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

The Park School *Active learning for life*

Annotation of 3 *Kytococcus sedentarius* Genes: *Ksed_* 04800, 14810, and 19630 involved in the same metabolic pathway

Gabrielle Baumgart, Mary Stevens, and Margaret Diamond, Ph.D.

The Park School of Buffalo and the Western New York Genetics in Research Partnership

Abstract

Three genes from the microorganism *Kytococcus sedentarius* (Ksed_04800, 14810, and 19630) 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), and enzymatic function (KEGG, MetaCyc, E.C .number) The Genbank proposed gene product name did not differ significantly from the proposed gene annotation for each of the three genes studied. Therefore, the genes appear to be have been correctly identified by the computer program.

Introduction

Kytococcus sedentarius is a gram positive bacterium that produces certain 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 an opportunistic pathogen that 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).

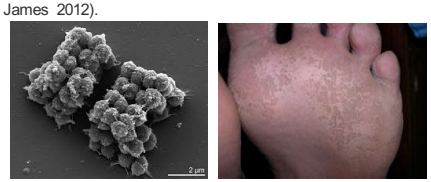

Figure I. Left panel, scanning electron micrograph of *Kytococcus sedentarius*. Right panel, pitted keratolysis of the sole of the foot, caused by *K. sedentarius*.

The initial proposed product for Ksed_04800 by GENI-ACT was homoserine O-acetyltransferase. This enzyme is part of the pathway for cysteine and methionine metabolism. The gene product proposal was supported by the top BLAST hits for the amino acid sequence, the high degree of conservation in most of the amino acid sequence, and the cellular location of the protein. As such, the proposed annotation as homoserine O-acetyltransferase, appears to be accurate.

GENI-ACT identified *Ksed_*14810 as homoserine dehydrogenase. This enzyme functions in the pathway for glycine, serine, and threonine metabolism. It is most similar to a hypothetical protein from *Proteiniclasticum reminis*, but the second hit (e value = 1e-84) was to homoserine dehydrogenase from *Clostridium ultunense*.

*Ksed_*19630 was identified by GENI-ACT as a gene that encodes aspartate semialdehyde dehydrogenase, an enzyme also in the glycine, serine, threonine and cysteine, methionine metabolic pathways. This gene product proposal was supported by the top BLAST hits for the amino acid sequence, its highly conserved amino acid sequence, and its apparent location in the cytoplasm.

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?
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 04800:

The initial proposed product of this gene by GENI-ACT was homoserine O-acetyltransferase which was confirmed by the top BLAST hits. WebLogo showed that, with the exception of the amino and carboxy terminal ends, its sequence has a high degree of similarity to proteins with the same function in other bacterial species. As would be expected from a protein involved in amino acid metabolism, cellular localization programs indicated its location in the cytoplasm; no signal sequence or transmembrane domains were found. The E.C. number for this enzyme is 2.3.1.31.

Kytococcus sedentarius 14810:

The initial proposed product of this gene by GENI-ACT was homoserine dehydrogenase. The top BLAST hit was a hypothetical protein, but the second hit was for a homoserine dehydrogenase. As shown by WebLogo, the amino acid sequence for Ksed_14810 is more conserved at the amino terminal end of the protein than at the carboxy terminus. Cellular localization data showed that this protein is located in the cytoplasm. It functions in the glycine, serine and threonine metabolic pathway, catalyzing a reaction producing L-homoserine from L-aspartate-semialdehyde. The E.C. number for this enzyme is 1.1.1.3.

Kytococcus sedentarius 19630:

The product for this gene as proposed by GENI-ACT is aspartate semialdehyde dehydrogenase, an enzyme involved in amino acid metabolism. Its identification was confirmed by the top BLAST hits. The WebLogo for this protein shows that the amino acid sequence is highly conserved. Cellular localization programs predicted that it is cytoplasmic. The E.C. number for this protein is 1.2.1.11.

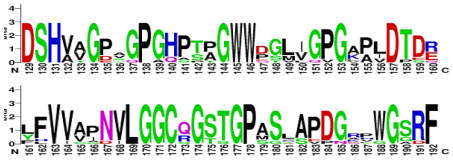

Figure II – Part of Ksed_04800 WebLogo showing highly conserved amino acid sequence.

Figure III – Part of Ksed_19630 WebLogo showing reflecting the high amino acid sequence conservation throughout.

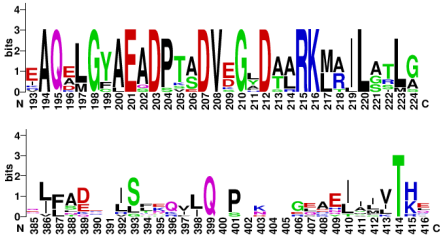

Figure IV - Ksed_14810 top panel shows the more highly conserved sequence towards the N terminus. Bottom panel shows the minimal sequence conservation towards the C terminus

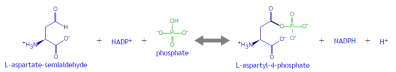

Figure V.- Reaction catalyzed by aspartate semialdehyde dehydrogenase. (Ksed_19630)

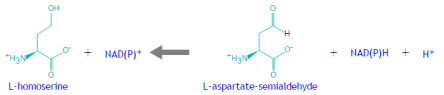

Figure VI – Reaction catalyzed by homoserine dehydrogenase (Ksed_14810)

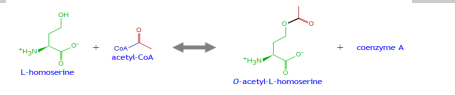

Figure VII - Reaction catalyzed by homoserine O-acetyltransferase (Ksed_04800)

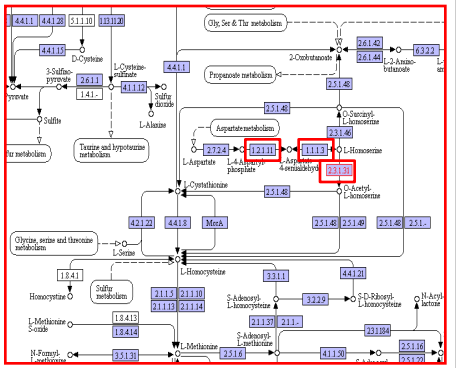

Figure VIII - Glycine, serine, threonine and cysteine, methionine metabolic pathways showing the three enzymes, E.C. 1.2.1.11, 1.1.1.3, and 2.3.1.31 in red boxes.

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.

Gene Locus	Geni-Act Product	Proposed Annotation
04800	Homoserine O-acetyltransferase	Homoserine O-acetyltransferase
14810	Homoserine dehydrogenase	Homoserine dehydrogenase
19630	Aspartate semialdehyde dehydrogenase	Aspartate semialdehyde dehydrogenase

References

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

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

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 with this project.