

# Annotation of the *Bifidobacterium longum* NCC2705 Genome at Locus Tags BL0031 and BL0033 to BL0035

Katharine Briggs, Milo Coen de Peck, Eli Curwin, Zachary Ellis, and Peter Hentschke

The Harley School, Rochester, NY and The Western New York Genetics in Research in Health Care Partnership



## Abstract

A group of four genes from *Bifidobacterium longum* NCC2705 (BL0031, BL0033, BL0034, and BL0035) were annotated using the collaborative genome annotation website GENI-ACT. The Genbank proposed gene product name for each 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. The results from this work did not contradict the Genbank proposed gene products in this group of genes.

## Introduction

*Bifidobacterium* species are obligate anaerobe, Gram-positive bacteria that make up a significant portion of the microflora in the infant and adult human gastrointestinal tract. This genus is one of the first to colonize a newborn babies gastrointestinal tract. *Bifidobacterium longum* is generally thought to be the most common species inhabiting the human intestine. These bacteria are known as “probiotic” because they provide healthy benefits to their hosts, such as providing water-soluble vitamins, preventing diarrhea, and reducing lactose intolerance and suppressing inflammatory responses. (Chaplin et. al, 2015)



Figure 1. An SEM image of *Bifidobacterium longum*.

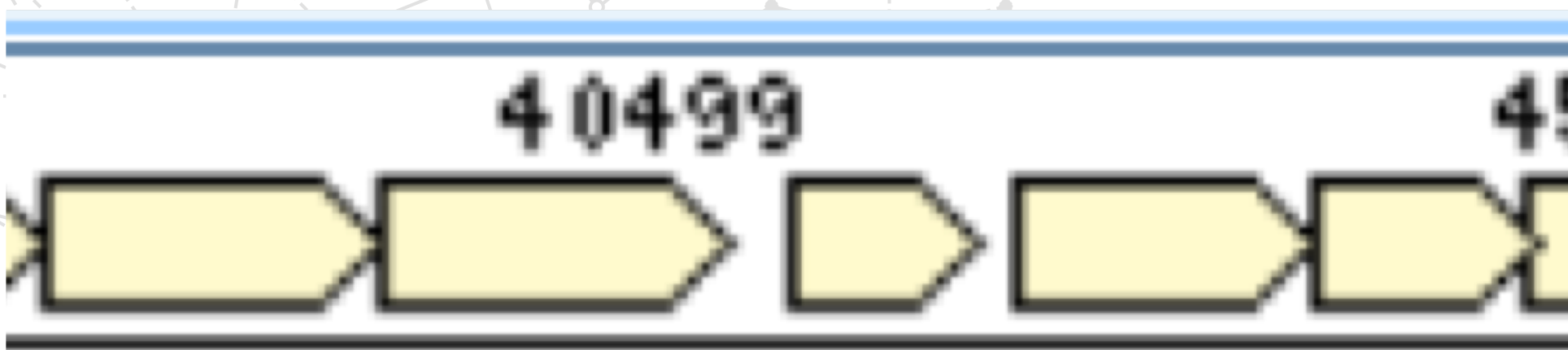


Figure 2. The locus neighborhood and relative position of the genes under investigation in this research

## Methods

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

### BL0031

According to GENI-ACT, this gene’s product name is sensor histidine kinase. The gene product proposal was later supported by a BLAST search excluding *Bifidobacterium longum*. In this search, SwissProt, a curated database, was used. It confirmed the gene product to be sensor histidine kinase.

### BL0033

The proposed product of this gene by GENI-ACT was a sugar ABC transporter substrate-binding protein. This gene product proposal was supported by the top BLAST hits for the amino acid sequence. The cellular location of this gene could not be located by Phobius, and PSORT-B. The Phobius graph given for my gene did not give enough information on the location of the gene, and there was only one predicted transmembrane helix. This helix did not even make a probability of 7.5. For PSORT-B, all of the scores were exactly the same which means the location of the gene product could not be located within the cell. Figure 6 shows a protein structure model from PDB for this gene product.

### BL0034

The initial proposed product of this gene by GENI-ACT was an Uncharacterized ABC transporter ATP-binding protein YtfR. This gene product proposal was supported by the top BLAST hits for the amino acid sequence, the cellular location of the amino acid sequence, and the enzymatic function of the amino acid sequence. This WebLogo graph from Locus Tag BL0034 (Fig. 4) shows the C terminal end of the protein. The first line is very well conserved with polar (green), acidic (red), and hydrophobic (black) amino acids. The second line is much less conserved and is made up of mostly polar amino acids. The third line is somewhat conserved and is made up of a majority of polar amino acids.

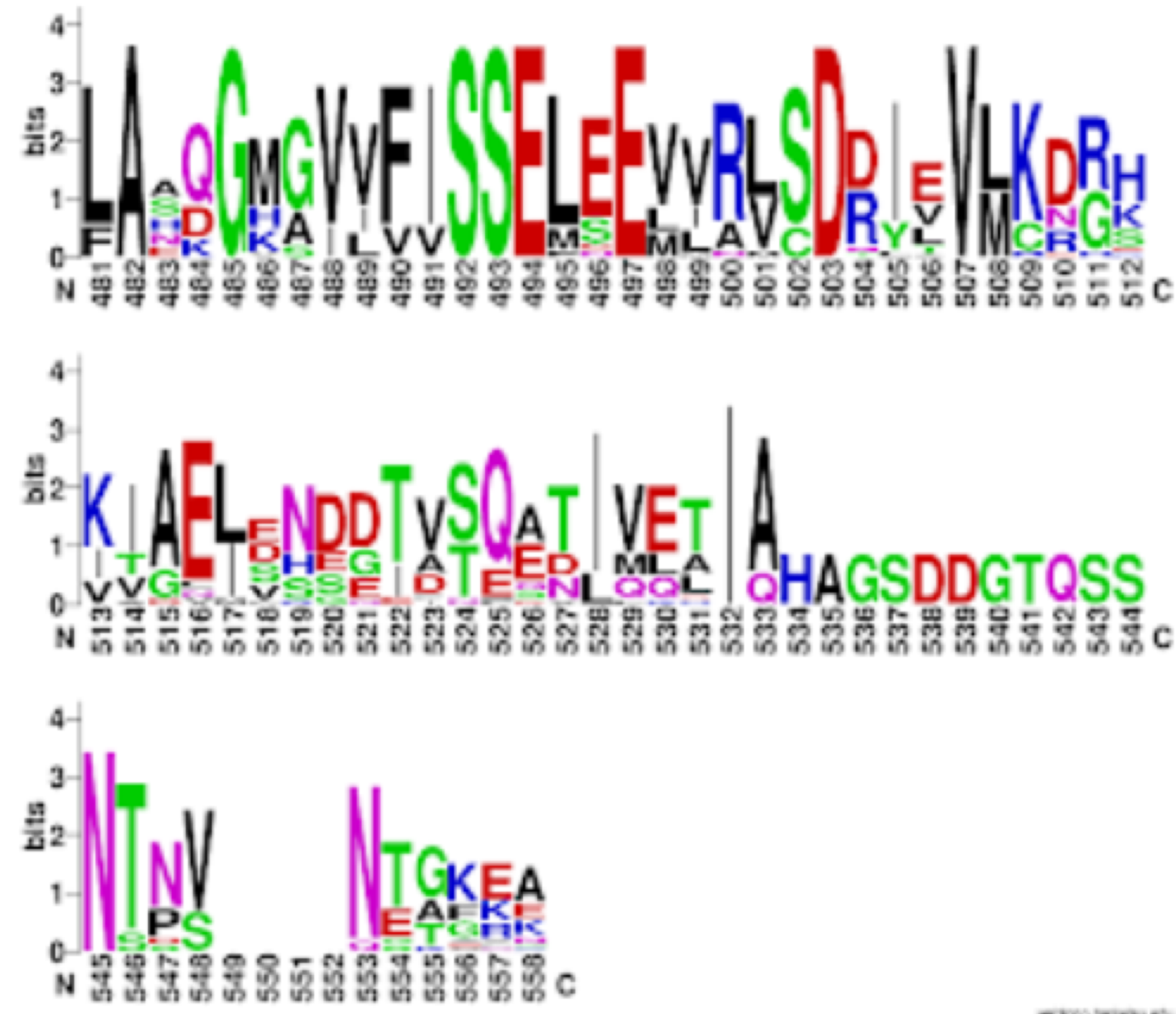


Figure 4. WebLogo of BL0034.

### BL0035

The product of the locus tag BL0035 was most likely called by the computer correctly: ABC transporter permease. The TMHMM test yielded a result of seven predicted TMHs or transmembrane helices (see Fig. 5). However, the P-SPORTB gave the result of a score of 10.00 for the cytoplasmic membrane. This means it is probably located in the cytoplasmic membrane. Helping confirm the hypothesis that the protein stays in the cytoplasmic membrane, SignalP predicted no likely cleavage site. All of this information concludes that the protein probably resides in the cytoplasmic membrane due to its score of 10 and also because there are seven TMHs but a lack of a cleavage site. This all shows that the computer was most likely correct about the product of the locus tag BL0035 being ABC transporter permease because the protein binds ATP inside of the cytoplasmic membrane and are part of a cassette of ATP-binding proteins.

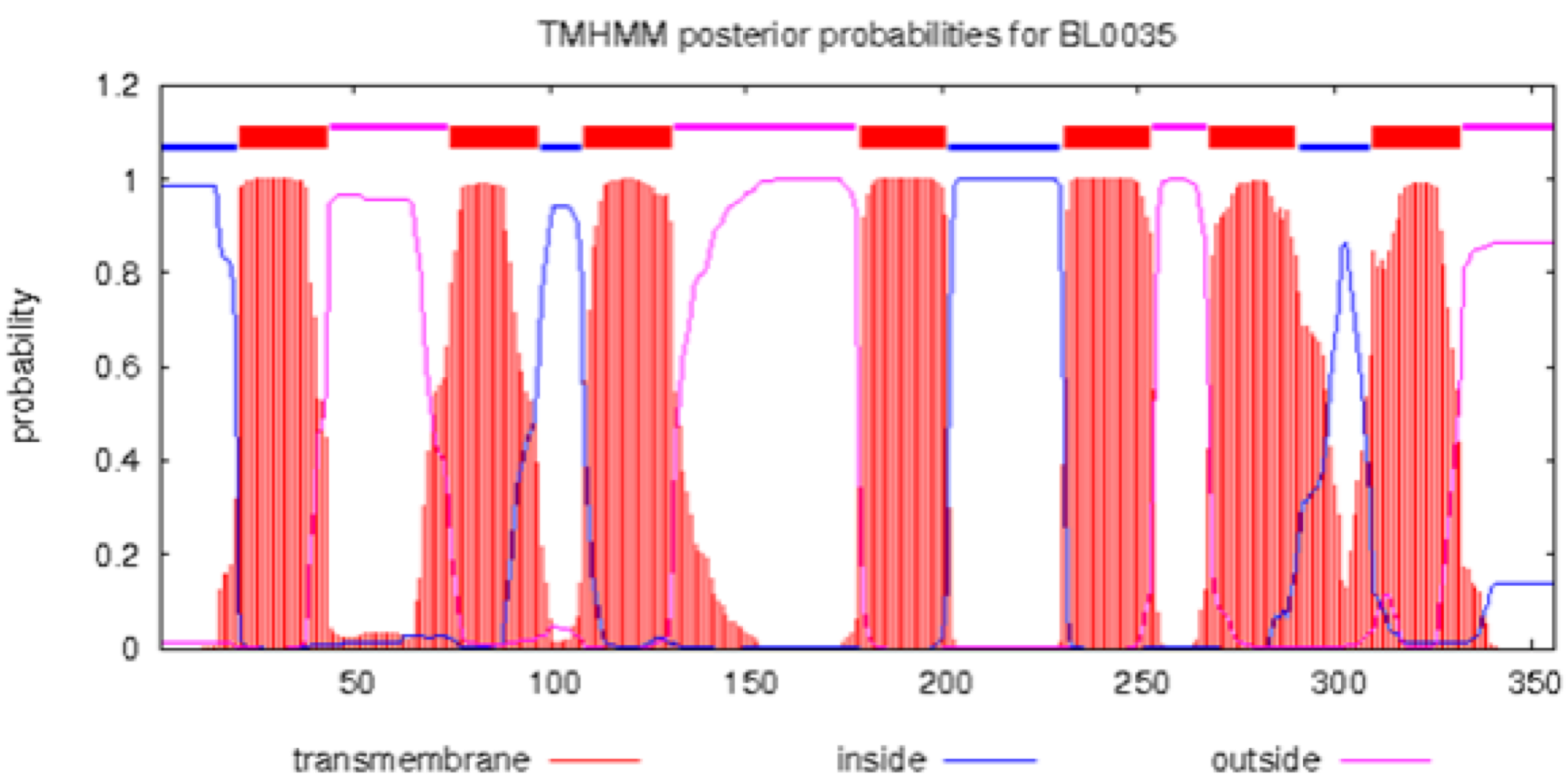


Figure 5. TMHMM for BL0035



Figure 6. The protein structure of the gene product of BL0033: galacto-N-biose-/lacto-N-biose I-binding protein (GL-BP) of the ABC transporter from *Bifidobacterium longum* in complex with lacto-N-biose.

## 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.

Gene Locus (Locus Tag)	Geni-Act Gene Products	Proposed Annotation
38212 (BL0031)	Sensor histidine kinase	Sensor histidine kinase
41952 (BL0033)	sugar ABC transporter substrate-binding protein	sugar ABC transporter substrate-binding protein
43076 (BL0034)	ABC transporter ATP-Binding protein YtfR	ABC transporter ATP-Binding protein YtfR
44619 (BL0035)	ABC transporter permease	ABC transporter permease

## References

Chaplin, Andrei V. et al. “Intraspecies Genomic Diversity and Long-Term Persistence of *Bifidobacterium Longum*.” Ed. Christian U. Riedel. *PLoS ONE*10.8 (2015): e0135658. *PMC*. Web. 16 May 2018.

Schell, Mark A. et al. “The Genome Sequence of *Bifidobacterium Longum*Reflects Its Adaptation to the Human Gastrointestinal Tract.” *Proceedings of the National Academy of Sciences of the United States of America* 99.22 (2002): 14422–14427. *PMC*. Web. 16 May 2018.

## Acknowledgments

Supported by an NIH Science Education Partnership (SEPA) Award - R25OD010536  
[www.buffalo.edu](http://www.buffalo.edu)