Annotation of the *Kytococcus sedentarius* Genome at Locus Tag Ksed_07720

Calvin Dimmig, Robert Wolfe, Dylan Sileo, Devon Eleey, Andrea DiChristopher and Robert O'Connor
Kenmore East High School and the Western New York Genetics in Research Partnership

University at Buffalo

Abstract

One specific gene, Ksed_07720, from the microorganism *Kytococcus sedentarius* was annotated using the collaborative genome annotation website GENI-ACT. The Genbank proposed that the gene product name was DNA Helicase II and this was assessed in terms of basic genomic information, sequence-based evidence from the amino acid sequence, structure-based evidence from the amino acid sequence, cellular localization data, enzymatic function, and the possibility of horizontal gene transfer. The Genbank proposed gene product name did not differ significantly from the proposed gene annotation for the specific gene and as such, the gene appears to be correctly annotated by in the databases.

Introduction

Kytococcus sedentarius is a nonmotile, free-living, Gram-positive bacterium. This organism is the only known producer of the antibiotics monensin A and B and can be found in various environments like human skin, groundwater and airline cabins (BACMAP). Thus, the importance of this research would lead to more knowledge on the function of this specific gene.

While there is some known information on *Kytococcus sedentarius*, much is also unknown about it. The gene products have been predicted for the *Kytococcus sedentarius* genome, but they have not been confirmed to be accurate. It is also unknown of the exact role that our gene of interest plays in the organism itself.

The purpose of this research was to confirm the GENI-ACT database's annotations for the unknown gene and locus_tag in the *Kytococcus sedentarius* genome.

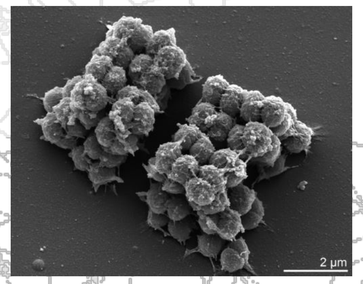

Figure 1 – Scanning electron micrograph of *Kytococcus sedentarius*

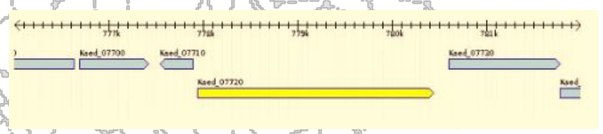

Figure 2 - The locus tag and relative position of the gene under investigation in this research

Methods

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

Modules	Activities	Question Investigated
Module 1: Basic Information	DNA Coordinates and sequence and amino acid sequence	What are the basics behind the DNA structure relevant to our gene annotation?
Module 2: Sequence-Based Similarity Data	BLAST, CDD, T-Coffee, and WebLogo	Is the protein annotated similar to other known proteins in other databases?
Module 3: Structure-Based Evidence	TIGRFAM and PFAM	Does the annotated protein have similar functionalities to other proteins?
Module 4: Cellular Localization Data	TMHMM, SignalP, PSORTb, and Phobius	Where is the protein annotated located?
Module 6: Enzymatic Function	KEGG and MetaCyc	What is the document or pathway in which the protein coded by the gene annotated function?
Module 8: Horizontal Gene Transfer	BLAST, T-Coffee, Phylogeny, and NCBI	Is there evidence that the gene has arisen by horizontal transfer as opposed to vertical transfer?

Results

The initial proposed product of this gene by GENI-ACT was DNA Helicase II. This gene product proposal was supported by the top BLAST hits for the amino acid sequence, and the cellular location of the amino acid sequence. The coordinates of our specific gene in the DNA sequence was determined to be 777948 to 780443 nucleotides, we further investigated this location in cellular localization. It was also determined that the nucleotide sequence which was 2496 nucleotides long, and the amino acid sequence which was 831 amino acids long. The amino acid sequence was used to compare proteins, and BLAST assisted us in researching any matches. Our closest match was DNA Helicase II with an alignment length of 720, a score of 518 bits, E-value of 1e-172. The first match has the highest chance of being similar to our protein due to it's low E-value and having a higher score than the second protein. This means that our protein that the gene codes for has a higher probability of being a DNA Helicase II rather than an ATP-Dependent Helicase Rep. We also used T-Coffee and BLAST together to determine the protein's multiple sequence alignment. In TIGRFAM suggested that the name of our name was "DNA Helicase II. The PDB presented an image displaying an x-ray diffracted macromolecular 3d structure of a DNA Helicase (Figure V). COG gave us results that gave the COG number being COG 0210 and the protein being apart of a superfamily of DNA or RNA helicase. COG also claimed that the protein was used in replication, recombination, and repair. Some observations from the WebLogo results were that the protein is well conserved throughout the whole sequence. We also found out that the Gram stain of the microbe is Gram positive.

The location of our gene was determined by using TMHMM and PSORT-B. Our Transmembrane topology graph (TMHMM) displays in figure III that our gene is outside the transmembrane with a probability score of 100% with 0 transmembrane helices predicted throughout the gene sequence. PSORT-B calculated that the gene is most likely localized in the cytoplasm with a final prediction depicting our localization score of 9.97 for cytoplasmic. KEGG displayed the process of DNA replication, specifically showing the role of a DNA Helicase II. MetaCyc displayed a pathway map in figure VI which depicts the reactions that our enzyme is involved with as well as an atom map .

```
# Ksed_07720 Length: 831
# Ksed_07720 Number of predicted TMHs:  0
# Ksed_07720 Exp number of AAs in TMHs: 0.0104
# Ksed_07720 Exp number, first 60 AAs:  0
# Ksed_07720 Total prob of N-in:        0.00047
Ksed_07720      TMHMM2.0          outside     1   831
```

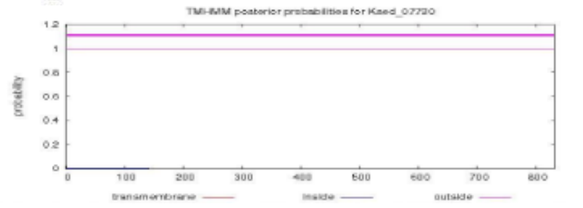

Figure 3 - TMHMM posterior probabilities

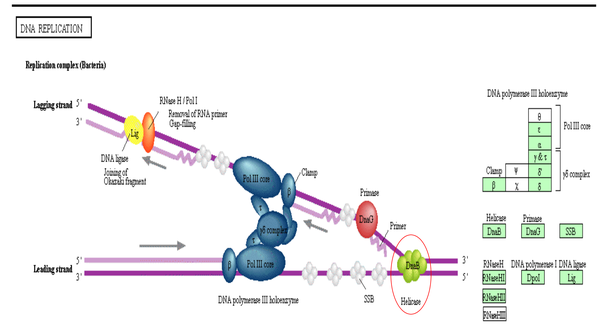

Figure 4 – Process of DNA replication with emphasis on DNA Helicase II

Figure 5 – X-ray diffracted DNA Helicase II

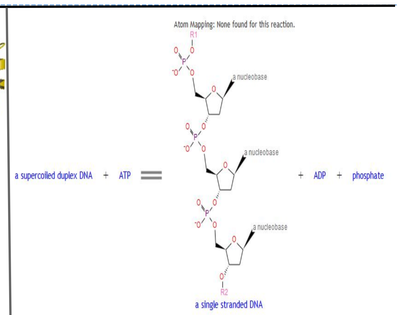

Figure 6 - Reaction pathway map

We used BLAST to find numerous other bacterium related to *Kytococcus sedentarius* like *Janibacter*, *Branchiibus*, and *Demetria*. Figure 7 displays a cladogram which shows that *Kytococcus* is in a clade with no other bacteria, but it shares a common ancestor with numerous other bacteria from the BLAST hits. Results from Phylogeny show that *Kytococcus sedentarius* has the same ancestors as from T-Coffee. We also used the Taxonomy Browser from NCBI and searched *Kytococcus sedentarius* which proved that it was from the phylum called Actinobacteria. Also apart of this Phylum are the bacteria that *Kytococcus* has similar ancestors to like *Streptomyces, Actinospica, and Kitasatospora* as seen from NCBI.

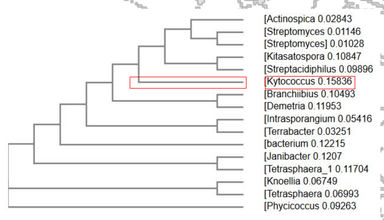

Figure 7 - *Kytococcus* and fifteen other bacteria from BLAST hits on a cladogram.

Conclusion

The proposed gene product from GENI-ACT did not differ from our initial hypothesis of our gene. It appears that the gene has been correctly annotated by the databases used, with a gene locus of 07720, the GENI-ACT gene product being DNA Helicase II, and our final proposed annotation leading us to a similar conclusion of DNA Helicase II. The proposed annotation is supported by our BLAST hits which compared our gene most closely to DNA Helicase II. TIGRFAM and PDB suggest that the name and structure of our gene is DNA Helicase II. This is further supported by our KEGG and MetaCyc pathway which both display DNA Helicase II. We concluded that our protein is not clustered with proteins from closely related bacteria, meaning that horizontal gene transfer is possible.

References

KEGG Pathway Database. KEGG, www.genome.jp/kegg/pathway.html.
"Kytococcus sedentarius DSM 20547." *BACMAP Genome Atlas*, BACMAP, bacmap.wishartlab.com/organisms/953.
"MetaCyc Metabolic Pathway Database." *MetaCyc*, 2016, metacyc.org/. Accessed 7 May 2017.
"NCBI." *NCBI*, National Center for Biotechnology Information, www.ncbi.nlm.nih.gov/. Accessed 12 May 2017.
PFAM. EMBL-EBI, 2014, pfam.xfam.org/. Accessed 7 Mar. 2017.
T-Coffee Multiple Sequence Alignment. EMBL-EBI, 2017, www.ebi.ac.uk/Tools/msa/tcoffee/. Accessed 24 Apr. 2017.
TMHMM Server v. 2.0. Center for Biological Sequence Analysis, 5 Jan. 2017, www.cbs.dtu.dk/services/TMHMM/. Accessed 13 Mar. 2017.

Acknowledgments

Supported by an NSF Innovative Technology Experiences for Students and Teachers (ITEST) Award - 1311902

www.buffalo.edu