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

Annotation of Five Genes in the DNA Mismatch Repair Pathway of *Kytococcus Sedentarius*

Joseph M. Gawron, Rama Dey-Rao, Ph.D., Stephen Koury, Ph.D.

Department of Biotechnical and Clinical Laboratory Sciences, School of Medicine and Biomedical Sciences, State University of New York University at Buffalo, 26 Cary Hall, 3435 Main Street, Buffalo, NY 14214

Abstract

Background information: *Kytococcus sedentarius* is an opportunistic pathogen that can survive both on and inside humans. Understanding the genetics of this organism and its biologic pathways can lead to better treatments in addition to possible uses of the protein products it produces.

Purpose: The purpose of this study was to use *in silico* gene annotation to characterize five genes in the DNA mismatch repair pathway of *K. sedentarius*.

Methods: This study used the IMG-ACT website to record data in a digital notebook. The gene details page was accessed for basic information on each gene. Databases such as BLAST, CDD, TIGRFAM, PDB, PSORT-B, MetaCyc, and Prosite were used to collect qualitative and quantitative data for the five genes, including similar genes in other organisms, functional families, crystal structures, localization, and the presence of necessary functional residues. Websites like Phylogeny.fr, Pfam, KEGG, TMHMM, SignalP, Web Logo, and Phobius were used to generate diagrams used in the analysis.

Conclusions: The proposed annotations for all five genes were confirmed. The genes were found to be phylogenetically conserved between *K. sedentarius* and multiple orthologs.

Introduction

Kytococcus sedentarius is a bacteria responsible for pitted keratolysis due to its unique proteolytic enzyme secretions [1]. Pitted keratolysis is characterized by the degradation of foot callus which usually leads to the production of foot malodour [2]. Clinical reports indicate *K. sedentarius* can also play a role in peritonitis and hemorrhagic pneumonia [3, 4]. *K. sedentarius* is able to degrade the keratin in foot callus by using its proteolytic enzymes and this process yields amino acid precursors to odorous molecules [5, 6]. In the presence of occlusive footwear, bacteria living on the foot receive heat, nutrients, and protection which allow them to grow [2]. Treatment methods include depriving the bacteria of nutrients by plugging the sweat glands, stopping the odorous molecules from evaporating off the skin, and binding the odorous molecules and precursors [4]. By understanding the metabolic pathways in *K. sedentarius* responsible for making the proteolytic enzymes and the odorous molecules, better treatments can be devised for pitted keratolysis. Previous interest in *K. sedentarius* has also been stimulated by the unique keratin proteases it produces. These degrading enzymes can potentially be used in industrial settings to degrade organic compounds [7].

K. sedentarius was separated from its original *Micrococcus* genus by a study which analyzed similarities in DNA and amino acid sequences across various members of the *Micrococcus* genus [8]. The methodology used in this study is very similar to the one used in the current study; however with modern databases, it is much easier and accurate to compare DNA and amino acid sequences, and there are many more sequences to compare as thousands more bacteria have been sequenced since 1995. Another study done in 2002 isolated and characterized two proteolytic enzymes that *K. sedentarius* uses to degrade keratin [1]. Research done in 2009 used DNA libraries and sequencing programs like BLAST for gene annotation of *K. sedentarius* [9]. All metabolic pathways were mapped and the genome was completely sequenced. However it is mentioned that the metabolic pathways were not reviewed and may contain errors. Also of note are the improvements in gene sequencing and annotation since this study. With more reliable programs and databases the gene annotation results could be improved [10]. Therefore the current study, which also uses gene annotation and similar programs, will be partially replicating some of the gene and pathway analysis done in the 2009 study.

The gene annotation process is inherently flawed due to the involvement of many sequence comparisons, limited databases, and programming mistakes. The current study hopes to discover more about genes that are involved in the DNA mismatch repair pathway. Here five genes that code for proteins in *K. sedentarius* are characterized using gene annotation. Multiple programs such as BLAST, TMHMM, PFAM, as well as many others were used in this study. This study hypothesized that the five genes called in the DNA mismatch repair pathway, would have functions that matched the ones predicted by the gene caller.

Materials and Methods

IMG-ACT Website

Mozilla Firefox was used to connect to the internet and was necessary for proper loading of the IMG-ACT web pages. The website login page was at http://img-act.jgi-psf.org/user/login.

Module 1: Basic Information

The gene details page was accessed to retrieve the DNA coordinates, nucleotide sequence, and amino acid sequence.

Module 2: Sequence Based-Similarity Data

The Basic Local Alignment Search Tool (BLAST) program was used to find similar proteins in other organisms besides *K. sedentarius* based on amino acid sequences. Clusters of Orthologous Groups (COGs) were also found on the BLAST results page using the Conserved Domain Database (CDD). It was started from http://blast.ncbi.nlm.nih.gov/Blast.cgi. Tree-Based Consistency Objective Function for Alignment Evaluation (T-Coffee) was used to create a multiple sequence alignment. T-Coffee was accessed using http://www.ebi.ac.uk/Tools/msa/tcoffee/. WebLogo was used to make a diagram that showed conserved amino acids between the sequences in the multiple sequence alignment. The site was accessed with http://weblogo.berkeley.edu/.

Module 3: Cellular Localization Data

Transmembrane Helices Hidden Markov Models (TMHMM) was used to find any transmembrane helices in the protein. Access to the site was available at http://www.cbs.dtu.dk/services/TMHMM/. SignalP was used for finding signal peptide sequences in the amino acid sequence. The webpage was at http://www.cbs.dtu.dk/services/SignalP/. PSORT-B was used to get estimations of the localization of the protein. It was accessed with http://www.psort.org/psortb/. Phobius was used to get a second opinion of the existence of transmembrane helices and signal peptides. It was accessed with http://phobius.sbc.su.se/.

Module 4: Alternative Open Reading Frame

This module was aimed at determining if the DNA coordinates given by the gene annotation computer program Glimmer were the most likely. This module was done by going to the gene details page and selecting the link Sequence Viewer For Alternate ORF Search.

Module 5: Structure-Based Evidence

TIGRFAM is a database that found the functional family for the protein product of each gene. It was accessed with http://blast.jcvi.org/web-hmm/. Pfam was able to give a family and clan for the query protein. It also made a sequence comparison between the Hidden Markov Model (HMM) sequence and the sequences of the matches it found. The website address was http://pfam.sanger.ac.uk/search. The Protein Data Bank (PDB) database was used to find the name, 3D crystal structure, and alignment of a protein that matched the query protein amino acid sequence. The site was accessed with http://www.rcsb.org/pdb/search/advSearch.do.

Module 6: Enzymatic Function

Kyoto Encyclopedia of Genes and Genomes (KEGG) is a database that has pathway maps which show where specific genes/proteins are involved. It was accessed with http://www.genome.jp/kegg/pathway.html. MetaCyc is another pathway database which stores pathways that have been experimentally evaluated. This was accessed with http://metacyc.org/. Expert Protein Analysis System (Expasy) was a database used to get more general information of the protein and it was accessed with http://enzyme.expasy.org/cgi-bin/enzyme/enzyme-search-ec.

Module 7: Duplication and Degradation

Prosite was used to find out if the amino acid sequence of a gene had the necessary residues and domains for its protein function. The website was accessed with http://prosite.expasy.org/scanprosite/.

Module 8: Horizontal Gene Transfer

A phylogenic tree was created to compare organisms that were evolutionary close to *K. sedentarius*. The website phylogeny.fr was accessed with http://www.phylogeny.fr/.

Ortholog Neighborhoods

The link Show neighborhood regions with this gene's bidirectional best hits on the gene details page was used to create a graphic that mapped the query gene as well as other genes in the DNA sequence of *K. sedentarius* which were close by. The graphic also included the mapping of four orthologs for comparison.

Results

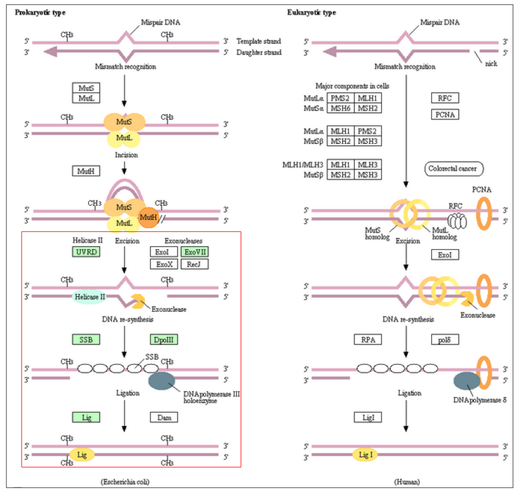

Figure 1: KEGG diagram of DNA mismatch repair pathway in prokaryotes (left) and eukaryotes (right). The five genes annotated in this study (red box) are highlighted in green.

Results

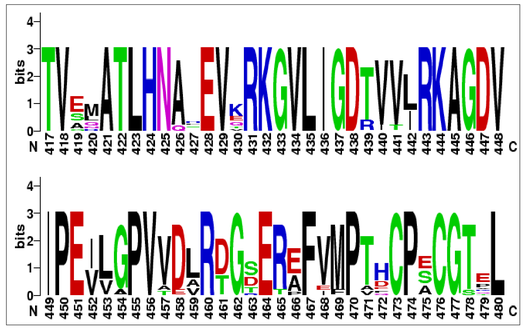

Figure 2: Portion of the WebLogo for gene OID 644991259 showing conservation. The larger the letter the more conservation between the collected homologous sequences and *K. sedentarius*.

Gene OID	TMHMM	SignalP	Phobius	PSORT-B
644991058	None	None	None	Cytoplasmic
644991171	None	None	None	Cytoplasmic
644992290	None	None	None	Cytoplasmic
644990318	None	None	None	Cytoplasmic
644991259	None	None	None	Cytoplasmic

Figure 3: Shows the results for TMHMM, SignalP, Phobius, and PSORT-B. No transmembrane helices or signal peptides were found in any of the amino acid sequences. Localization for all the proteins was determined to be cytoplasmic.

Gene OID	TIGRFAM	E values
644991058	uvrD: DNA helicase II	1.6e-217
644991171	single-strand binding protein	8.3e-06
644992290	exodeoxyribonuclease VII, large subunit	1.1e-49
644990318	DNA polymerase III, beta subunit	1.6e-78
644991259	None	None

Figure 4: Results from TIGRFAM which gave functional families for the protein products. All families matched predicted functions with the exception of ligase which had no family.

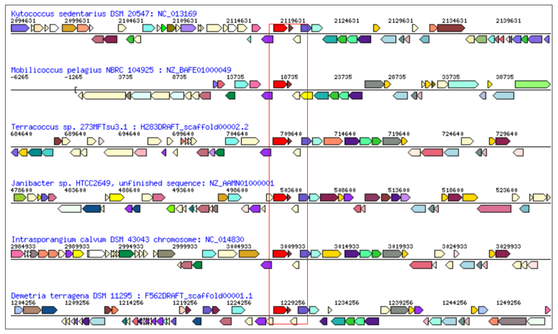

Figure 5: Gene neighborhood diagram showing gene OID 644992290 (red triangle) location in *K. sedentarius* and five orthologs. Alignment of the gene along with its dark purple neighboring gene in all the species (red box) indicates no evidence for horizontal gene transfer.

Results

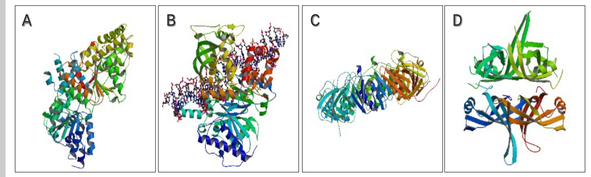

Figure 6: PDB crystal structures of (A) helicase, (B) ligase, (C) polymerase subunit, and (D) single stranded DNA binding protein.

Gene OID	Protein Product	EC Number
644991058	ATP-dependent DNA helicase PcrA	3.6.4.12
644991171	Single stranded DNA-binding protein	None
644992290	Exodeoxyribonuclease VII large subunit	3.1.11.6
644990318	DNA polymerase III, beta subunit	2.7.7.7
644991259	DNA ligase, NAD-dependent	6.5.1.2

Figure 7: The Gene OID numbers, protein products, and EC numbers for the five genes.

Conclusions

This study confirmed the proposed annotations of five genes involved in the DNA mismatch repair pathway of *K. sedentarius*. Searches with BLAST found the same amino acid sequences present in other organisms indicating a high amount of conservation. A high amount of sequence conservation was also seen in the WebLogos. All protein products were determined to have an cytoplasmic localization, from which they could interact with the DNA. The clans and functional families of the genes corresponded to their proposed functions. Pathway inquiries confirmed the genes functioned in other pathways involved in DNA repair and replication, in addition to mismatch repair. No evidence was found for the possibility of pseudogenes or horizontal gene transfer. The high conservation of these genes found in *K. sedentarius* and its orthologs supports the necessity of DNA repair regardless of an organism's natural environment. *K. sedentarius* is able to survive on humans where there is less environmental stress and more protection from factors like cold temperatures and ultraviolet radiation. However this protection has not resulted in the loss of proteins involved in DNA repair which is evidence for their continued importance.

Literature Cited

1. Longshaw, C.M., et al., *Kytococcus sedentarius, the organism associated with pitted keratolysis, produces two keratin-degrading enzymes.* Journal of Applied Microbiology, 2002. 93(5): p. 810-816.
2. James, A.G., D. Cox, and K. Worrall, *Microbiological and biochemical origins of human foot malodour.* Flavour and Fragrance Journal, 2013. 28(4): p. 231-237.
3. Chaudhary, D. and S.N. Finkle, Peritoneal dialysis-associated peritonitis due to *Kytococcus sedentarius.* Peritoneal Dialysis International, 2010. 30(2): p. 252-253.
4. Levenga, H., et al., *Fatal hemorrhagic pneumonia caused by infection due to Kytococcus sedentarius - a pathogen or passenger?* Annals of Hematology, 2004. 83(7): p. 447-449.
5. Hasegawa, Y., M. Yabuki, and M. Matsukane, *Identification of new odoriferous compounds in human axillary sweat.* Chemistry & Biodiversity, 2004. 1(12): p. 2042-2050.
6. Kanlayavattanakul, M. and N. Lourith, *Body malodours and their topical treatment agents.* International Journal of Cosmetic Science, 2011. 33(4): p. 298-311.
7. Ha, M., et al., *Comparison of the proteolytic activities of new commercially available bacterial and fungal proteases toward meat proteins.* Journal of Food Science, 2013. 78(2): p. C170-C177.
8. Stackebrandt, E., et al., *Taxonomic dissection of the genus Micrococcus: Kocuria gen. nov., Nesterenkonia gen. nov., Kytococcus gen. nov., Dermacoccus gen. nov., and Micrococcus Cohn 1872 gen. emend.* Int J Syst Bacteriol, 1995. 45(4): p. 682-92.
9. Sims, D., et al., *Complete genome sequence of Kytococcus sedentarius type strain (541(T)).* Standards in Genomic Sciences, 2009. 1(1): p. 12-U4.
10. Liu, W. and H.W. Xie, *Predicting potential cancer genes by integrating network properties, sequence features and functional annotations.* Science China-Life Sciences, 2013. 56(8): p. 751-757.