Annotation of the *Kytococcus Sedentarius* Genome from DNA Coordinates 35601 to 38458

Nathan Kollatz, Madison Stevens, Tyler Rohde, Rachael Bilek and Jeff Salerno
Lake Shore High School and the Western New York Genetics in Research Partnership

University at Buffalo

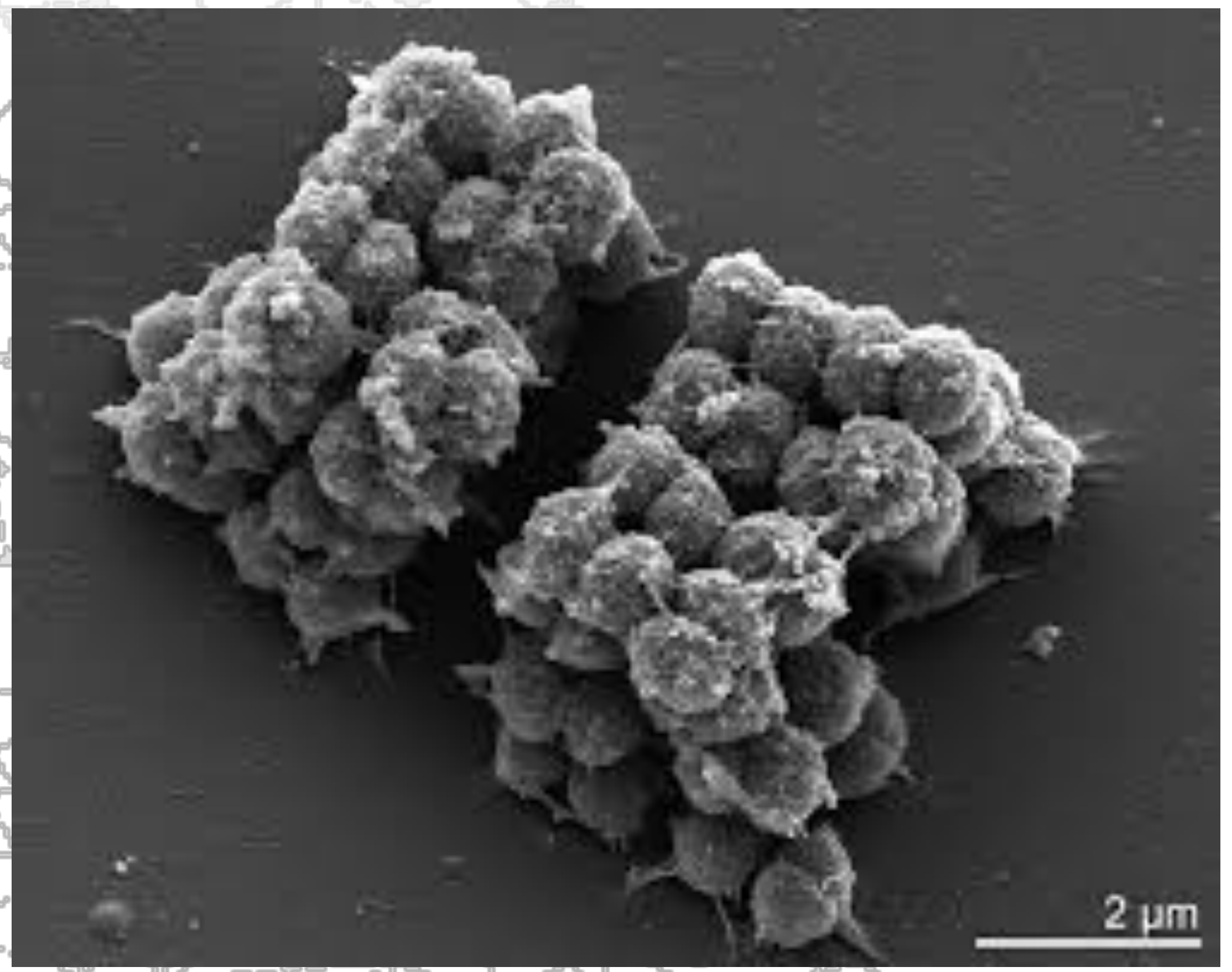
Abstract

A group of consecutive 3 genes from the microorganism *Kytococcus sedentarius* (Ksed_00320 – Ksed_0340) 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, potential alternative open reading frames, enzymatic function, presence or absence of gene duplication and degradation, the possibility of horizontal gene transfer, and the production of an RNA product. 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 in the r database.

Introduction

Kytococcus sedentarius is a strictly aerobic, non-motile, non-encapsulated, and non-endospore forming gram positive coccoid bacterium, found predominantly in tetrad formation. This organism is classified as a chemoheterotroph, as it requires methionine and several other amino acids for growth. Originally 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 monensin A and monensin B (Sims et al., 2009). *Kytococcus sedentarius* has been implemented as the etiological agent of a number of opportunistic infections including valve endocarditis, hemorrhagic 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 *Dermacoccaceae* within the actinobacterial suborder *Micrococcineae*, which has yet to have been thoroughly studied utilizing bioinformatics (Sims et al., 2009).



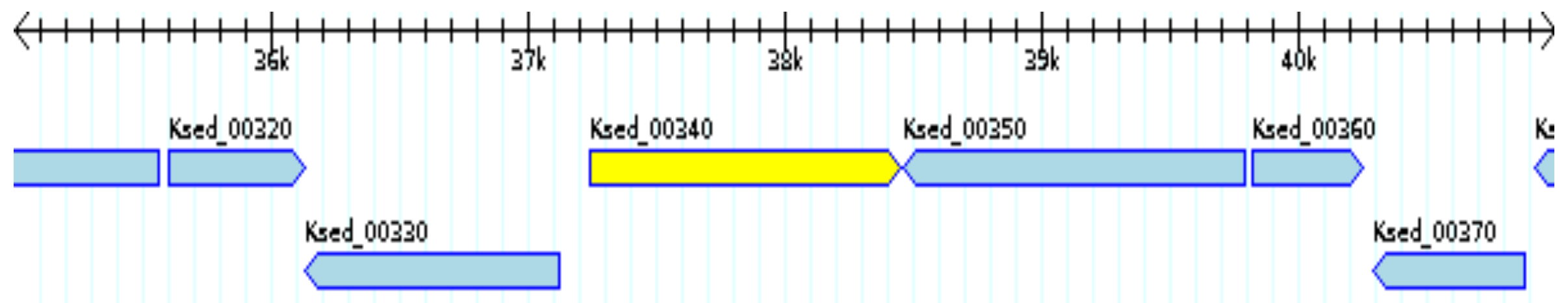


Figure 1. The locus tags and relative position of the genes under investigation in this research

Methods

Modules of the GENI-ACT (http://www.geni-act.org/) were used to complete Kytococcus sedentarius genome annotation . The modules 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- Cellular Localization Data	Gram Stain, TMHMM, SignalP, PSORT, Phobius	Is my protein in the cytoplasm, secreted or embedded in the membrane?
Module 4- Alternative Open Reading Frame	IMG Sequence Viewer For Alternate 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 there functional domains in my protein?
Module 6- Enzymatic Function	KEGG, MetaCyc, E.C. Number,	In what process does my protein take part?
Module 7- Gene Duplication/ Gene Degradation	Paralog, Pseudogene	Are there other forms of my gene in the bacterium? Is my gene functional?
Module 8- Evidence for Horizontal Gene Transfer	Phylogenetic Tree,	Has my gene co-evolved with other genes in the genome?
Module 9- RNA	RFAM	Does my gene encode a functional RNA?

Results

Ksed_00320:
The initial proposed product of this gene by GENI-ACT was a hypothetical protein. This gene product proposal was supported by the top BLAST hits for the amino acid sequence. The product had gene coordinates from 35601 to 36134 and a sequence length of 534 bp, and an amino acid sequence length of 177 aa. With a score of 333 and an E value of 2e-115 the results are a very close match the presence of well-curated protein functional domains within the amino acid sequence, the transmembrane topography of the amino acid sequence, and the cellular location of the amino acid sequence. As such, the proposed annotation is a hypothetical protein.

Ksed_00330:
The initial proposed product of this gene by GENI-ACT was a short-chain dehydrogenase. This gene product proposal was supported by the top BLAST hits for the amino acid sequence, since the positives and identities are a 87% match, with the E-Value at 0.0. and the enzymatic function of the amino acid sequence. With minor differences, the proposed annotation is a short-chain dehydrogenase.

Ksed_00340:
The initial proposed product of this gene by GENI-ACT was a glutathione-dependent formaldehyde dehydrogenase. This gene product proposal was supported by the top BLAST hits for the amino acid sequence. The gene product had the gene coordinates 37247 to 38458 with a nucleotide sequence length of 1212 bp and an amino acid sequence length of 403 aa. With a score of 823 bits and an E-value of 0.0, this was a 100% match with the gene. The proposed gene product top COG hit was COG1063. This hit was named Threonine dehydrogenase. It had an E-value of 2.81e-89. In my WebLogo i had a majority of large letters that was consistent with few gaps in the alignment.

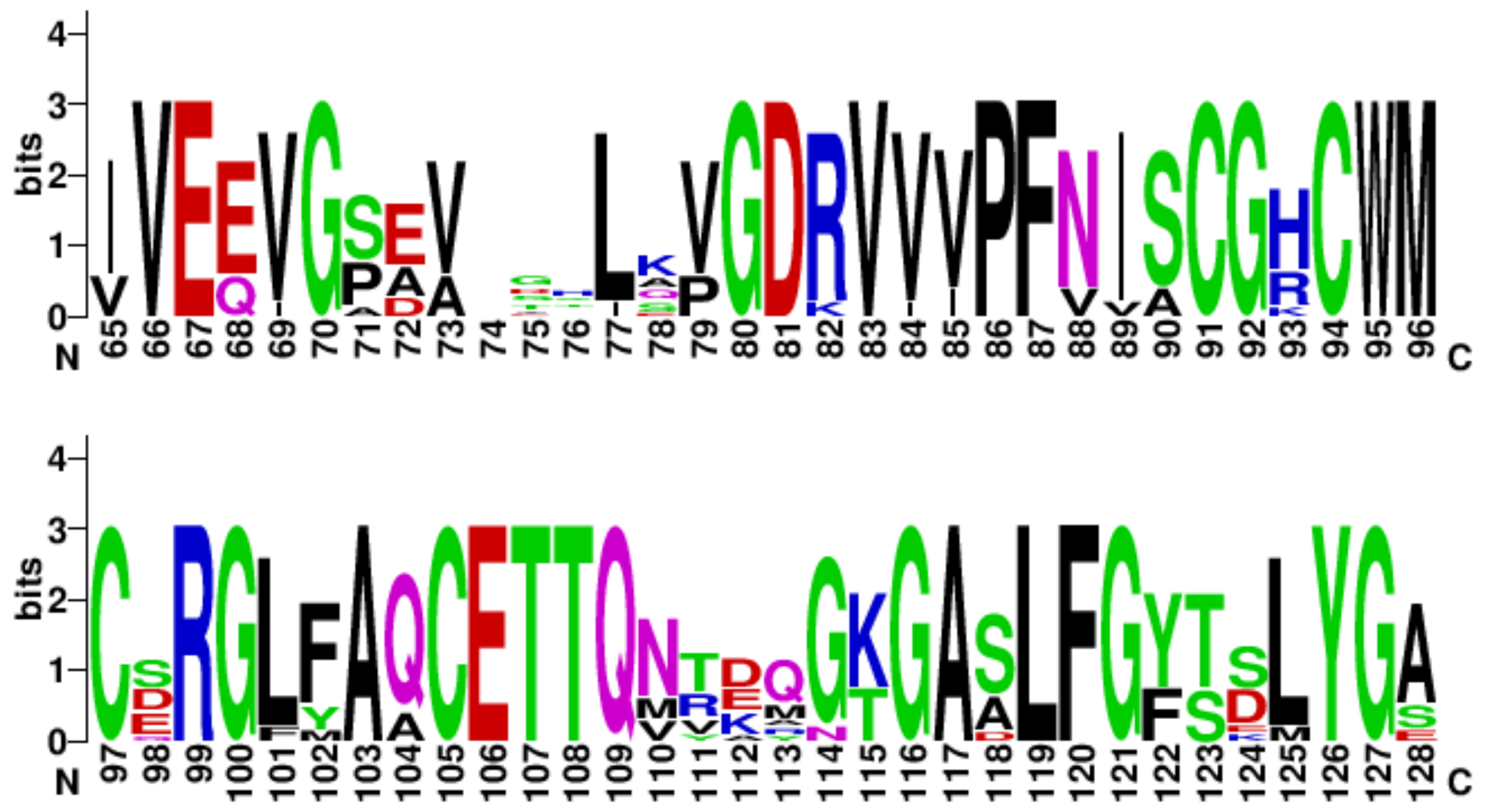
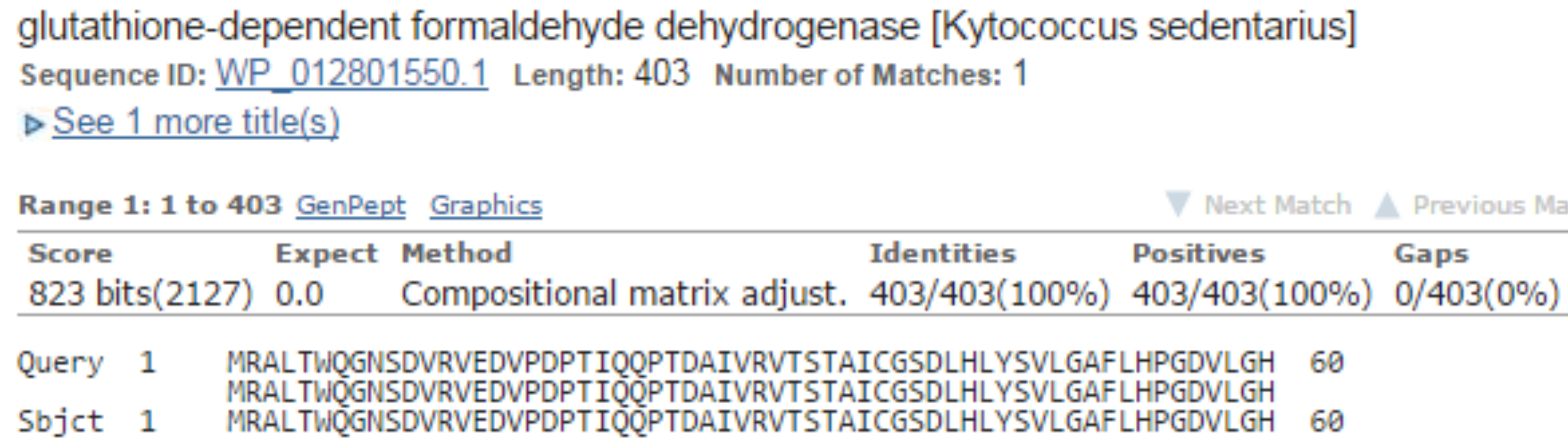


Figure 2. - Kytococcus Sedentarius 00340 HMM logo. This provides a quick overview of the features of a profile of HMM while conserving as much information as possible.



glutathione-dependent formaldehyde dehydrogenase [Kytococcus sedentarius]
Sequence ID: WP_012801550.1 Length: 403 Number of Matches: 1
▶ See 1 more title(s)

Range 1: 1 to 403 GenPept Graphics

Score	Expect	Method	Identities	Positives	Gaps
823 bits(2127)	0.0	Compositional matrix adjust.	403/403(100%)	403/403(100%)	0/403(0%)

```
Query  1    MRALTWQGNSDVRVEDVPDPTIQQPTDAIVRVTSTAICGSDLHLYSVLGAFLHPGDVLGH  60
            MRALTWQGNSDVRVEDVPDPTIQQPTDAIVRVTSTAICGSDLHLYSVLGAFLHPGDVLGH
Sbjct  1    MRALTWQGNSDVRVEDVPDPTIQQPTDAIVRVTSTAICGSDLHLYSVLGAFLHPGDVLGH  60
```

Figure 3. - Kytococcus Sedentarius 00340 BLAST results top hit. BLAST is used to rapidly identifyamino acid sequences that are related to a query sequence submitted by an investigator. Here, the Ksed_00340 results show the top hit in the BLAST search. The results show a 100% match with a score of 823 bits.

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.

Locus Tag	Predicted Function	Change
Ksed_00320	hypothetical protein	No change
Ksed_00330:	short-chain Dehydrogenase	No change
Ksed_00340:	glutathione-dependent formaldehyde dehydrogenase	No change

References

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

Acknowledgments

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

www.buffalo.edu