University at Buffalo *The State University of New York* | REACHING OTHERS

BUFFSCI BUFFALO ACADEMY OF SCIENCE

NSF

Annotation of the *Kytococcus sedentarius* Genome from DNA Coordinates 712204 to 719468

Mahjabin Ali, Tashrif Chowdhury, Nabilah Narwin, Gregory Turner, and Jennifer Livergood

Buffalo Academy of Science Charter School and Western New York Genetics in Research Partnership

Abstract

A group of consecutive 5 genes from the microorganism *Kytococcus sedentarius* (Ksed_07140 – Ksed_07180) 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 and amino acid sequence-based similarity data. The Genbank proposed gene product name did not differ significantly from the proposed gene annotation for most of the genes in the group. These genes appear to be correctly annotated by in the database. One gene locus tag was found to be incorrectly identified in the Genbank. A proposed gene name has been found.

Introduction

Kytococcus sedentarius is believed to cause pitted keratolysis, a condition that erodes one's feet. This condition is characterized by malodor and erosions on the soles of the feet. *Kytococcus sedentarius* was originally isolated from a playground slide in 1944. This organism is classified as a chemoautotroph, as it obtains energy from the oxidation of carbon compounds, including amino acids. *Kytococcus sedentarius* is a member of the family *Dermatococcaceae* within the order *Micrococcales*, and class *Actinobacteria*. The genomic sequence has yet to be completely studied (Sims et al., 2009). It is important to annotate the genome completely to be able to understand more about the underlying biology of the organism.

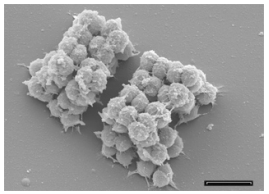

Photo credit: Dr. Manfred Rohde at Helmholtz Centre for Infection Research, Braunschweig, Scale = 2 µm.

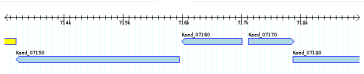

Figure 1. The *Kytococcus* locus tags under investigation in this research. Locus tag numbers are indicated for the genes that have been annotated.

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?

Results

Ksed_07140

The initial proposed product of this gene by GENI-ACT was a nucleoside-diphosphate-sugar epimerase. This gene product proposal was supported by the top BLAST hits for the amino acid sequence (Figure 3). The weblog for Ksed_07140 showed areas of hydrophobic aminio acids separated by polar and basic amino acids.

nucleoside-diphosphate sugar epimerase [Mobilicoccus pelagius]
Sequence ID: ref|WP_009482784.1| Length: 334 Number of Matches: 1
See 1 more title(s)

Range 1: 2 to 330 GenPept Graphics

Score	Expect	Method	Identities	Positives	Gaps
364 bits(935)	4e-121	Compositional matrix adjust.	209/335(62%)	241/335(71%)	18/335

```
Query  3    RVLVTGASGMLGGAVARALAARGDEVTVMQRRRS----ASGLREVLGSVDDPAAVARAVA  58
            RVLVTGASGMLGG VARALAARGD+VTV+QRR S      +G+REVLG V D  AVARA+A
Sbjct  2    RVLVTGASGMLGGVARALAARGDVTVLQRRPSGLAETAGVREVLGDVVDAEAVARAMA  61

Query  59   GQDAVVHLAAKVDVVGPAAEFERTNIGGTGLLLDAAQRAGVGRFVHTSSPSVAHAGAALV  118
            GQDAVVHLAAKV+V+G   ++ R N+ GT+ +LDAA+R  V RFVHTSSPSVAHAGA+LV
Sbjct  62   GQDAVVHLAAKVNVMGRWEDYVRVNVVGTRTMLDAARRHAVPRFVHTSSPSVAHAGASLV  121

Query  119  GVEAEPASPEHARGHYARTKAVAELDALA------RDTEGLAVTAVRPHLVWGPGDTQLV  172
            G GA P  PE ARG+YAR+KA+E  ALA        RD  G AVT +RPHLVWGPGDTQLV
Sbjct  122  GEGAGPTDPEQARGNYARSKALSERIALAADARNVRDGIGPAVTVIRPHLVWGPGDTQLV  181

Query  173  GRLVQRAHAGRLPWVGDGAALVDSTYVDNARDALVAALDALAGPRNAGSAHGRAFVVTND  232
            R++HRA AGRLP +G GAALVD+TYVDNA DA VAALD      R        G +FVVTN
Sbjct  182  ARVIRRAQAGRLPVIGSGAALVDTTYVDNAVDAFVAALDRCEAVR------GESFVVTNG  235

Query  233  EPRPVGELIGSILRASGAPLPRLRIGTGAAGAAGHAVDGAVLASQRVHPITE--PPLTHF  290
            EPRPV E+  +  RA GA     RI T  A AAG  VD      +R+  + E  PPLT F
Sbjct  236  EPRPVSEIFAAWARAGGADPSPTRIPTRLAYAAGGVVDAVTAVRERIGTVDEADPPLTRF  295

Query  291  LVEQLTTAHWFSQAETHAALGWRPAVSLDDGLERL  325
            LVEQLTTAHWF Q   T   ALGW P V+LD+G ERL
Sbjct  296  LVEQLTTAHWFDQRRTREALGWTPRVTLDEGFERL  330
```

Figure 3. The top nr database blast hit of Ksed_07140

Figure 4. Ksed_07140 Weblogo results. Hydrophobic amino acids are in black, polar amino acids are green and basic amino acids are blue.

Ksed_07150

The initial proposed gene product was acyl-CoA synthetase (AMP-forming)/AMP-acidligase II. The top two BLAST search results were a hydrolase, and both had an E-value of 0 and high scores. Research shows that a hydrolase and synthetase have a function of assembling molecules in the cell. This was a large protein of 922 amino acids in length.

Ksed_07170

GENI-ACT indicated, that the gene product name was to be ABC-type multidrug transport system, ATPase component. The top BLAST hits for this were ABC Transporter (Figure 5), and the top COG hit was ABC-type Na+ transport system, ATPase component NatA. Further research will need to be done to determine if Ksed_07170 is the ATPase domain of a sodium ABC transporter in *Kytococcus sedentarius*.

ABC transporter [Phycicoccus sp. Root101]
Sequence ID: ref|WP_056920105.1| Length: 256 Number of Matches: 1
See 1 more title(s)

Range 1: 7 to 254 GenPept Graphics

Score	Expect	Method	Identities	Positives	Gaps
326 bits(836)	1e-108	Compositional matrix adjust.	166/250(66%)	197/250(78%)	2/250(0%)

```
Query  7    AMEWRGLTKAFGGQVAVDGVTLAVPRGAVLGLVGPNGAGKTTALSMATGLLAPDAGTVEV  66
            A+  RGLTK FG  VAVD + L VPRG+  GLVGPNGAGKTT LSMATGLL PD GT  V
Sbjct  7    ALGLRGLTKRFGDMVAVDHLDLDVPRGSFFGLVGPNGAGKTTTLSMATGLLRPDQGTAHV  66

Query  67   LGHDVWSDPRGAKALLGVLPDGMRVFDRLSGRELLTYTGRLRGMDEAVIAERSAELLQAL  126
            LG DVWSDP  AKA +GVLPDG+R+FDRL+GR+L+TY G LRGMD  ++ +R+ ELL AL
Sbjct  67   LGADVWSDPVAAKARIGVLPDGLRLFDRLTGRQLVTYAGLLRGMDRELVRQRTDELLAAL  126

Query  127  GLEGAGNKMVDYSAGMTKKIGLACALIHAPRLLVLDEPLEAVDPVSAQGIRELLAGYAA  186
            GL  A  K+V+DYSAGMTKK+ LACALIHAPRLLVLDEP EAVDPVSA+ IR +L  +AA
Sbjct  127  GLTDAAEKLVIDYSAGMTKKVTLACALIHAPRLLVLDEPFEAVDPVSARTIRGILEDFAA  186

Query  187  NGGTVVISSHVMELVESLCSHVAIMREGQILREGTVEEVREGDSLQNTFIRLVGREAAGA  246
            NGGTVV+SSHVM+LVE +CSHVAI+  GQ+   GTV++VR G SL + F+ LVG +
Sbjct  187  NGGTVVLSSHVMDLVERICSHVAIIAAGQVRAHGTVDDVRAGMSLDDRFLDLVGQGT--D  244

Query  247  GEGLSWLRSS  256
            EGL+WLR+S
Sbjct  245  VEGLAWLRTS  254
```

Figure 4. The top named nr database blast hit of Ksed_07170

Ksed_07180

According to Geni-Act the proposed function of the gene is hypothetical. The top nr blast hits are all hypothetical for Ksed_07180, but further down the list there are matches thought to be transporters (Figure 5). Since Ksed_07170 may be an ATPase component of an ABC transporter, Ksed_07180 may be the permease component of the same transporter. With the help of Dr. Koury, the amino acid sequence was analyzed using Phobius, a tool from the cellular localization module of Geni-Act, that can identific transmembrane helixes and signal peptides. Ksed_07170 was found to have 13 transmembrane helicies, suggesting it is embedded in the cytoplasmic membrane of *K. sedentarius* (Figure 6).

Sequences producing significant alignments:

Select: All None Selected:0

Alignments Download GenPept Graphics Distance

	Max score	Total score	Query cover	E value	Ident	Accession
hypothetical protein [Kytococcus sedentarius]	994	994	100%	0.0	100%	WP_012802192.1
hypothetical protein [Microlunatus phosphovorus]	283	283	99%	3e-84	39%	WP_013862081.1
hypothetical protein [Microbacterium sp. UCD-TDU]	234	234	99%	7e-66	35%	WP_017828117.1
hypothetical protein [Aestuariimicrobium kwangyangense]	231	231	96%	5e-65	36%	WP_022910699.1
hypothetical protein [Luteipulveratus mongoliensis]	229	229	99%	7e-64	34%	WP_052594024.1
hypothetical protein [Microbacterium sp. Root1433D1]	228	228	99%	2e-63	35%	WP_055870008.1
hypothetical protein [Microbacterium sp. Ag1]	226	226	99%	9e-63	35%	WP_046747299.1
transporter [Arthrobacter sp. M2012083]	224	224	96%	3e-62	34%	WP_017200122.1
transporter [Arthrobacter sp. ATCC 21022]	221	221	96%	6e-61	34%	WP_062097153.1

Figure 5. List of the top nr database blast hits matching Ksed_07170

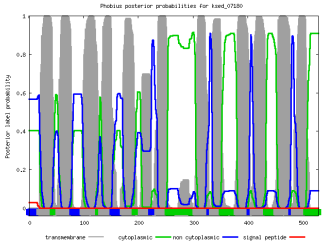

Figure 6. The Phobius results for Ksed_07170.

Conclusion

The GENI-ACT proposed gene product did not differ significantly from the proposed gene annotation for most of the genes in the group. One gene will need further research to determine if the proposed gene product is correct. More research needs to be done

Gene Locus	Geni-Act Gene Product	Proposed Annotation
Ksed_07140	nucleoside-diphosphate-sugar epimerase	nucleoside-diphosphate-sugar epimerase
Ksed_07150	acyl-CoA synthetase (AMP-forming)/AMP-acidligase II	more information is needed to propose an new gene name.
Ksed_07160	3-oxoacyl-(acyl-carrier-protein) synthase III	No data to support or refute
Ksed_07170	ABC-type multidrug transport system, ATPase component	ABC-type multidrug transport system, ATPase component
Ksed_07180	hypothetical	ABC transporter permease component

References

Longshaw et al. (2002). Kytococcus sedentarius, the organism associated with pitted keratolysis, produces two keratin-degrading enzymes. *Journal of Applied Microbiology*, 93(5): 810-816.

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

Acknowledgments

We would like to thank Dr. Stephen Koury and Dr. Rama Dey-Rao for their assistance and training for this program. Supported by NSF ITEST Strategies Award Number 1311902