University at Buffalo *The State University of New York* | REACHING OTHERS

SEPA SCIENCE EDUCATION PARTNERSHIP AWARD

NIH National Institutes of Health *Turning Discovery Into Health*

Annotation of the *Bifidobacterium longum* Genome from DNA Coordinates 73478 to 882086

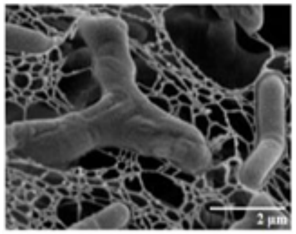
Paul Gaston, Ashley Reed, Ali Robeson, Reilly Stewart, Emily Wunder, Dennis Youngs and Brenda Travis

Penn Yan Central School District and the Western New York Genetics in Research Partnership

Abstract

Our team consisted of four annotators who worked on four sections of *Bifidobacterium longum,*'s genome, a gram positive bacillus, to discover possible gene functions. Four GeniAct modules were used to obtain the information summarized in the results and in the conclusion. Sections of this bacteria's genome were hand annotated to confirm that the computer correctly identified the amino acid sequences and the sequence products. The four proteins are penicillin binding protein which are essential for bacterial cell wall synthesis, the FtsW RodA SpoVE protein determines rod shape and functions in cell division, phosphatase a regulatory enzyme, dipeptidyl peptidase a membrane protease. Our team also included two researchers who were instrumental in understanding protein function and in developing the poster design.

Introduction



microbenet.cdc.gov

Bifidobacterium longum is a Gram positive bacterium, one of 32 species in the genus *Bifidobacerium*[1]. This is a beneficial bacteria that aids in human digestion. It is found in infant guts but represents only 3-6% of adult gut bacteria. It is a rod-shaped anaerobe and is non-pathogenic.

The whole genome of Bifidobacerium Longum is 2,256,646 base pairs, and is predicted to code for 1,730 protein[1]. Both Penn Yan posters together show annonotations for 8 coding sequences this represents about .5% of the total sequence in bifidobacterium.

The genes that are reported on this poster BLLJ0060 to BLLJ 0062, and BLLJ 0065 had DNA coordinates stated in the title, containing 7,167 base pairs. The section codes for 2385 amino acids.

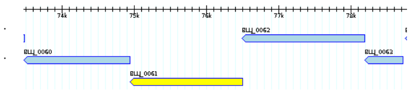


Figure: Locus tags and coordinates for Bifidobacterium longum genes that were manually annotated.

Methods and Materials

Modules 1-4 of the GENI-ACT (http://www.geni-act.org/) were used to complete *Bifidobacterium longum* genome annotation. Those 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- Structure-Based Evidence	TIGRfam, Pfam, PDB	Are there functional domains in my protein?
Module 4- Cellular Localization Data	Gram Stain, TMHMM, SignalP, PSORT, Phobius	Is my protein in the cytoplasm, secreted or embedded in the membrane?

Results

Bifidobacterium longum 0060:

The initial proposed product of these genes by GENI-ACT was penicillin-binding protein. This gene product was well conserved as supported by the WebLogo. There is a good probability that the protein contains a signal peptide. The predicted D score of 0.571 (Fig.I), which is above the 0.45 probability cutoff, supports the signal peptide.

Bifidobacterium longum 0061:

The initial proposed product of these genes by GENI-ACT was Uncharacterized FtsW (cell division)-like protein. This gene product proposal was supported by the top BLAST hit from SwissProt database for the amino acid sequence. This is further supported by a well conserved WebLogo. TMHMM (Fig II) clearly indicates 12 transmembrane helices containing hydrophobic amino acids. PSORT-B returns a score of 10.0 for cytoplasmic membrane.

Bifidobacterium longum 0062:

The initial proposed product of this gene by GENI-ACT was predicted phosphoprotein phosphatase, top BLAST hit from NR. It was supported based on the WebLogo (Fig.III) as being well conserved. The final PSORT prediction is that phosphoprotein phosphatase has 2 transmembrane helixes.

Bifidobacterium longum 0065:

This gene product was a predicted peptidase S9, supported by the top hit of BLAST. Further evidence indicated a dipeptidyl aminopeptidase /acylaminoacyl peptidase including a well-curated functional domain (Fig.IV) within the amino acid sequence which was shown in the results of the T-Coffee and WebLogo tests. After an examination of the cellular localization of this gene, it became evident that this gene was located in the cytoplasmic membrane and function is similar to the superfamily of peptidases.

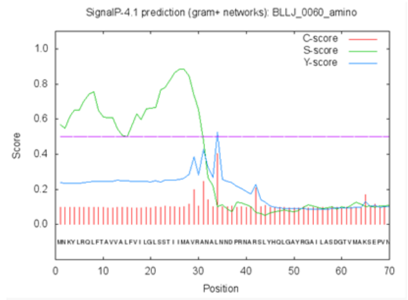


Figure I – *Bifidobacterium longum* 0060: SignalP results predicting signal peptide with D score of 0.571.

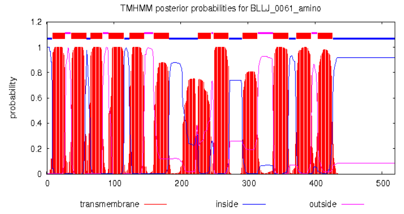


Figure II – *Bifidobacterium longum* 0061: TMHMM results showing 12 transmembrane helices

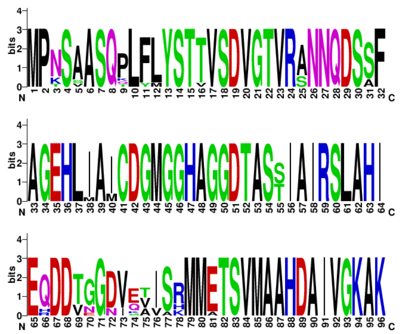
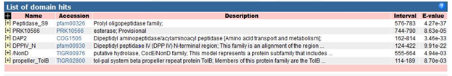


Figure III – *Bifidobacterium longum* 0062: WebLogo (portion) results show a well conserved protein. Gaps occur from 519 - 531 and 555 -



List of domain hits

Name	Accession	Description	Interval	E-value
[+] Peptidase_S9	pfam00326	Prolyl oligopeptidase family;	576-783	4.27e-37
[+] PRK10566	PRK10566	esterase; Provisional	744-790	8.63e-05
[+] DAP2	COG1506	Dipeptidyl aminopeptidase/acylaminoacyl peptidase [Amino acid transport and metabolism];	162-814	3.46e-33
[+] DPPIV_N	pfam00930	Dipeptidyl peptidase IV (DPP IV) N-terminal region; This family is an alignment of the region ...	124-422	9.91e-22
[+] /NonD	TIGR00976	putative hydrolase, CocE/NonD family; This model represents a protein subfamily that includes ...	555-664	4.94e-03
[+] propeller_TolB	TIGR02800	tol-pal system beta propeller repeat protein TolB; Members of this protein family are the TolB ...	114-189	8.70e-03

Figure IV – *Bifidobacterium longum 0065* Conserved Domain results showing a COG name of Dipeptidyl aminopeptidase/acylaminoacyl peptidase and an E-value of 3.46e-33

Conclusion

The GENI-ACT proposed gene product did not differ significantly from the proposed gene annotation for each of the genes in the four sectors and as such, the genes appear to be correctly annotated by the computer database. The most interesting gene product out of the 8 that were manually annotated by our school was BLLJ_0060 which was the only signal peptide of them all.

Gene Locus	Geni-Act Gene Products	Proposed Annotation
0060	Penicilin binding protein	Penicilin binding protein
0061	FtsW RodA SpoVE	FtsW RodA SpoVE
0062	Phosphotase 2C	Phosphotase 2C
0063	Dipeptidyl peptidase	Dipeptidyl peptidase

References

1. Schell et al., 2002. The genome sequence of *Bifidobacterium longum* reflects its adaptation to the human gastrointestinal tract. *PNAS* 99(22): 14422-14427.

2. Alekseeva et al., 2015. Isolation and purification of recombinant serine/threonine protein kinases of the strain *Bifidobacterium longum* B379M and investigation of their activity. *Biochemistry (Moscow)* 80(10): 1303-1311.

3. Ventura et al., 2007. Genomics as a means to understand bacterial phylogeny and ecological adaptation: the case of bifidobacteria. *Antonie van Leuwenhoek* 91(4): 351-372.

4. EMBL-EBI, InterPro: protein sequence analysis and classification

Acknowledgments

Supported by NIH SEPA Award R25ODO10536

We would like to thank for assistance with teaching and research Dr. Stephen Koury University at Buffalo. Ms. Irka Elsevier: volunteer for her untiring support. Mr. David Pullen: Principal Penn Yan Academy