**Annotation of the *Kytococcus sedentarius* Genome from DNA Coordinates 886597 to 899459**

Amina Adillahi, Amina Ali, Filmon Asmelash, Gospel Djongara, Radhika Chapagain and Jeffrey Besinger

Newcomer Academy at Lafayette High School and the Western New York Genetics in Research Partnership

---

**Abstract**

A group of consecutive 5 genes from the microorganism *Kytococcus sedentarius* (Ksed_08680 – Ksed_08770) were annotated using the collaborative genome annotation website GEN-I-AC-T. The objective of this study was to introduce students to techniques involved in determining gene function by doing that the computer correctly annotated the *Kytococcus sedentarius* genome. The gene product name was proposed by GenBank and each gene was assessed using various search tools (**BLAST**, **TigR** and **N**). The GenBank proposed gene product name did not offer significant from the proposed gene annotation for each gene in the group.

**Methods and Materials**

**Modules of the GENI-AC-T** were used to complete *Kytococcus sedentarius* genome annotation.

The modules are described below:

- **Module 1**: Basic Information Module
  - DNA Coordinates and Sequence, Protein Sequences
  - What is the sequence of my gene and protein? Where is it located in the genome?

- **Module 2**: Sequence-Based Similarity Data
  - T-Coffee, Pfam,ign3
  - It is a sequence similarity to other sequences in GenBank?

- **Module 3**: Cellular Localization Data
  - Gram Stain, TIGRfam, SignalPSORT, Phobius
  - Is my protein in the predicted compartment or not?

- **Module 4**: Alternative Open Reading Frame
  - IMGS/Sequence Viewer For Alternative 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 these functional domains in my protein?

**Results**

**Kytococcus sedentarius 08780**

Ksed_08780 was characterized as containing a DNA binding domain, which was confirmed by BLAST and TIGRfam. Pfam match characterized the sequence as containing a Helix-Turn-Helix domain. We did some additional research and determined that the Helix-Turn-Helix domain is a protein structure that binds to DNA. Despite the fact that the domains had different names, their functions are similar, therefore Pfam results could be considered supporting the original annotation.

**Figure 4**: Results of Pfam for Ksed_08780 showing a significant match to “Helix-Turn-Helix domain”

**Kytococcus sedentarius 08740**

Ksed_08740 was characterized as containing a delta-1-pyrroline-5-carboxylate dehydrogenase, group 1 which was confirmed by BLAST, PDB, TIGRfam and PDB. BLAST results show a value as low as 0.0 showing an extremely high amount of sequence similarities.

**Figure 5**: Shows an example of a “Helix-Turn-Helix domain” binding to DNA

**Kytococcus sedentarius 08700**

Ksed_08700 was characterized as containing a DN-A binding domain, which was confirmed by BLAST, and TIGRfam. Pfam match characterized the sequence as containing a Helix-Turn-Helix domain. We did some additional research and determined that the Helix-Turn-Helix domain is a protein structure that binds to DNA. Despite the fact that the domains had different names, their functions are similar, therefore Pfam results could be considered supporting the original annotation.

**Figure 6**: Results of BLAST for Ksed_08700 showing a values of 0.0

**Kytococcus sedentarius 08680**

Ksed_08680 was characterized as containing a lipM domain, which was confirmed by Pfam. TIGRfam also revealed that the gene contains 3 regions of the m-in-a-domain sequence that would result in the transmembrane composition of lipids and transmembrane helixas.

**Figure 3**: Results of TIGRfam for Ksed_08680 indicating 3 possible transmembrane helixas

**Kytococcus sedentarius 08770**

Ksed_08770 was characterized as containing a NAD-glutamate dehydrogenase which was confirmed by BLAST and Pfam. TIGRfam indicated that the sequence was likely to have a histidine domain, a protein function that supports the hypothesis that glutamate dehydrogenase has a histidyl binding site which confirms the original annotation.

**Figure 8**: Results of Pfam for Ksed_08770 showing e-values of 0.0

**Kytococcus sedentarius 08770**

Ksed_08770 was characterized as containing an NAD-glutamate dehydrogenase which was confirmed by BLAST and Pfam. TIGRfam indicated that the sequence was likely to have a histidine domain, a protein function that supports the hypothesis that glutamate dehydrogenase has a histidyl binding site which confirms the original annotation.

**Figure 9**: Shows a portion of weblogo and 10 additional NAD-glutamate dehydrogenases with a high amount of conservation.

---

**Conclusion**

The GENI-AC-T 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. TMMHMM predicted additional functions regarding cellular localization of Ksed_08680 and Ksed_08770.

**References**

Sims et al. (2009). Complete genome sequence of *Kytococcus sedentarius* type strain (S411). Standards in Genomic Sciences, 12 - 20.

**Acknowledgments**

Supported by NSF-TEST Strategies Award Number 1131902