDVSGEGVQQALLKILEGTVASVPPQGGRKHPHQEFVQIDTR  240

Query  241  NILFIVGGAFAGLDDIVAAREGKGGLGFGANLSVRREEGDLLRQIRPEDLVNFGLIPEFI  300
            NILFIVGGAFAGLDDIVAAREGKGGLGFGANLSVRREEGDLLRQIRPEDLVNFGLIPEFI
Sbjct  241  NILFIVGGAFAGLDDIVAAREGKGGLGFGANLSVRREEGDLLRQIRPEDLVNFGLIPEFI  300

Query  301  GRVPVVAAVSPLDADALVDILTGPRNALVRQYQHLFRLDGVELEFTEDAVRAVAEQAMLR  360
            GRVPVVAAVSPLDADALVDILTGPRNALVRQYQHLFRLDGVELEFTEDAVRAVAEQAMLR
Sbjct  301  GRVPVVAAVSPLDADALVDILTGPRNALVRQYQHLFRLDGVELEFTEDAVRAVAEQAMLR  360

Query  361  RTGARGLRAIMEEVLQPVMFDVPSDDRIERVVVTAEVVADNVNPTIVRRETKSTKRPKRK  420
            RTGARGLRAIMEEVLQPVMFDVPSDDRIERVVVTAEVVADNVNPTIVRRETKSTKRPKRK
Sbjct  361  RTGARGLRAIMEEVLQPVMFDVPSDDRIERVVVTAEVVADNVNPTIVRRETKSTKRPKRK  420

Query  421  SA  422
            SA
Sbjct  421  SA  422
```

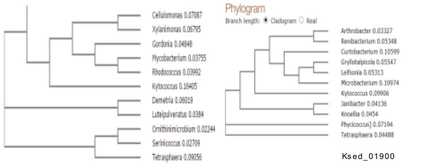
The alignment shows that my protein/amino acid sequence exactly matches the translation of the DNA sequence at IMG/EDU. Therefore there is no evidence of a frameshift mutation or a premature stop codon.



Ksed_09100 is gram-positive bacteria and based on this TMHMM result, no transmembrane helixes could be found. The computer showed that no signal peptide could be predicted. Its cytoplasmic score was 9.97 out of 10. This gene is closely related to Kytococcus.

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.



A Phylogram Tree of Kytococcus sedentarius

Based on the computer's annotations on my gene I can confidently say that Ksed_09100 does not significantly differ from the other genes. Its phylogenetic tree shows that the gene is highly conserved in organisms closely related to Kytococcus, therefore there was no evidence of horizontal transfer. The characteristic GC percent was 72% and the average GC% of the gene was 69%. There was also no evidence of a frameshift mutation or stop codon.

References

Sims et al. (2009). Complete genome sequence of Kytococcus sedentarius type strain (541T). *Standards in Genomic Sciences*,12-20.
NCBI PubMed (1997)

Acknowledgments

Supported by NSF ITEST Strategies Award Number 1311902