**Abstract**

A group of four consecutive genes from the microorganism *Kytococcus sedentarius* (Locus tag: Ksed_05820 – Ksed_05850) were annotated using genome annotation modules in GENI-ACT (Genomics Educational National Initiative- Annotation Collaboration Toolkit). The objective was to determine if the genes had been annotated correctly in GenBank. The proposed gene product name for each gene was evaluated by exploring modules 1, 2, 3, and 4 that include the basic information of the gene, amino acid sequence based similarity data, cellular localization data, and structure based evidence. The proposed gene product names for all four genes 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 in the database.

**Introduction**

*Kytococcus sedentarius* is a gram positive bacteria that is responsible for the disease pitted keratolysis that can become serious in people with immunocompromised systems. It is known that automated annotations include up to 35% error rates. This was one of the many reasons for undertaking manual annotation of the genes. Our objective was to make sure the gene was identified correctly and its function was predicted accurately.

According to S. Pospisil, et al. (1998) *Kytococcus sedentarius* is a producer of the antibiotic monensin, an antibiotic often used in animal food products. *Kytococcus sedentarius* is also suspected to be responsible for the death of a 55 year old man who had undergone chemotherapy. With chemotherapy, the man was severely immunocompromised, thus *K. sedentarius* broke its way into the blood stream, ultimately reaching and invading the lungs, causing major, and fatal, hemorrhagic pneumonia (Levenga, et al. 2003).

**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 1: Basic Information**
  - **Activity:** Chuk Coordinates and Sequence, Protein Sequence
  - **Question Investigated:** What is the sequence of my gene and protein? Where is it located in the genome?

- **Module 2: Sequence-Based Data**
  - **Activity:** Blast, CDD, T-Coffee
  - **Question Investigated:** Is my sequence similar to other sequences in the database?

- **Module 3: Cellular Localization Data**
  - **Activity:** SignalP, PSORT-B
  - **Question Investigated:** Is my protein in the cytoplasm, secreted or extracellular?

- **Module 4: Alternative Open Reading Frame**
  - **Activity:** TIGRfam, Pfam, PDB
  - **Question Investigated:** Are there functional domains in my protein?

- **Module 5: Structure-Based Evidence**
  - **Activity:** each specific tool
  - **Question Investigated:** Are there conserved regions in my protein or localizations?

**Results**

**Ksed_05820**

The initial proposed product of this gene by GENI-ACT was Pyridoxamine 5'-phosphate oxidase. This gene product was supported by the BLAST hits along with CDD and T-Coffee results. There does not seem to be a difference in the product determined by both the proposed person and personal research. Web logo provides info that the gene also has a range of well-conserved amino acids. Using TMHMM cellular localization tool, there were no transmembrane helices, predicted. SignalP tool indicated a “YES” with a cleavage site between position 16 and 17 for the presence of a signal peptide. However, the c, s, and y plots in the graphical output do not support that conclusion. The D score however does exceed the cut-off value and thus the prediction is a “yes”. The final prediction by PSORT-B with a score of 8.91, is extracellular.

**Ksed_05840**

The initial proposed product of this gene by GENI-ACT was a phosphoribosylaminomazole-succinocarboxamidase synthase. This gene product proposal was supported by BLAST and CDD searches (nr database). T-COFFEE analysis followed by WebLogo revealed that the protein was well conserved indicating important functionality.

**Ksed_05850**

The initial proposed product of this gene by GENI-ACT was a death-on-curing protein. BLAST search against the nr database returned a moderately high significant match to a similar protein in *S. cinnamonensis*. Multiple alignment of amino acid sequence from ten orthologs using T-Coffee followed by WebLOGO indicates the protein to be well conserved. TMHMM, SignalP and Phobius do not show any transmembrane helices or signal peptide. PSORTb shows the highest score for cytoplasmic membrane. The cellular localization of the protein is unclear and needs further investigation. According to the top hit of TIGRfam this gene is in a Death-On-curing protein family.

**Conclusions**

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 automated methods.

**References**

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

**Acknowledgments**

We would like to thank Drs. Rama Dey-Rao and Stephen Koury for assisting with the training and implementation of this project.

Supported by NSF ITEST Strategies Award Number 1311902