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

Annotation of the *Kytococcus sedentarius* Genome from Locus Tag Ksed_05850 to Ksed_06000

Ian Ferer*, Jessica Gill*, Elizabeth Koman* and Stacy Rominger (*indicates equal contribution)

Orchard Park High School, 4040 Baker Road, Orchard Park, NY 14127 and the Western New York Genetics in Research Partnership

Abstract

A group of 3 nonconsecutive genes from the microorganism *Kytococcus sedentarius* (Ksed_05850, Ksed_05990, and Ksed_06000) were annotated using the genome annotation website GENI-ACT. The proposed gene product name for each gene was evaluated based on several gene qualities, including basic genomic information, amino acid sequence-based similarities, similarities in structure from the amino acid sequence, predicted cellular locations, potential alternative open reading frames, enzymatic functions of the proposed product, presence or absence of gene duplication and degradation in paralogs or pseudogenes, and the possibility of horizontal gene transfer. The proposed gene product name was not significantly different from the gene annotation for each of the genes in the group, and therefore, the genes have likely been correctly annotated by the database.

Introduction

Kytococcus sedentarius is a non-motile, Gram-positive bacterium (containing a simple cell wall with more peptidoglycan than Gram-negative bacteria). It can act as an opportunistic pathogen in humans, and it has been isolated from areas including the human skin, airplane cabins, and groundwater. *Kytococcus sedentarius* is a spherical bacterium (also known as coccoid) and occurs mainly in tetrad form. It does not form endospores (resistant cells that form when some essential nutrient is lacking in the surrounding environment and can resume metabolism when the environment becomes favorable again) and it is non-encapsulated. It was first isolated in 1944 and thrives in NaCl solutions with concentrations up to 0.10.

According to Sims et al. (2009), *Kytococcus sedentarius* is an important bacterium to be researched for numerous reasons. It is the only known producer of the antibiotics monensin A and B, which is used consistently in the beef industry to prevent intestinal parasites from infecting livestock. *Kytococcus sedentarius* is strictly aerobic, meaning it can only survive in oxygenated environments for cellular respiration, and it is strictly chemoheterotrophic, meaning it must consume organic molecules in order to obtain energy. *Kytococcus sedentarius* requires several amino acids, including methionine, as it cannot produce these amino acids itself. However, much is also unknown about *Kytococcus sedentarius*. The gene products have been predicted for the *Kytococcus sedentarius* genome, but they have not been confirmed to be accurate. Additionally, the genomic sequencing of *Kytococcus sedentarius* can help to understand the process by which certain products can be obtained and how the bacteria's products can be determined from the annotations.

The purpose of this research was to confirm the GENI-ACT database's annotations for unknown genes in the *Kytococcus sedentarius* genome. The results of this study are listed for the genes Ksed_05850, Ksed_05990, and Ksed_06000.

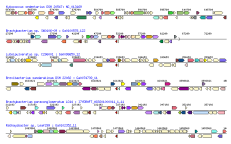

Figure 1. The ortholog neighborhood for Ksed_05850, Ksed_05990, and Ksed_06000.

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:

*Module 9 (RNA) was not completed.

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?
Module 4- Alternative Open Reading Frame	IMG Sequence Viewer For Alternate ORF Search	Has the amino acid sequence of my protein been called correctly by the computer?
Module 5- Structure-Based Evidence	TIGRfam, Pfam, PDB	Are there functional domains in my protein?
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?
Module 9- RNA	RFAM	Does my gene encode a functional RNA?

Results

Kytococcus sedentarius 05850:

The initial proposed product of this gene by the GENI-ACT database was a death-on-curing protein. The top BLAST hits (in both the nr and swissprot databases) had the name death-on-curing protein, the top COG hit had the name prophage maintenance system killer protein (a member of the death-on-curing family), the top TIGRFAM hit had the name death-on-curing family protein, and the top Pfam hit was the Fic/DOC family (death-on-curing). All of these findings support the conclusion that *Kytococcus sedentarius* 05850 is a gene for a death-on-curing protein.

One unique finding for the *Kytococcus sedentarius* 05850 protein was that the PSORTb database determined that the most probable location for this protein was the cytoplasmic membrane. However, data outputs from other databases determined that this protein had no transmembrane helices or signal peptides. The contradictory location predictions provide insufficient evidence to conclude the true location of the protein, so more research would have to be performed to accurately determine the location of this protein.

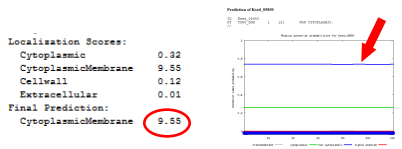

Figure 2 – Cellular localization data for *Kytococcus sedentarius* 05850. The predicted location differs between the PSORTb (membrane) and Phobius databases (no transmembrane helices).

Kytococcus sedentarius 06000:

A 801 base pair DNA sequence was analyzed. The longest open reading frame was identified as having 266 amino acids. Using Swissprot database, similarity with Polyamine aminopropyltransferase from Chromobacterium violaceum was found with a similarity score of 58.2 bits and an e-value of 1e-08. CDD search identified SpeE domain (Spermidine synthase [Amino acid transport and metabolism]) with e-value 1.01e-07.

According to T-COFFEE and WebLogo, the area between the amino terminus and the middle of the protein shows a high frequency of the amino acids in the alignment with several proteins from other microorganisms, indicating good homology. The carboxyl terminus tends to have a weaker homology than the first part of the protein. It appears that polar amino acids are the most common in the sequence, however basic, hydrophobic, and acidic are also present.

Localization of this protein is most likely cytoplasmic since PSORT-B cytoplasmic score was the highest (7.50), while other scores were below 2.0. According to PDB database, the potential structure of Kytococcus sedentarius 06000 is similar to transferase of spermidine synthase Corynebacterium glutamicum (PDB 3GJY, Figure 1).

Based on KEGG database, this protein should be involved in the metabolism of proline and arginine (Figure 2). Other databases, such as ExPASy, agree that this protein is most probably enzyme spermidine synthase. Multiple analyses demonstrated the absence of horizontal transfer of this gene. Thus, based on these analyses, this gene can be annotated as spermidine synthase.

Figure 3. Crystal structure of a probable spermidine synthase from Corynebacterium glutamicum.

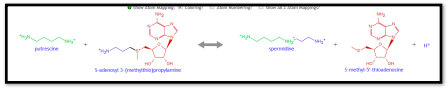

Figure 4. Most probable metabolic reaction controlled by spermidine synthase, product of Kytococcus sedentarius 06000.

Kytococcus sedentarius 05990:

The proposed product of this gene that resulted by GENI-ACT was aspartate aminotransferase. Aspartate is supported by SwissProt, the COG Name, the EC Name, and the Paralog Gene Product Name. With the combination of results from all of these programs the proposed annotation of this gene is aspartate aminotransferase.

Aminotransferase is an enzyme that catalyzes a reaction between amino acids and oxoacids by transferring amino groups. Aspartate aminotransferase, also known as AST, is an enzyme that can be found in liver and heart cells. This is released into the bloodstream when the heart or liver is damaged.

This gene is most likely to be found in the cytoplasm as indicated by PSORT-B. The cytoplasmic score for Ksed_05990 is 7.5. This is, by far, the highest score of all the results. The cytoplasmic membrane score was 1.15, cell wall was 0.62, and extracellular was 0.73. A cell's membrane has channels to let materials in and out of the cell, and the nucleus contains all genetic material ("Cytoplasm"). The cytoplasm in a cell is responsible for everything else.

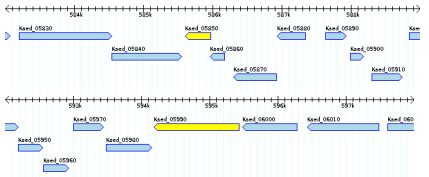

Figure 5. Gene locations of Ksed_05850, Ksed_05990, and Ksed_06000 in the *Kytococcus sedentatrius* genome.

Conclusion

Each annotated gene (Ksed_05850, Ksed_05990, and Ksed_06000) had a proposed product from the GENI-ACT database that was not significantly different from the proposed products of the other databases, so it can be accurately concluded that the gene products originally proposed by GENI-ACT were correct. Ksed_05850 was confirmed to produce a death-on-curing protein, Ksed_05990 was confirmed to produce asparatate aminotransferase, and Ksed_06000 was confirmed to produce spermindine synthase.

References

"Cytoplasm, Organelles Golgi Lysosomes – by Moose and Doc." Breast Cancer Moose and Doc. N.p., 2016. Web. 03. May 2016. <http://breast-cancer.ca/cytopagaeam/>.

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

Acknowledgments

Supported by NSF ITEST Strategies Award Number 1311902, Stephen Koury, Ph.D, Rama Dey-Rao Ph. D., Patricia Masso-Welch Ph. D., Danise Wilson, Mr. Jonathan Wolf