Annotation of the *Bifidobacterium longum* Genome from DNA Coordinates 68401 to 73481

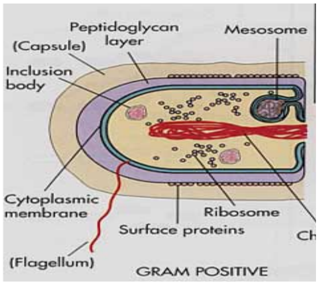
Jennalyn Conley, Lin Dickinson, Donald Griner, Erica Jensen, Ashley Owens, Victoria Wilson and Brenda Travis

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

Abstract

Four sections of *Bifidobacterium longum,*'s genome, a gram positive bacillus, were annotated 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 sortase, glutamine amidotransferase which speeds up the removal of the ammonia group from a glutamine and then transfers it[4], and two serine/threonine protein kinases which move a phosphate group from the amino acids serine or threonine[2]. Our team consisted of the four annotators and also two researchers who were instrumental in understanding protein function and in developing the poster design.

Introduction



microdigitalproteus.com

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 Lonum is 2,256,646 base pairs, and is predicted to code for 1,730 proteins[1]. Both Penn Yan posters together show anonotations for 8 coding sequences. This represents about .5% of the total sequence in bifidobacterium.

The genes that are reported on this poster BLLJ0056 to BLLJ 0059 had DNA coordinates stated in the title, containing 4,800 base pairs. The section codes for 1, 596 amino acids.

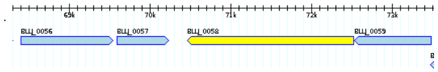


Figure : Locus tags and coordinates for Bifidobacterium longum genes that were manually annotated.

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 0056:
The initial proposed product by GENI-ACT was class E Sortase and is located in the cell wall. This is supported by the top BLAST hit from the NR database for the amino acid sequence. All modules completed support this product as a member of the Sortase superfamily. Sortase is a transpeptidase protein that attaches surface proteins to the cell wall[2]. TMHMM clearly indicates 3 transmembrane helices. PSORT-B returns a score of 10.0 for cytoplasmic membrane supporting that this protein will be found in the cell membrane.

Bifidobacterium longum 0057:
Glutamine amidotransferase was supported by the top blast hit for the amino acid sequence which is supported by a well conserved T-Coffee alignment (Fig.II). Glutamine Amidotransferase is located in the cytoplasm ,which is supported by PSORTb which predicts the cytoplasmic score to be 9.97.

Bifidobacterium longum 0058:
The initial proposed product of these genes by GENI-ACT was serine/threonine protein kinase. The gene product proposal was supported by the top BLAST hit in NR exclusion of bifidobacterium. The E-value is zero. This suggests that the sequence is highly conserved in a number of different organisms. PSORTb predicts that this protein is found in the cytoplasm.

Bifidobacterium longum 0059 The initial proposed product was a predicted serine/threonine protein kinase, supported by the top hit of BLAST which analyzes the amino acid sequence. Further evidence includes a well-curated functional domain within the amino acid sequence shown in the results of the T-Coffee and HMMLogo (Fig.IV). The similarities between the initial proposed gene in GENI-ACT and various test results indicate a serine/threonine protein kinase PKnA, found in the cell membrane.

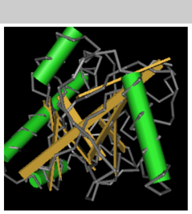


Figure I – *Bifidobacterium* 0056 Sortases are cysteine transpeptidases, that anchor surface proteins to peptidoglycans of the bacterial cell wall envelope.

```
BLLJ_0057                          MTDSAHILVVDNYDSFVYTIVGYLQTLGATVDVVRNDAIDPAAPGVLDGY
gi|524040084|emb|CCY18304.1|       MTDSARILVVDNYDSFVYTIVGYLKTLGATVTVVRNDAIDPSQPGVIDEY
gi|524120660|emb|CCY95018.1|       MTDSAHILVVDNYDSFVYTIVGYLQTLGATVDVVRNDAIDPAAPGVLDGY
gi|644611207|ref|WP_025341493.1|   MTDSARILVVDNYDSFVYTIVGYLKTLGATIDVVRNDAIDPAAPGVLDGY
gi|705393794|ref|WP_033492765.1|   MTDSARILVVDNYDSFVYTIVGYLKTLGATVDVVRNDAIDAAKPGLLDDY
gi|705447856|ref|WP_033517662.1|   MTDSAHILVVDNYDSFVYTIVGYLKTLGATVDVVRNDAIDPAQPQVLDGY
gi|705457893|ref|WP_033522470.1|   MTDSAHILVIDNYDSFVYTIVGYLETLGATVDVVRNDAIDPSDPTVTDGY
gi|727804403|ref|WP_033891444.1|   MTDSAHILVVDNYDSFVYTIVGYLKTLGATVDVVRNDAIDPAAQGVLDGY
gi|736513475|ref|WP_034529464.1|   MTDSAHILVIDNYDSFVYTIVGYLKTLGATVDVVRNDAVDPHDEHVTDGY
gi|759444633|ref|WP_043166639.1|   MTDSAHILVVDNYDSFVYTIVGYLKTLGATVDVVRNDAVDPSDEHVTDGY
gi|763217050|ref|WP_044089514.1|   MTDSARILVVDNYDSFVYTIVGYLKTLGATVDVVRNDAIDPAAPGVLDGY
                                   *****:***:*************:*****: ******:*.    : * *

BLLJ_0057                          DGVLISPGPGAPAESGASEDMIRLCAASGVPMFGVCLGLQALAEVYGCTV
gi|524040084|emb|CCY18304.1|       DGVLISPGPGAPAESGASEDMIRLCAKLDKPMFGVCLGMQALAEVFGATV
gi|524120660|emb|CCY95018.1|       DGVLISPGPGAPAESGASEDMIRLCVASGVPMFGVCLGLQALAEVYGCTV
gi|644611207|ref|WP_025341493.1|   DGVLISPGPGAPAESGASEDMIRLCAATGVPMFGVCLGLQALAEVYGCTV
gi|705393794|ref|WP_033492765.1|   DGVLISPGPGAPAESGASEDVIRRCAAEKKPMFGVCLGLQALAEVFGCTV
gi|705447856|ref|WP_033517662.1|   DGVLISPGPGAPADSGASEDMIRRCASERKPMFGVCLGLQALAEVYGCTV
gi|705457893|ref|WP_033522470.1|   DGVLISPGPGAPAESGASEDMIRLCAEQNKPMFGVCLGLQALAEVYGCTV
gi|727804403|ref|WP_033891444.1|   DGVLISPGPGAPADSGASEDMIRLCAANDVPMFGVCLGLQALAEVYGCTV
gi|736513475|ref|WP_034529464.1|   DGVLISPGPGAPADSGASEDMIRLCAKQGKPMFGVCLGLQALAEVYGATV
gi|759444633|ref|WP_043166639.1|   DGVLISPGPGAPAESGASEDMIRLCARQNKPMFGVCLGLQALAEVYGCTV
gi|763217050|ref|WP_044089514.1|   DGVLISPGPGAPAQSGASEDMIRLCAAAGVPMFGVCLGLQALAEVFGCTV
                                   *************:******:** *.    ********:******:*.**
```

Figure II – *Bifidobacterium* 0057 The results of a T-Coffee multiple sequence alignment. Showing a well conserved protein.

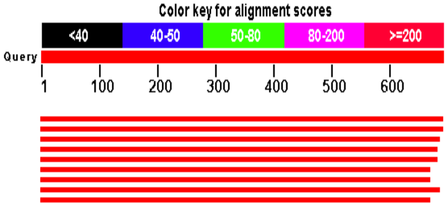


Figure III –*Bifidobacterium* 0058 color key alignment score was close to 100% coverage indicating a high quality alignment.

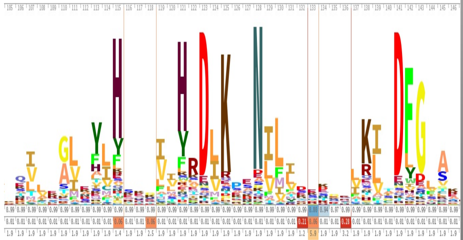


Figure IV –*Bifidobacterium* 0059 HMM LOGO results show that were key functional residues at G8,G10,K30,E48,D123,K125,N128,D141,R252 within the amino acid sequence.

Conclusion

Two of these four proteins BLLJ0056 and BLLJ0059 seem to be located in the membrane while the two others BLLJ 0057 and BLLJ0058 are located in the cytoplasm. Our hand annotations do confirm the computer's identification of the gene products.

Gene Locus	Geni-Act Gene Products	Proposed Annotation
0056	Predicted class E Sortase	Sortase Family
0057	Glutamine amidotransferase	Glutamine amidotransferase
0058	Serine/threonine protein kinase	Kinase domain
0059	Serine/threonine protein kinase	serine/threonine protein kinase PKnA

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 R25OD010536
We would like to thank for assistance with teaching and research Dr. Stephen Koury University at Buffalo's. Ms. Irka Elsevier, volunteer for her untiring support. Mr. David Pullen Principal Penn Yan Academy