Annotation of the *Bifidobacterium longum* Genome at Locus Tags BL0036 to BL0038, BL0041, and BL0043)

Emmy Kaza, Ella Novy-Marx, Henry Michael, Micah Smith, Hannah Snyder 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

University at Buffalo

Abstract

A group of 5 genes from the microorganism *Bifidobacterium longum* (BL0036, BL0037, BL0038, BL0041, and BL0043) 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. We did not find any results that contradicted the gene product names predicted by Genbank.

Introduction

Bifidobacterium longum develops in gut-bacteria and has a commonly known ability to digest and consume oligosaccharide structures. Gene products from this genome function as glycosidase and oligosaccharide transporters, anti-inflammatories for intestinal cells, and control the amount of Enterobacteriaceae in feces. It additionally serves as a manipulator with human milk glycans and intestinal microbiota.

Bifid microbes have been discovered in stools of healthy infants proving that human milk contains types of bifidogenic factors that grow *Bifidobacterium*. These microbes inhabit the intestinal tract and health of the gut serving as protection against feeding, birth, antibiotics, probiotics, and more. *Bifidobacterium longum* is one of the most abundant of the 500 bacteria found in the human digestive tract. Found in the gastrointestinal tract of humans and many other mammals, *Bifidobacterium longum* is a rod shaped bacterium that makes lactic acid from the fermentation of various sugars.

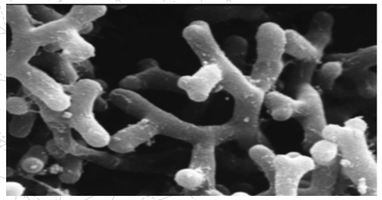

Figure 1. 3D Image of the *Bifidobacterium longum* organism.

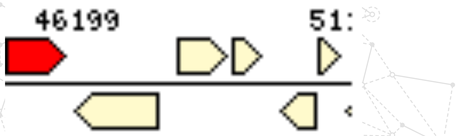

Figure 2. The locus neighborhood and relative position of the genes under investigation in this research

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

BL0036: GENI-Act proposed that the gene product for BL0036 was an ABC transporter permease, an integral transport permease located in the cellular membrane. The TMHMM search showed seven predicted transmembrane helices. (See Fig. 3). This indicates that this protein may be located in the membrane. SignalP predicted no signal peptides, indicating that the protein is most likely not secreted from the cell. PSORT-B gave a Cytoplasmic Membrane score of 10.0, with all other scores reported as 0. This strongly suggests that the protein is located in the membrane

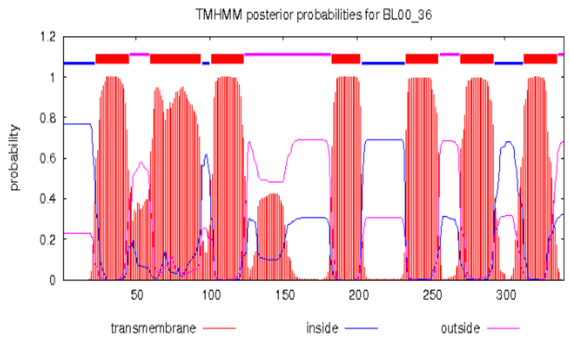

Figure 3. Graph of the probability of transmembrane helices, at certain amino acid locations for BL0036

BL0037: The WebLogo for BL0037 shows the amino acid alignment of my protein with similar proteins from other organisms. (Fig. 4.) Each letter represents an amino acid, and the larger the letter, the more consistently occurring the amino acid is across the proteins. You can tell how well conserved they are because the letters that represent the amino acids are consistently large. The only exception is the first row of alignment slots. You notice that there are many blanks. This is because one of the proteins doesn't have many amino acids in those spaces.

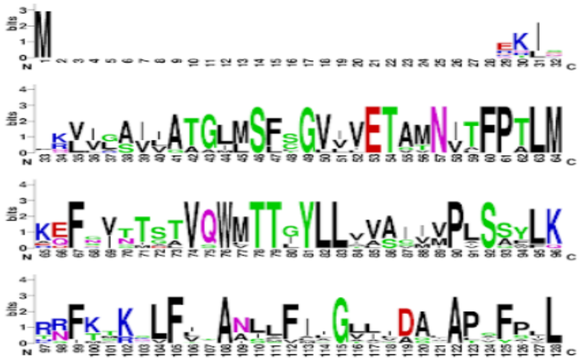

Figure 4. WebLogo for BL0037.

BL0038: The HMM logo for my protein, BL0038, is shown in Figure 5. It shows the key functional/structural residues. For this particular segment of my HMM logo the key functional/structural residues are G40, G41, G43, G45, I47, and G48. These amino acids, demonstrated by wide tall letters, indicate major amino acids.

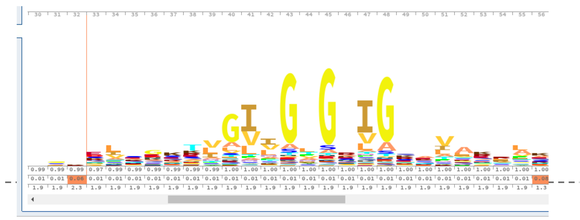

Figure 5. HMM Logo with key functional/structural residues of protein BL0038

BL0041: BL0041 has a very short query sequence, with only 121 amino acids. There were 4 helices predicted in the program TMHMM. In PSORT-B, the only score that concluded with results was the cytoplasmic score of 10.00. For BL0041, the pairwise alignment did not match up with any clan. (Fig. 6) The score of this pairwise alignment was 64, and an E-value of 1.6e-17. The Pfam number for the top his was 04020. The Pfam name for this hit was "Phage_holin_4_2".

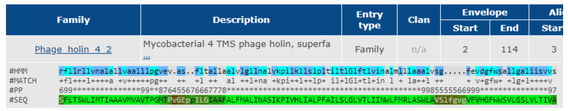

Figure 6. The Pfam pairwise alignment for BL0041.

BL0043: BL0043 was analyzed and no Shine-Delgarno sequence was found upstream of the called start codon, but an alternative start codon was found downstream with a Shine-Delgarno sequence 5 nucleotides upstream. (See Fig.7)

A BLAST search was performed to determine if the alternative start codon would be a better match, and the original called start codon appeared to be a better match. This was determined under the circumstances that the original start codon had a higher score and lower e-value compared to the alternative start codon. The BLAST search resulted in two top hits; one with a gene product of an ABC Transporter from the *Bifidobacterium* sp.12 1 47BFAA organism. The second hit gene product name was Cobalt ABC Transporter from the *Bifidobacterium breve* organism.

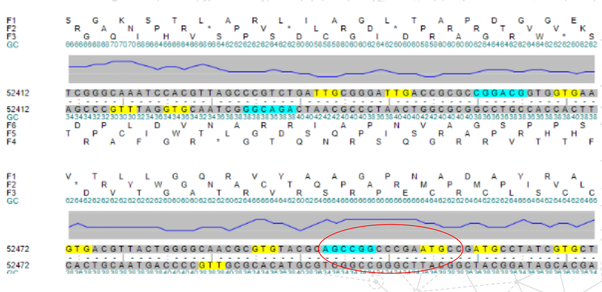

Figure 7. Alternative Open-Reading Frame results for BL0043 depicting alternative start codon.

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
45686 (BL0036)	Sugar ABC transporter permease	Sugar ABC transporter permease
46931 (BL0037)	Lincomcyn resistance protein LmrB	Lincomcyn resistance protein LmrB
48777 (BL0038)	D-isomer specific 2-hydroxyacid dehydrogenase	D-isomer specific 2-hydroxyacid dehydrogenase
51370 (BL0041)	MULTISPECIES: membrane protein	MULTISPECIES: membrane protein
52211 (BL0043)	ABC Transporter	ABC Transporter

References

Underwood, Mark A. et al. "*Bifidobacterium Longum* Subspecies *infantis*: Champion Colonizer of the Infant Gut." *Pediatric research* 77.0 (2015): 229–235. *PMC*. Web. 16 May 2018.

Acknowledgments

Supported by an NIH Science Education Partnership (SEPA) Award - R25ODO10536

www.buffalo.edu