Annotation of the *Lactobacillus acidophilus* Genome at Locus Tags LBA0034 And LBA0035)

Melissa Moore, Kioralys Colon, Julia DeLand, and Marie Schuh
Dunkirk High School, Dunkirk, NY and The Western New York Genetics in Research and Health Care Partnership

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

University at Buffalo

Abstract

Two genes from the microorganism *Lactobacillus acidophilus* (LBA0034 - LBA0035) 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, and cellular localization data. The Genbank proposed gene product name did not differ significantly from the proposed gene annotation for each of the genes and as such, the genes appear to be correctly annotated by in the r database. However, the amino acid sequences did differ between GENI-ACT and IMG/M for LBA0035, with GENI-ACT in the end proving to be incorrect for the number of amino acids coded for by the gene. The LBA0034 and LBA0035 genes code for antibiotics and dipeptide metabolism respectively.

Introduction

Gene annotation is important in order to understand a genome. Genome annotation is the process of locating the genes and coding regions as well as determining what the genes actually do. A gene's annotation can be submitted to and stored in genomic databases. Many organisms have had their entire genome sequenced, however this is not all there is to a genome project. Annotation also involves describing different regions of the code and identifying which regions can actually be called genes. Overall, gene annotation is important because it gives us a better understanding of genes and helps us gain knowledge.

Lactobacillus acidophilus is a species of bacteria that can ferment sugars into lactic acid. It grows at pH values below pH 5.0 and grows best around 37C. It naturally occurs in the gastrointestinal tract and mouth of humans and many animals. Some strains of *L. acidophilus* are considered to be probiotic, or healthy, when consumed. These strains are often used in many dairy products such as yogurt. The Mayo Clinic has published a list of disorders for which *L. acidophilus* has been tested to aid with, including vaginal disorders, allergic skin reactions, anemia, and many more. According to the list, "There is good evidence supporting the use of *L. acidophilus* for the treatment of vaginal infections…[And] taking Lactobacillus by mouth along with itraconazole has been shown to help treat scaly, itchy rashes" (Mayo Clinic, 2010). *Lactobacillus* is considered "the most commercially important genus of lactic acid-fermenting bacteria" (Lee, 2009).

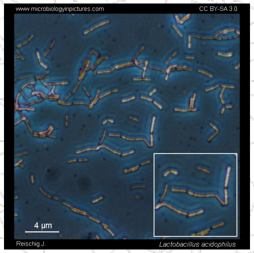


Figure 1 - *Lactobacillus acidophilus*, negative phase contrast under microscope (Microbiology in Pictures website.)

Methods

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

Results

LBA0034:

The initial proposed product of this gene by GENI-ACT was an AAC(3) family N-acetyltransferase. This gene product proposal was supported by the top BLAST hits for the amino acid sequence, however it is questionable based on how poorly conserved the majority of the gene sequence is, especially at the beginning of the multiple sequence alignment WEBLOGO, as can be seen in Figure 2. The region features however include an antibiotic in the CDD, as shown in Figure 4.

LBA0035:

The initial proposed amino acid sequence in GENI-ACT for this gene was a length of 470 amino acids, however when completing a BLAST on this sequence, it returned no matches. When using the IMG/M amino acid sequence, its length was 466 amino acids, which yielded results in a BLAST search. This leads to the potential conclusion that GENI-ACT has an error for the amino acid sequence, and potentially other information. This gene codes for a dipeptidase, an enzyme that breaks down dipeptides.

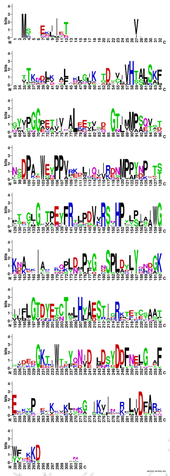


Figure 2 - WEBLOGO for LBA0034. This shows that there is very little conservation at the beginning of the amino acid sequence, which may be a sign that there is an alternative open reading frame (however we did not get to test this theory.) There is no set amino acid chemical pattern or dominant property throughout the sequence.

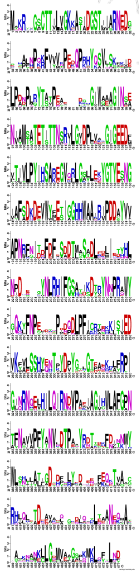
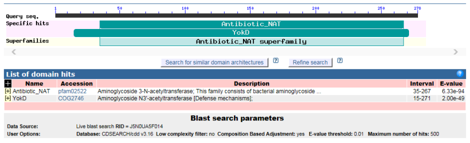


Figure 3 – WEBLOGO for LBA0035. This shows that there is very high conservation and many frequent amino acids in alignment. There are a lot of polar amino acids (green) and very few acidic or basic amino acids. This makes sense when considering the function of the protein that LBA0035 codes for.



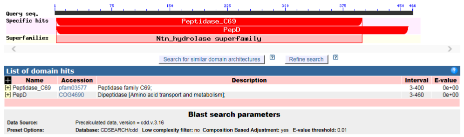
List of domain hits

Name	Accession	Description	Interval	E-value
[+] Antibiotic_NAT	pfam02522	Aminoglycoside 3-N-acetyltransferase; This family consists of bacterial aminoglycoside ...	35-267	6.33e-94
[+] YokD	COG2746	Aminoglycoside N3'-acetyltransferase [Defense mechanisms];	15-271	2.00e-49

Blast search parameters

Data Source: Live blast search RID = J5N0UA5F014
User Options: Database: CDSEARCH/cdd v3.16 Low complexity filter: no Composition Based Adjustment: yes E-value threshold: 0.01 Maximum number of hits: 500

Figure 4 – Domain hits for LBA0034 shows that it codes for antibiotic proteins, as it is part of the Antibiotic-NAT superfamily. This correlates with *L. acidophilus* having potential antibiotic properties.



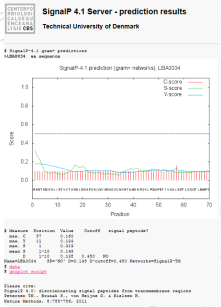
List of domain hits

Name	Accession	Description	Interval	E-value
[+] Peptidase_C69	pfam03577	Peptidase family C69;	3-400	0e+00
[+] PepD	COG4690	Dipeptidase [Amino acid transport and metabolism];	3-460	0e+00

Blast search parameters

Data Source: Precalculated data, version = cdd.v.3.16
Preset Options: Database: CDSEARCH/cdd Low complexity filter: no Composition Based Adjustment: yes E-value threshold: 0.01

Figure 5 – LBA0035 belongs to the Ntn_hydrolase superfamily, which correlates with its protein's proposed metabolic function.

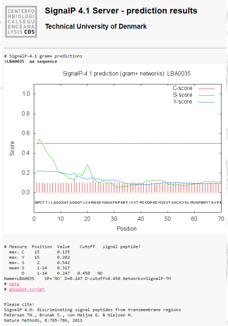


```
# SignalP-4.1 gram+ predictions
>LBA0034  aa sequence

# Measure  Position  Value    Cutoff   signal peptide?
  max. C    57       0.160
  max. Y    11       0.133
  max. S     1       0.319
  mean S     1-10    0.145
       D     1-10    0.138   0.450   NO
Name=LBA0034    SP='NO' D=0.138 D-cutoff=0.450 Networks=SignalP-TM
# data
# gnuplot script

Please cite:
SignalP 4.0: discriminating signal peptides from transmembrane regions
Petersen TN., Brunak S., von Heijne G. & Nielsen H.
Nature Methods, 8:785-786, 2011
```

Figure 6 – The SignalP shows no signal peptides for LBA0034.



```
# SignalP-4.1 gram+ predictions
>LBA0035  aa sequence

# Measure  Position  Value    Cutoff   signal peptide?
  max. C    15       0.135
  max. Y    15       0.202
  max. S     2       0.542
  mean S     1-14    0.317
       D     1-14    0.247   0.450   NO
Name=LBA0035    SP='NO' D=0.247 D-cutoff=0.450 Networks=SignalP-TM
# data
# gnuplot script

Please cite:
SignalP 4.0: discriminating signal peptides from transmembrane regions
Petersen TN., Brunak S., von Heijne G. & Nielsen H.
Nature Methods, 8:785-786, 2011
```

Figure 7 – The SignalP shows no signal peptides for LBA0035, however it has a higher S-score.

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. However, there does seem to be an error with the amino acid sequence proposed by GENI-ACT for LBA0035 which should be corrected.

Gene Name for *Lactobacillus acidophilus*	Proposed Gene Function
LBA0034	AAC(3) family N-acetyltransferase (antibiotic properties)
LBA0035	Dipeptidase (dipeptide metabolism)

References

Lee, 2009, Lactobacillus (Wikipedia). https://en.wikipedia.org/wiki/Lactobacillus

Mayo Clinic, 2010, Acidophilus: Evidence. http://www.mayoclinic.org/drugs-supplements/acidophilus/evidence/hrb-20058615.

Acknowledgments

We would like to thank Dr. Stephen Koury, Dr. Rama Dey-Rao, and Jennifer Cooke for their support and guidance with this project. Supported by an NIH Science Education Partnership (SEPA) Award - R25ODO10536

www.buffalo.edu