Annotation of the *Kytococcus sedentarius* Genome from DNA Coordinates 468392 to 471684

Michael Hagleberger, Shawn Kirsch, Joshua Minser, Elisabeth Rebmann, Lindys VonOhlen and Laura O’Donnell

Attica High School and The Western New York Genetics in Research Partnership

Abstract

A group of consecutive 5 genes from the microorganism *Kytococcus sedentarius* (Ksed, 04710 – Ksed, 04750) were annotated using the collaborative genome annotation website GENI-ACT. The Genbank proposed gene product name for each gene was assessed in terms of the general genomic information, amino acid sequence-based similarity data, structure-based evidence from the amino acid sequence, cellular localization data, and potential alternative open reading frames. The Genbank proposed gene product name 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.

Introduction

*Kytococcus sedentarius* is a strictly aerobic, non-motile, non-encapsulated, and non-endospore forming gram positive cocoid bacterium, found predominantly in tetrad formation. This organism is encapsulated, and non-endospore forming gram positive coccoid bacterium, found predominantly in tetrad formation. This organism is isolated from a microscope slide submerged in sea water in 1944, *Kytococcus sedentarius* grows well in sodium chloride at concentrations less than 10% (w/v).

According to Sims et al. (2009), *Kytococcus sedentarius* is a microorganism of interest for several reasons. This bacterium is a natural source of the oligoketide antibiotics monomycin A and monomycin B (Sims et al., 2009). *Kytococcus sedentarius* has been implemented as the eukaryotic agent of a number of opportunistic infections including valve endocarditis, hemophagocytic pneumonia, and pitted keratolysis (Sims et al., 2009). Finally, the phylogeny of this microorganism is a source of interest, as it is a member of the family Deltaproteobacteria within the actinobacterial suborder Mycobacterales, which has yet to have been thoroughly studied utilizing bioinformatics (Sims et al., 2009).

Methods and Materials

Modules of the GENI-ACT (http://www.geni-act.org) were used to complete the *Kytococcus sedentarius* genome annotation. The modules are described below:

<table>
<thead>
<tr>
<th>Modules</th>
<th>Activities</th>
<th>Questions Investigated</th>
</tr>
</thead>
<tbody>
<tr>
<td>Module 1: Basic Information</td>
<td>DNA Coordinates and Sequence, Protein Sequence</td>
<td>What is the sequence of my gene and protein? Where is it located in the genome?</td>
</tr>
<tr>
<td>Module 2: Sequence Similarity Data</td>
<td>Blast, COG, TCOFFEE, Weblogo</td>
<td>Is my sequence similar to other sequences in GenBank?</td>
</tr>
<tr>
<td>Module 3: Cellular Localization Details</td>
<td>Gram Stain, TAH/MVM, SignalP, PSORTb</td>
<td>Is my protein in the cytoplasm, extruded or embedded in the membrane?</td>
</tr>
<tr>
<td>Module 4: Alternative Open Reading Frame</td>
<td>IMG Sequence Viewer, For Alternative ORF Search</td>
<td>Has the amino acid sequence of my protein been called correctly by the computer?</td>
</tr>
<tr>
<td>Module 5: Structure Based Evidence</td>
<td>TIGR sites, Pfam, PDB</td>
<td>Are there functional domains in my protein?</td>
</tr>
</tbody>
</table>

Results

*Kytococcus sedentarius* 04710:

The COG hits for Ksed, 04710 suggest it may be a hypothetical protein. According to the proposed enzymatic activity of this gene product, it is likely to be a serine protease inhibitor, the cleavage of peptide bonds in proteins. The gene was consecutively found to be most similar to phospholipid-binding protein, and structure based evidence. No alternate open reading frames were detected. With the lack of evidence discovered, it is possible that this gene was called wrong. It may refer to as a hypothetical protein.

*Kytococcus sedentarius* 04720:

The protein product for this gene binds ATP, opioids and phosphatidylethanolamine. This gene is thought to be a serine protease inhibitor, the cleavage of peptide bonds in proteins. The gene was consecutively found to be most similar to phospholipid-binding protein, and structure based evidence. No alternate open reading frames were detected. With the lack of evidence discovered, it is possible that this gene was called wrong. It may refer to as a hypothetical protein.

*Kytococcus sedentarius* 04730:

The protein product predicted for this gene is a cytoplasmic enzyme based on its TIGR03083. TIGR03083 is a family of uncharacterized proteins that are predicted to have enzymatic functions. Its PFAM, PF01716 is a member of the super family DinB.2 which is an uncharacterized family of potential enzymes. It is also predicted to be cytoplasmic because it contained no transmembrane helices or signal peptides. There were no COG hits and according to TCOFFEE and WEBLOGO it was fairly well conserved.

*Kytococcus sedentarius* 04740:

The protein product predicted for this gene is a cytoplasmic enzyme, and most likely a racemase enzyme, as shown through its Pfam of D-threo-3-hydroxyaspartate dehydratase and aldolase. The low amount of conservation of this large gene could be due to the predicted enzymatic activity of this gene product, it is likely to be a serine protease inhibitor, the cleavage of peptide bonds in proteins. The gene was consecutively found to be most similar to phospholipid-binding protein, and structure based evidence. No alternate open reading frames were detected. With the lack of evidence discovered, it is possible that this gene was called wrong. It may refer to as a hypothetical protein.

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. The exception to this is the 04710 gene locus which may be a hypothetical protein.

References

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

Acknowledgments

Supported by NSF ITEST Strategies Award Number 1311902

Figure 1 – Cellular Localization Data for *Kytococcus sedentarius* (04720, 04730, 04740, and 04750). All are predicted to be cytoplasmic proteins. No significant cellular localization data was available for the gene locus 04710, which is predicted to be a hypothetical protein. Due to the predicted enzymatic functions, it is probable that these enzymatic pathways are involved in reactions within the cytoplasm of the bacterial cell.