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

The Park School
Active learning for life

Annotation of 2 *Kytococcus sedentarius* genes: *Ksed*_19620 and 25490 that encode an aspartate kinase

Meghan Francis, Xiaoying Xu, and Margaret Diamond, Ph.D.
The Park School of Buffalo and the Western New York Genetics in Research Partnership

Abstract

Two genes from the microorganism *Kytococcus sedentarius* (Ksed_19620 and 25490) were annotated using the collaborative genome annotation website GENI-ACT. The Geni-Act proposed gene product identification for each gene was examined in terms of the general genomic information, amino acid sequence-based similarity data (BLAST, CDD, T-Coffee, and WebLogo), structure-based evidence from the amino acid sequence (TIGRfam and Pfam), cellular localization data (TMHMM, SignalP, and Phobius), enzymatic function (KEGG, MetaCyc, E.C. number), gene duplication (paralog), and evidence for horizontal gene transfer. The Genbank proposed gene product name did not differ significantly from the proposed gene annotation for each of the two genes studied. Therefore, the genes appear to have been correctly identified by the computer program.

Introduction

Kytococcus sedentarius is a gram positive bacterium that produces cert ain antibiotics. It is aerobic and requires some amino acids for growth. This bacterium was originally found in a marine environment. Even though it is often overlooked it is known to have caused serious illnesses such as valve endocarditis and pneumonia It also causes foot odor and pitting of foot calluses (pitted keratolysis) (Sims 2009; James 2012) .

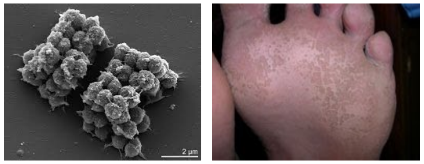


Figure 1. Left panel, scanning electron micrograph of *Kytococcus sedentarius*. Right panel, pitted keratolysis of the sole of the foot, caused by *K. sedentarius.*

*Ksed*_19620 was identified as an aspartate kinase protein; its function is to transfer a phosphate group from ATP to L-aspartate. BLAST showed that this protein is most similar to other aspartate kinase proteins identified in many other bacterial species. In addition to BLAST, other programs (as listed in Materials and Methods) were used to analyze and annotate the gene *Ksed*_19620, including WebLogo, cellular localization, duplication , and enzymatic function.

*Ksed*_25490 was also identified as an aspartate kinase. It was found to be most similar to a protein sequence from *Holophaga foetida* with a high degree of similarity to many other bacterial protein sequences as shown by the e- value 1e-148. This gene was characterized using the same methods as for *Ksed*_19620.

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?
Module 4- Cellular Localization Data	Gram Stain, TMHMM, SignalP, PSORT, Phobius	Is my protein in the cytoplasm, secreted or embedded in the membrane?
Module 6- Enzymatic Function	KEGG, MetaCyc, E.C. Number,	In what process does my protein take part?
Module 7- Gene Duplication/Gene Degradation	Paralog, Pseudogene	Are there other forms of my gene in the bacterium? Is my gene functional?
Module 8- Evidence for Horizontal Gene Transfer	Phylogenetic Tree,	Has my gene co-evolved with other genes in the genome?
Final Annotation	Review data from all modules	Does the student proposed name of the gene agree with that proposed by the automated computer annotation? Are any changes proposed to the pipeline annotation?

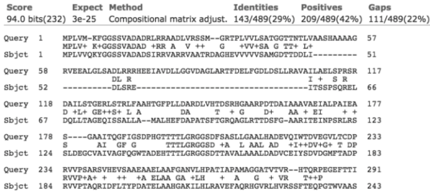
Results

Kytococcus sedentarius 19620:

Comparison to the non-redundant protein database using protein BLAST showed that *Ksed*_19620 is most similar to other identified aspartate kinase proteins and that it has an e value of 2e-171. This is a very high value which means that the compared protein sequences are very similar. WebLogo also illustrated that it was highly conserved throughout. It most likely is in the cytoplasm since no transmembrane domains or signal sequence were identified.

Kytococcus sedentarius 25490:

Comparison of the amino acid sequence of *Ksed*_25490 to the non-redundant protein database using BLAST showed a high degree of similarity to other aspartate kinase proteins found in many other bacterial species. Like *Ksed*_19620, it contained 2 functional domains, and is located in the cytoplasm, as expected given its function in amino acid metabolism. These two genes, while not identical, were found to be paralogs and catalyze the same reaction. Their E.C. number is 2.7.2.4.



```
Score           Expect  Method                        Identities     Positives      Gaps
94.0 bits(232)  3e-25   Compositional matrix adjust.  143/489(29%)   209/489(42%)   111/489(22%)

Query  1    MPLVM-KFGGSSVADADRLRRAADLVRSSM--GRTPLVVLSATGGTTNTLVAASHAAAAG  57
            MPLV+ K+GGSSVADAD +RR A  V ++   G   +VV+SA G TT+ L+
Sbjct  1    MPLVVQKYGGSSVADADSIRRVARRVAATRDAGHEVVVVVSAMGDTTDDLI---------  51

Query  58   RVEEALGLSADLRRRHEEIAVDLLGGVDAGLARTFDELFGDLDSLLRAVAILAELSPRSR  117
                      DL R                                    I +  S R
Sbjct  52   ----------DLSRE------------------------------------ITSSPSQREL  66

Query  118  DAILSTGERLSTRLFAAHTGFPLLDARDLVHTDSRHGAARPDTDAIAAAVAEIALPAIEA  177
            D +L+ GE++S+ L A          DA     T + G    D+   AA + EI    + +
Sbjct  67   DQLLTAGEQISSALLA--MALHEFDAPATSFTGRQAGLRTTDSFG-AARITEINPSRLRS  123

Query  178  S----GAAITQGFIGSDPHGTTTTLGRGGSDFSASLLGAALHADEVQIWTDVEGVLTCDP  233
            S     AI  GF G      TTTLGRGGSD +A  L AAL AD  +I++DV+G+ T DP
Sbjct  124  SLDEGCVAIVAGFQGWTADEHTTTLGRGGSDTTAVALAAALDADVCEIYSDVDGMFTADP  183

Query  234  RVVPSARSVHEVSAAEAAELAAFGANVLHPATIAPAMAGGATVTVR--HTQRPEGEFTTI  291
            RVVP+A+ +  ++  +A ELAA GA +LH   +  A   G  + VR   T++P
Sbjct  184  RVVPTAQRIDFLTYPDATELAAHGAKILHLRAVEFAQRHGVRLHVRSSFTEQPGTWVAAS  243
```

Figure II – Pairwise alignment of *Ksed*_19620 and 25490 amino acid sequences

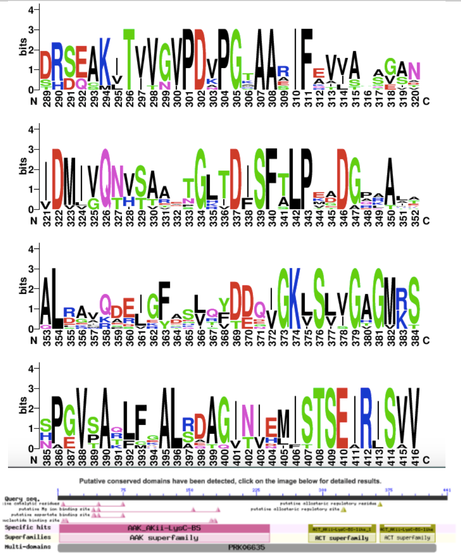


Figure III - The top panel shows part of the WebLogo for Ksed_19620, showing fairly highly conserved amino acid sequence. The bottom shows the two domains identified in this protein.

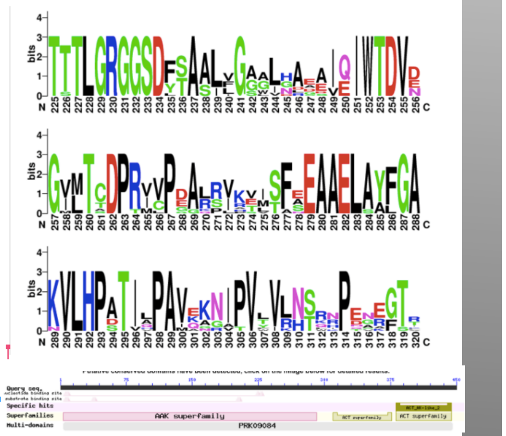


Figure IV - The top panel shows part of the WebLogo for *Ksed*_25490, showing fairly highly conserved amino acid sequence. The bottom shows the two domains identified in this protein.

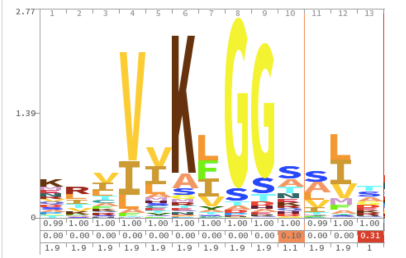


Figure V – Amino terminal end of the HMM logo showing highly conserved residues in this protein family.

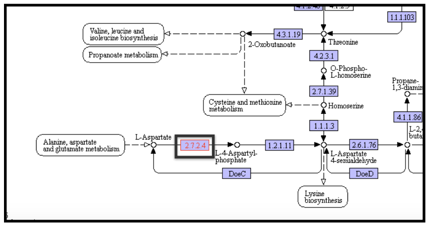


Figure VI – *Ksed*_19620 and 25490 have the same E.C. number, 2.7.2.4, shown in the black box in the glycine, serine, threonine metabolism pathway.

Conclusion

The GENI-ACT proposed gene products did not differ significantly from the proposed gene annotation for each of the genes analyzed and as such, the genes appear to have been correctly identified by the computer program as follows:

Gene Locus	Geni-Act Product	Proposed Annotation
19620	Aspartate kinase	Aspartate kinase
25490	Aspartate kinase	Aspartate kinase

References

James et al. (2013). Microbiological and Biochemical origins of foot malodour. *Flavour and Fragrance Journal,* 28: 231-237

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

http://www.regionalderm.com/Regional_Derm/Pfiles/pitted_keratolysis.html

Acknowledgments

Supported by NSF ITEST Strategies Award Number 1311902. Thanks to Dr. Stephen Koury and Dr. Patricia Masso-Welch for their invaluable help.