Annotation of the *Kytococcus sedentarius* Genome from Ksed_09020 to Ksed_09080

Kimberly Van Knapp, Spencer Wrazen, Andrea DiChristopher and Robert O'Connor
Kenmore East High School and the Western New York Genetics in Research Partnership
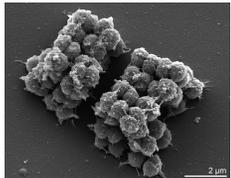
Abstract

A group of consecutive 3 genes from the microorganism *Kytococcus sedentarius* (Ksed_09020 – Ksed_09080) were annotated using the collaborative genome annotation website GENI-ACT. The Genbank proposed a gene product name for each gene was assessed in terms of their general genomic information, amino acid sequence-based similarity data, structure-based evidence from the amino acid sequence, cellular localization data. Numerous different databases including Blast, TIGRfam, PDB and Phobius were used in order to thoroughly investigate our genes and the functions they had so that we could either validate or invalidate our hypothesis. For the three researched genes, our proposed hypothesis were all confirmed by the information we gathered by using these different programs that we stored in our gene notebooks.

Introduction

A study of *Micrococcus* has led to a significant revision of this genus in 1995. As part of this change, the species *Micrococcus sedentarius* was renamed *Kytococcus sedentarius*. *Kytococcus sedentarius* is an aerobic, catalase positive, oxidase negative, spherical, gram-positive organism. It usually appears in tetrads or in cubical packets and maybe surrounded by a slimy gram-negative layer on smear.

Nevertheless, the bacterium has been documented to be a causative organism in various infections (Karsten Becker). Cases of pneumonia have been linked to Kytococcus Sedentarius during neutropenia following induction chemotherapy for acute myeloid leukemia (Levenga H.). It is most sensitive to ampicillin, cephalothin, ciprofloxacin, oxacillin, penicillin, tetracycline, and vancomycin (D. Chaudhary).

Current study involving the gene annotation process has not been entirely accurate as a result of the many sequence comparisons in the databases. Each database revolves around a hypothesis as to what it thinks it is, not what it is for sure. The programs in the modules on Geni-Act are designed to create an output of information to the individual using it, however, it may often at times lead to programming mistakes. Individuals currently working on improving this study are trying to discover improved ways to make the programs more accurate in the sense that these databases deliver better and more accurate information overall.

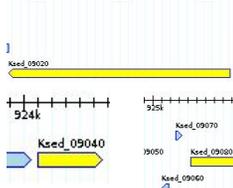Figure 1 shows K. Sedentarius under a high powered microscope

Figure 2 shows the three consecutive genes studied

Methods and Materials

Modules of the GENI-ACT (http://www.geni-act.org/) were used to complete *Kytococcus sedentarius* genome annotation. The modules used 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, on membranes, secreted or embedded in the membrane?

Each database in the modules are responsible for discovering more information on the gene being studied. The databases range from providing information on location, gene families, what it is, and what the function of it is. Images are also produced in these modules, which provide a visual understanding and help further the comprehension of the information being allocated.

Results

Kytococcus sedentarius 09020:

Aminopeptidase N (APN) is a Zn2+ dependent membrane-bound enzyme that degrades special proteins and peptides with an N-terminal neutral amino acid. It also an enzyme, related with tumorigenesis, immune system, pain etc. In humans, this enzyme can serve as a receptor for human viruses like the human coronavirus which is an important cause of upper respiratory tract infections. Aminopeptidase N has been associated with the growth of different human cancers and is commonly known to be a suitable target for anti-cancerous therapy. Defects in this gene appear to be a cause of various types of leukemia or lymphoma. The aminopeptidase N (APN) exists in two forms such as the membrane aminopeptidase N and the soluble aminopeptidase N. Bacteria display several aminopeptidasic activities which may be localised in the cytoplasm, on membranes, associated with the cell envelope or secreted into the extracellular media (7). psortB predicted a cytoplasmic localization for Ksed_09020, as shown in Figure 3 below. No transmembrane helixes were detected by TMHMM or Phobius, further supporting the cytoplasmic localization of Ksed_09020 The substrate specificity of this particular aminopeptidase in *Kytococcus sedentarius* is unknown.

```
Localization Scores:
     Cytoplasmic               9.97
     CytoplasmicMembrane       0.00
     Cellwall                  0.01
     Extracellular             0.02
Final Prediction:
     Cytoplasmic               9.97
```

Figure 3. The psortB output for Ksed_09020. Results clearly indicate that the aminopeptidase encoded by Ksed_09020 is located in the cytoplasm.

Kytococcus sedentarius 09040:

Ksed_09040 has 462 nucleotides and 153 amino acids in the sequence. The results of using the GenBank database of genes showed that the *Kytococcus sedentarius* gene Ksed_09040 most closely resembles the ribose-5-phosphate isomerase protein. The family of this protein is a member of the RpiB/LacA/LacB subfamily (TIGR00689) but lies outside the RpiB equivalog (TIGR01120) which is also a member of that subfamily. This protein directly catalyzes the reversible conversion between ribose-5-phosphate and ribulose-5-phosphate in the pentose-phosphate pathway. The enzyme has been given the EC number 5.3.1.6. TMHMM and SignalP indicate that the protein can be found in the cytoplasm and is not a trans-membrane protein. Using WebLogo, the gene sequence is shown to be well conserved throughout most of the sequence. This is shown by the overall height of the stacks in the image.
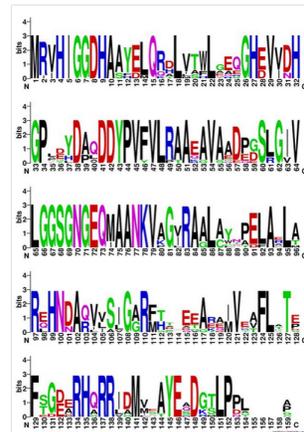
Figure 4. WebLogo of the T-Coffee alignment of orthologs of Ksed_09040. High conservation is seen throughout most of the alignment.

Kytococcus sedentarius 09080:

Ksed_09080 has 1359 nucleotides and 452 amino acids in its sequence. Based on what was found using the Geni-Act databases, it is a peptidyl prolyl cis-trans isomerase enzyme. It catalyzes the cis-trans isomerisation of proline-imidic peptide bonds in oligopeptides. It's gene product name is a trigger factor which is a ribosomal protein that provides a protective nano environment for nascent polypeptides at the ribosome[1]. No helixes were present in further investigations through using TMHMM and further supporting that by using SignalP. Since the lines on the graph do not pass the top threshold line, it predicts that there is no results found for it to be a trans membrane protein and further concluding it is found in the cytoplasm.

SignalP-4.1 gram+ predictions
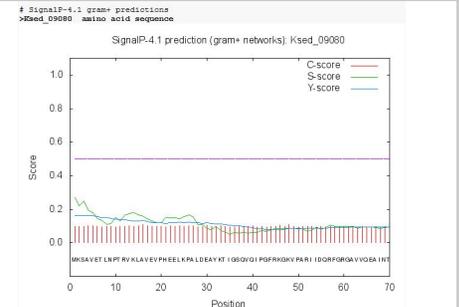>Ksed_09080 amino acid sequence

Figure 5. SignalP output for Ksed_09040. No signal peptide is present since it doesn't cross the threshold line for Ksed_09080
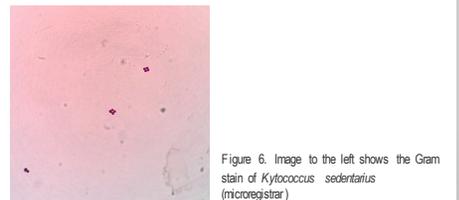
Figure 6. Image to the left shows the Gram stain of *Kytococcus sedentarius* (microregistrar)

Conclusions

Based on the results found using the databases in Geni-Act, Ksed_09020, Ksed_09040 and Ksed_09080 had their cellular localization found in the cytoplasm. All of the results are fully supported based upon the findings through Geni-Act databases. Ksed_09020 was determined to be an aminopeptidase. Ksed_09040 was determined to be ribose-5-phosphate isomerase. Ksed_09080 was determined to be peptidyl prolyl cis-trans isomerase.

References

1. http://medical-dictionary.thefreedictionary.com/peptidyl+prolyl+cis-trans+isomerase+D
2. http://onlinelibrary.wiley.com/doi/10.1111/j.1349-7006.2010.01826.x/full
3. http://www.ncbi.nlm.nih.gov/pubmed/17346152
4. http://www.creativebiomart.net/symbolsearch_ANPEP.htm?gclid=CKWLuID_1MwCFVhbhqodFOwC6A
5. http://www.ncbi.nlm.nih.gov/Structure/cdd/cddsrv.cgi?uid=TIGR02133
6. http://www.genecards.org/cgi-bin/carddisp.pl?gene=RPIA
7. http://www.ncbi.nlm.nih.gov/pubmed/8703509#

Acknowledgments