

Annotation of the *Bifidobacterium longum* NCC2705 Genome at Locus Tags BL0045 and BL0047 to BL0049

Erin Szuromi, Diya Tanwar, Daniel Villani, Daniel Wishengrad, and Peter Hentschke
The Harley School, Rochester, NY and The Western New York Genetics in Research in Health Care Partnership



SEPA
SCIENCE EDUCATION
PARTNERSHIP AWARD
Supported by the National Institutes of Health

UB
University
at Buffalo

Abstract

A group of four genes found in the bacterium *Bifidobacterium longum* (from BL0045-BL0049, excluding BL0046) 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, and enzymatic function. For the most part, the Genbank proposed gene product name did not differ significantly from the proposed gene annotation for each of the four genes in the group and as such, the genes appear to be correctly annotated by the computer database.

Introduction

Bifidobacterium longum is a Gram-positive, catalase-negative, and rod-shaped bacterium found in the human gastrointestinal tract. It is a microaerotolerant anaerobe, and it is one of the earliest colonizers of the gastrointestinal tract of infants. It represents about 3% of the bacteria in an adult's gastrointestinal tract, but it represents about 90% of an infant's gastrointestinal tract. It is a part of the gut flora, and it produces lactic acid which is said to prevent growth of pathogenic organisms.

Bifidobacterium longum is also used as a probiotic. It is in many supplements and dairy products to help promote better health levels and happier moods. This probiotic doesn't form spores, it cannot travel freely, and it thrives best in environments with a lack of oxygen. *Bifidobacterium longum* also gives anti-allergy effects, reduces the amount of harmful bacteria, and it improves the intestinal environment.

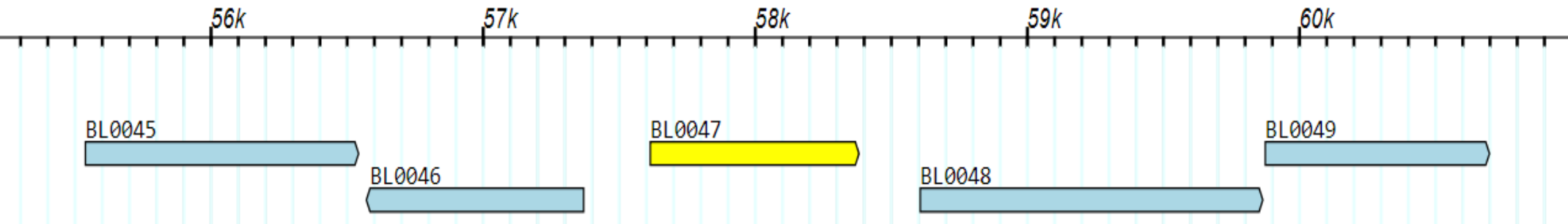


Figure 1. The locus tags and relative position of the genes under investigation in this research

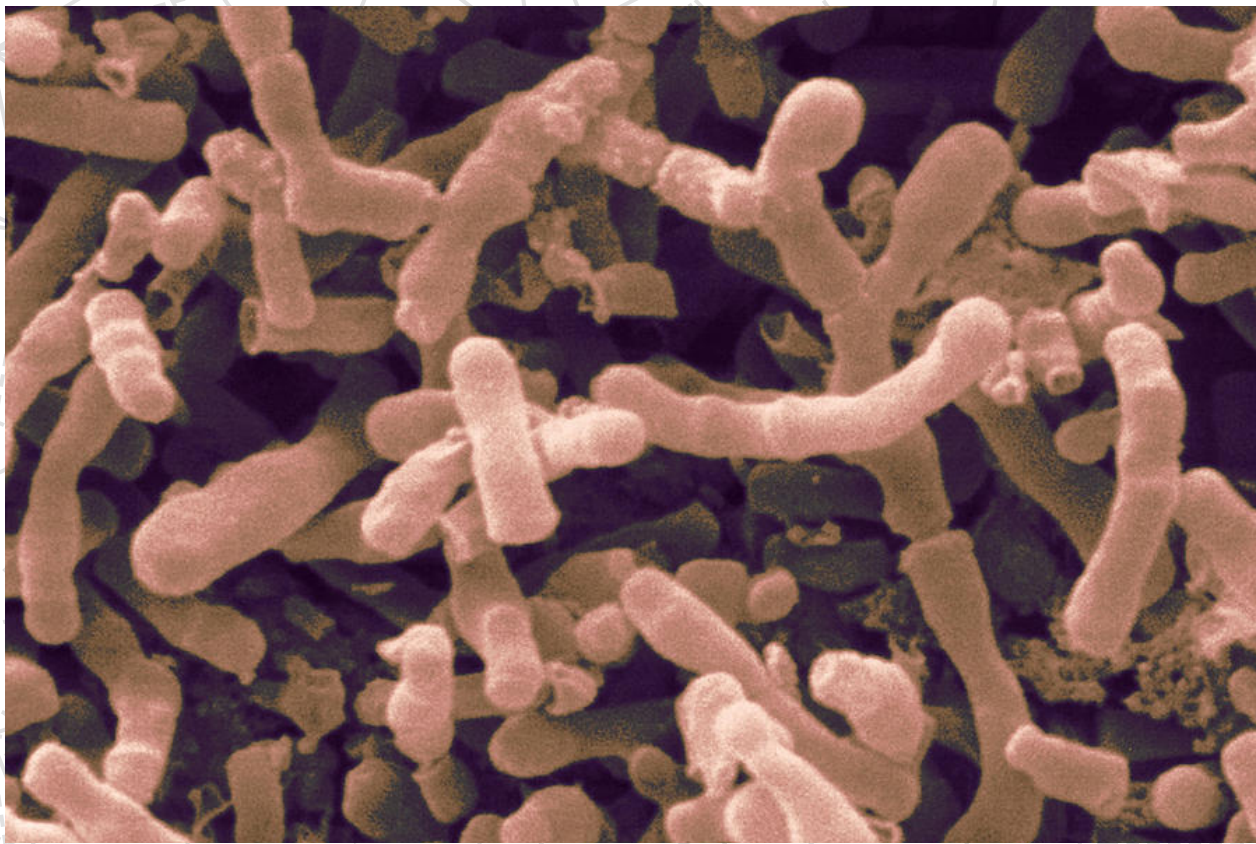


Figure 2. This is an SEM of a colony of *Bifidobacterium longum*.

Methods

Modules of the GENI-ACT (<http://www.geni-act.org/>) were used to complete *Bifidobacterium longum* NCC2705 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?

Results

BL0045:
The gene BL0045 is located from DNA coordinates 55688 to 56692. BL0045 is located on the top strand in the positive direction. It has 1005 base pairs in its nucleotide sequence. This gene was predicted to be a putative glycosyltransferase. The program WebLogo constructs the multiple alignment made previously in Module 2 in a way that is easier to read. (See Fig. 3). Each section of the multiple alignment will be reduced to the most common amino acid occurring. Each large letter represent a very common amino acid that occurs in the multiple alignment of the similar proteins. The different colors in the WebLogo represent different types of amino acids. Polar amino acids are represented by green, basic are blue, acidic are red, and hydrophobic are black. Given there are single tall letters in many positions, this protein is highly conserved. There is a noticeably large amount of hydrophobic molecules located in the last three lines at the C terminal end.

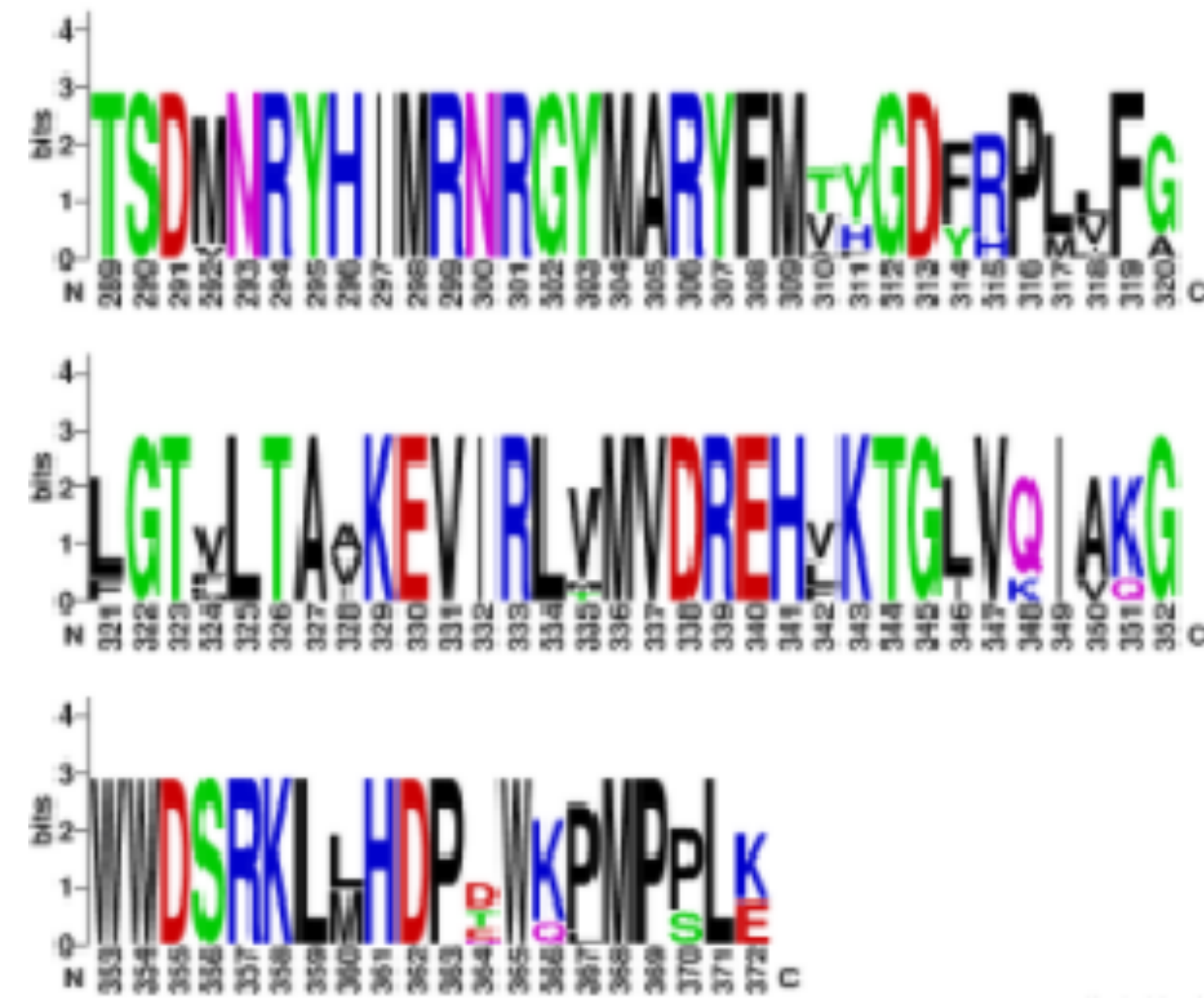


Figure 3. The C-terminal end of the Weblogo of gene BL0045.

BL0047:
This gene is located on the DNA coordinates from 57764 to 58531. It has 768 base pairs in its nucleotide sequence. The predicted product of BL0047 was polyphosphate glucokinase. The results from PSORT-B predicted that the protein was in the cytoplasm. This is because the cytoplasmic score was the highest, at 9.67, while the cytoplasmic membrane score was .01. Also, there were no transmembrane helices, so that also suggests that the protein was not in the membrane. Also, the protein isn't meant to be secreted because there was no sign of a signal peptide. The results from the E.C Number show that the E.C name is polyphosphate-glucose phosphotransferase. When analyzing the alternate open reading frame of the gene, there was a Shine-Dalgarno sequence 7 base pairs upstream of the potential start codon (Fig. 4). This suggests that the software predicted the start codon correctly, and the reading frame was correct. Also, after looking at the DNA coordinates of the potential start and stop codons, it was concluded that the given coordinates were the same as the coordinates on the reading frame.

BL0048:
This gene is located from DNA coordinates 58756 to 60015 on the leading strand of the organism's DNA in the positive direction. The nucleotide sequence is 1260 base pairs in length. This gene's proposed product was identified as Aminotransferase through GENI-ACT, and was supported by both BLAST and PFAM searches. Aminotransferase is a cytoplasmic enzyme used in the synthesis of amino acids. This cellular location was supported by the PSORTb result, and also supported by the lack of transmembrane helices and signal peptides. In PFAM, this gene product had a PFAM name of Aminotransferase class I and II in the clan of PLP Aminotran. It had a PFAM score of 96.3 and an E-value of 1.9e-27. The pairwise alignment and the HMM logo contained key structural residues including G45, N159, P160, D187, Y190, G162, R229, G227, G231, and R345. Part of the HMM logo of this gene is shown in Figure 5.

BL0049:
The gene BL0049 is located from DNA coordinates 60024 to 60848. The gene is found on the top strand in the positive direction. The gene is 825 nucleotides long. The evidence from the tools in Module 5 contradict each other. TMHMM says that there are no predicted transmembrane helices, however PSORTb predicts it is part of the cytoplasmic membrane and Phobius corroborates this information. This is odd because when there are no predicted transmembrane proteins, the protein is not predicted to be in the membrane. A protein found in the membrane is likely to have several transmembrane helices. Nevertheless, I conclude that the protein is part of the cytoplasmic membrane because of the results from the latter two tools. The predicted crystal structure of the protein is shown in Figure 6

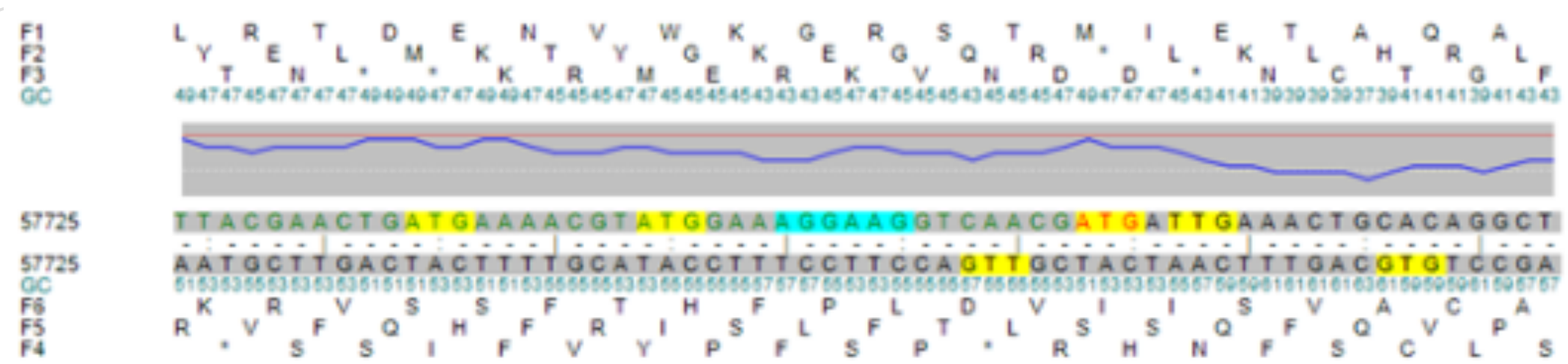


Figure 4. Alternate Open Reading Frame for BL0047.

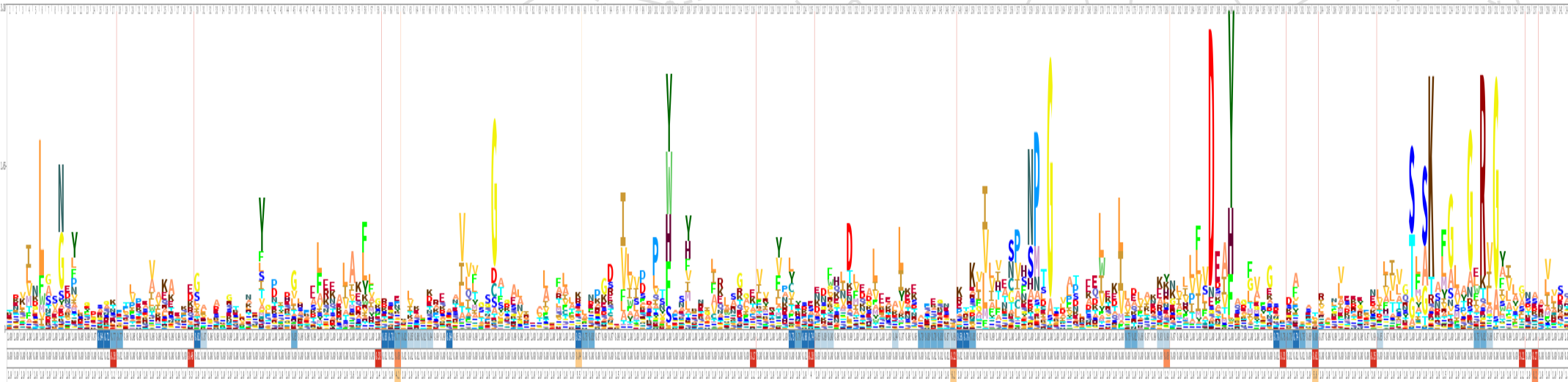


Figure 5. Part of the HMM logo of the gene BL0048.

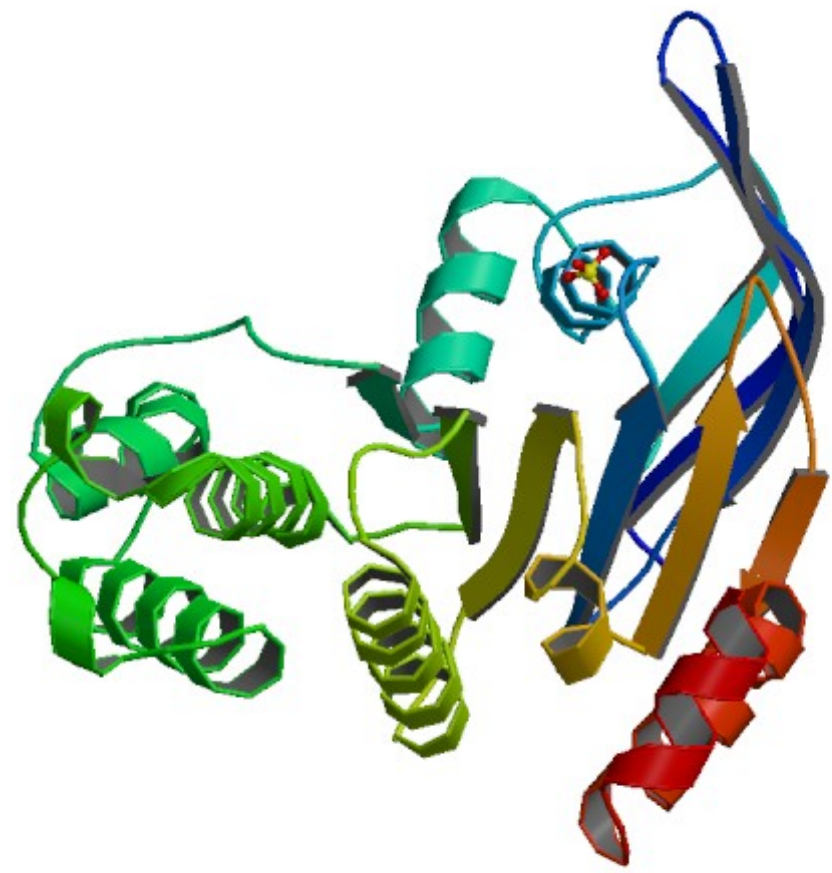


Figure 6. Proposed crystal structure of gene BL0049.

Conclusion

For the most part, 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.

Gene Locus (Locus Tag)	Geni-Act Gene Products	Proposed Annotation
55688 (BL0045)	Putative Glycosyltransferase	Glycosyltransferase family 2
57764 (BL0047)	Polyphosphate glucokinase	Polyphosphate–glucose phosphotransferase
58756 (BL0048)	Aminotransferase	Aminotransferase
60024 (BL0049)	ABC transporter ATP-binding protein	ABC transporter ATP-binding protein

References

Sugahara, H. et al. Probiotic *Bifidobacterium longum* alters gut luminal metabolism through modification of the gut microbial community. *Sci. Rep.* 5, 13548; doi: 10.1038/ srep13548 (2015).

Acknowledgments

Supported by an NIH Science Education Partnership (SEPA) Award - R25OD010536
A special thanks to Dr. Stephen Koury for all his help.
www.buffalo.edu