

SI Text

Organism. *Lactobacillus iners* AB-1 is a vaginal isolate from a healthy woman. For DNA isolation for genome sequencing, it was cultured on Columbia blood agar (VWR, DF0792-17) supplemented with 50 ml/L sheep's blood (Cedarlane, DSB050) at 37°C under strict anaerobic conditions.

Genome assembly. We received 621,824 reads from 454 pyrosequencing ranging between 100 bp and greater than 600 bp (Next-Gen Sequencing Facility at Iowa State University, Iowa). The reads were trimmed for low quality bases and leftover adapter sequences using the python script, `sff_extract.py` (downloadable at: http://bioinf.comav.upv.es/sff_extract/index.html). Of the total reads after trimming, the 68,626 reads greater than 500 bp were used to generate a preliminary assembly using the Minimus assembler (AMOS package) (1). The resulting contigs were used as long reads in combination with the approximately 12 million short (35 nt) paired reads from Illumina (Illumina, Inc. San Diego, California) for paired-end assembly by the Velvet assembly program (2). Contigs greater than 200 bp were retained, and these 47 contigs were used as the backbone for gap closure and scaffolding.

Gap closure and scaffolding. To close gaps between contigs, a method of “read walking” was employed. A BLAST database of all 454 reads greater than 500 bp was created using formatdb. Contig ends were BLASTed (3) against the database to find reads with a perfect overlap of at least 300 bp. If no reads matched, the end of the contig was trimmed back until a match could be found, or the required overlap was reduced. Reads matching the ends of contigs were compared to the database of 454 reads using BLAST in order to extend the contig ends until another contig was reached. If no matching reads were uncovered in the dataset of reads greater than 500 bp, a second database including reads greater than 400 bp (262,847 reads) was used. Once inter-contig reads were obtained, gaps between contigs were joined using Minimus (1). This read-walking approach was similarly used to place the rRNA operons by walking out of the contigs identified as 16s and 23s sequence by BLAST to the NCBI non-redundant database. Reads matching the ends of these contigs were imported into JalView (4) where the multiple sequence alignments could be partitioned into separate sequence paths into contig ends. Additional contig joinings used sequence information from an unfinished *Lactobacillus iners* project in the GenBank database (GenBank accession number: ACLN00000000). Mauve was used to evaluate the alignment and contig order between the two datasets (5). Initially, a coverage of 5 reads was required to extend a contig, and upon completion of gap filling all contigs were evaluated for expected sequence coverage and “mate-happiness” between Illumina reads as described below.

By read-walking, the genome was assembled into seven contiguous pieces which were then ordered into a single scaffold of 2 contigs and a total of 1,301,173 bp. Scaffolding order was determined based on maximum synteny with the most closely related organisms, *L. johnsonii* and *L. gasseri*, using the PGA platform (6). The contigs were also oriented and ordered based on expected positions of the rRNA operons (7). Optical Mapping (OpGen, Inc., Madison, Wisconsin) validated contig order and orientation, as well as the expected genome size of 1.3 Mb (Figure S3). Primers were designed for the ends of each scaffolded contig and long-range PCR (Expand Long

Template PCR System, Roche) was used to verify connections between neighboring contigs as predicted by optical mapping. PCR products were electrophoresed on a 1% agarose gel and bands were isolated, gel purified, and submitted for dideoxy chain termination sequencing (London Regional Genomics Institute, London, Ontario). Sanger reads were used to extend contig ends, and read-walking was again performed to close remaining gaps. The majority of contig ends were rRNA operons and the PCR and Sanger sequences confirmed these rRNA placements.

Validating Genome Coverage. The final scaffold was independently validated by separately mapping the original Illumina paired reads and the 454 reads and then assessing the expected read coverage. Using Novopaired (www.novocraft.com), 97.7% of the Illumina paired reads mapped back to the scaffold with unique alignment. The expected distance between read pairs, termed “mate-happiness”, was also evaluated from the Novopaired output to identify areas of compression (mates too close together) or expansion (mates too far apart). We used in-house developed software and the BLAST algorithm to map the 454 reads to the scaffold using cutoffs of 98% ID, 95% read length excluding reads with insertions or deletions compared to the scaffold. Coverage and mate happiness for each final scaffold was visualized using R (8) (Figures S12 and S13). One region identified by optical mapping suggested ~12,000 bases missing in the final assembly (Figure S3). This region had good coverage and mate-happiness by the Illumina reads, but low coverage of 454 reads suggesting short tandem repeats which could not be resolved.

Annotation. Open reading frames (ORFs) greater than 100 nt were predicted using GeneMark (9) and by Glimmer (10). The translated ORF predictions were compared to the NCBI non-redundant database (nrdb) using BLASTp to evaluate gene predictions. There were several cases of GeneMark predicted ORFs that were shorter in length compared to homologous sequences in the database, and the corresponding Glimmer prediction better matched the length of the proteins in the database. In these instances, the Glimmer prediction was preferentially retained over the GeneMark prediction in order to prevent overestimation of truncated pseudogenes.

Several tools combined with custom-created Perl scripts were used for the manual annotation of predicted ORFs including: BLAST to NCBI nrdb, COG (Clusters of Orthologous Groups of proteins) (11), and LaCOG (Lactobacillales-specific Clusters of Orthologous protein coding Genes) (12) databases. In addition, automatic annotation results were collected from RAST (Rapid Annotation using Subsystem Technology) (13). Metabolic predictions were made by KAAS (KEGG Automatic Annotation Server) (14). HMMER (<http://hmmer.org/>) was used to search the sortase database (http://bamics3.cmbi.kun.nl/sortase_substrates) for sortase substrates (15). The predicted ORFs were also submitted to Pfam (16) and TMHMM (<http://www.cbs.dtu.dk/services/TMHMM/>) for conserved domain and transmembrane domain predictions respectively. A list of Pfam domains related to adhesion adapted from Kankainen et al. (17) was used to search the predicted ORFs with *E* values of 1e-3 or less considered significant (Table S6). A summary of Surface and exported proteins were identified using SLEP (18). Transporter genes were predicted and annotated by TransAAP

(<http://www.membranetransport.org/>). Transfer RNA (tRNA) genes were identified using tRNAscan-SE (19) Annotation predictions were manually validated.

Prediction of highly expressed genes (using the codon adaptation index, CAI). Highly expressed genes show a biased codon usage pattern, and the CAI provides an indication of expression (20). Using the EMBOSS (21) cai tool, the ORF sequence of the generally highly expressed small and large subunits of the ribosomal proteins were used as a reference for calculating the codon adaptation index (CAI) of all the predicted genes in the *L. iners* genome.

Comparative Genomics. Comparative genomics of *L. iners* was performed largely using sequence acquired from the GenBank database for *L. crispatus* JV-V01 (GenBank accession: ACKR00000000), *L. johnsonii* NCC 533 (NC_005362), *L. gasseri* ATCC 33323 (NC_008530), *L. acidophilus* NCFM (NC_006814), and *L. delbrueckii* subsp. *bulgaricus* ATCC 11842 (NC_008054).

Ortholog predictions for each pair of organisms was predicted by InParanoid (22) using the default settings, and custom Perl scripts were developed in-house to merge the datasets into a four-way comparison of overlapping ortholog predictions. The lists were submitted to the 4-way Venn Diagram Generator (<http://www.pangloss.com/seidel/Protocols/venn4.cgi>) and the predicted values were used to generate a figure. Unique proteins were defined as those without ortholog predictions. For simplicity, excluded from the total count were predicted orthologous groups between organism pairs that were not a one-to-one orthologous pairing (if a protein from organism A had more than one predicted ortholog from organism B, and vice versa). Other homology predictions are based on BLAST results to the NCBI nrdb.

Prediction of horizontally acquired genes. Predicted protein sequences from *L. iners* AB-1 and from the organisms listed above were compared to the NCBI nrdb by BLAST. Excluding self-hits and hits to the same species, genes were identified as foreign if the three most significant hits ($E \leq 1E^{-20}$) were a genus other than *Lactobacillus* with the most significant hit having at least 60% protein identity to the query sequence.

Construction of phylogenetic tree of select lactobacilli. Ribosomal protein sequences were chosen and tested for vertical transmission as outlined by Ciccarelli et al. (23). The following sequences contained clear orthologs in all the organisms and were vertically transmitted: ribosomal protein L1, L10, L11, L13, L15, L16, L17, L18, L2, L20, L21, L22, L23, L24, L27, L29, L3, L4, L5, L6, S10, S11, S12, S13, S20, S3, S4, S5, S7, S8, S9. These concatenated sequences were sent for phylogenetic analysis using the 'one click' mode at www.phylogeny.fr (24). This is a "default" mode which uses a pipeline already set up to run and connect programs recognized for their accuracy and speed (MUSCLE for multiple alignment, Gblocks for alignment curation, PhyML for phylogeny and finally TreeDyn for tree drawing) to reconstruct a robust phylogenetic tree from a set of sequences. The trees of cholesterol-dependent cytolysins (Figures S5 and S6) were created by the same method.

Cell-wall protein isolation. Protein isolation was performed using the method adapted from Cole, et al. (25). Briefly, *L. iners* was grown in 25ml MRS broth for 96 hours. The culture was

spun down, the supernatant was removed, and the pellet was frozen at -20°C until isolation. After thawing, the pellet was washed two times using TES buffer, resuspended in a 10% lysozyme solution in TES, and incubated at 37°C with shaking for 2h. After centrifugation, the supernatant containing the solubilized cell wall-associated proteins was collected. Proteins were precipitated using 10% TCA before separating by size by electrophoresis on SDS-PAGE, and bands were visualization by coomassie blue staining. For Western Blot analysis, the SDS-PAGE electrophoresed proteins were transferred to a polyvinylidene difluoride (PVDF) membrane for immuno-blotting as described below.

Mass spectrometry. After isolation of proteins, separation of by SDS-PAGE, and coomassie staining as described above, the visibly stained bands were isolated for mass spectrometry. In-gel digestion was performed using a MassPREP automated digester station (PerkinElmer). Gel pieces were Coomassie destained using 50mM ammonium bicarbonate and 50% acetonitrile (silver de-stained using a 50 mM sodium thiosulphate 5 hydrate and 15 mM potassium ferricyanide solution), which was followed by protein reduction using 10 mM dithiothreitol (DTT), alkylation using 55 mM iodoacetamide (IAA), and tryptic digestion. Peptides were extracted using a solution of 1% formic acid and 2% acetonitrile and lyophilized. Prior to mass spectrometry analysis, dried peptide samples were re-dissolved in a 10% acetonitrile and 0.1 % TFA (trifluoroacetic acid).

MALDI matrix, α -cyano-4-hydroxycinnamic acid (CHCA), was prepared as 5 mg/mL in 6 mM ammonium phosphate monobasic, 50% acetonitrile, 0.1 % trifluoroacetic acid and mixed with the sample at 1:1 ratio (v/v).

Mass Spectrometry data were obtained using a 4700 Proteomics Analyzer, MALDI TOF TOF (Applied Biosystems, Foster City, CA, USA). Data acquisition and data processing were respectively done using 4000 Series Explorer and Data Explorer (both from Applied Biosystems). The instrument is equipped with a 355 nm Nd:YAG laser; the laser rate is 200 Hz. Reflectron positive ion mode was used and the instrument was calibrated at 50 ppm mass tolerance. Each mass spectrum was collected as a sum of 1000 shots.

Protein identification was made using a local MASCOT (<http://www.matrixscience.com>) server and a custom-build peptide mass database constructed from the predicted protein sequences of *L. iners* AB-1. Search criteria allowed for up to 3 missed cleavage sites and a 50 ppm tolerance, and searches were performed using constant carbamidomethyl and variable oxidation modifications.

Rabbit polyclonal antisera. Polyclonal antisera was raised in rabbits using formalin-fixed whole cells of *Lactobacillus iners* AB-1, by ProSci Incorporated (Poway, CA) following a standard two-month protocol.

Whole-cell lysates were prepared from overnight cultures of each of the following ten strains: *Staphylococcus aureus* subsp. Newman, *Escherichia coli* subsp. Nissle, *Atopobium vaginae*, *Gardnerella vaginalis*, *Lactobacillus rhamnosus* GR-1, *Lactobacillus reuteri* RC-14, *Lactobacillus crispatus*, *Lactobacillus gasseri* 33323, *Lactobacillus jensenii* RC-28, and *Lactobacillus johnsonii* ATCC 11506. For each culture, the cells from 1 ml of overnight culture were harvested by centrifugation, resuspended in phosphate-buffered saline (PBS), then lysed using a bead beater. Unlysed cells were subsequently removed by centrifugation and the lysates from all ten cultures were combined. Nitrocellulose membranes (5 × 5 cm squares) were

incubated in the lysate for 1h at room temperature, with the membranes being turned every 15 min. After removing the lysate and washing the membranes 3 × 10 min with tris-buffered saline (TBS), the membranes were blocked for 1 h in 10% skim milk. The polyclonal antisera was incubated with 5 blocked membranes for 60 to 90 min, with the membranes being turned over every 15 min. The membranes were replaced with 5 new membranes and incubation repeated. This process was repeated for a total of 20 membranes, with 100 µL samples being taken after each 5-membrane incubation for analysis by dot blotting.

Immuno-blot assay. Immuno-blot assays were performed using the Bio-Rad Goat Anti-Rabbit IgG (H+L) AP Immun-Blot® Assay Kit and the Bio-Rad Alkaline Phosphatase Conjugate Substrate Kit according to the recommendations of the manufacturer. The completeness of the absorption procedure was analyzed using the same ten strains that were used for the absorption protocol (see above) plus *Lactobacillus iners* AB-1.

Transmission electron microscopy and immunogold labeling. Overnight monocultures of *L. iners* grown in MRS and self-collected vaginal swabs from healthy volunteers were prepared using a standard method for electron microscopy and immunogold labeling at the Transmission Electron Microscopy Facility, Department of Microbiology and Immunology at the University of Western Ontario. After embedding samples in LR white resin and thin slicing, the sections were placed on 400 mesh nickel grids and treated with the anti-*L. iners* primary rabbit antisera. After washing, samples were then incubated with goat anti-rabbit IgG conjugated to 10nm gold particles (British Biocell International, EM.GAR10 for labeling. Samples were stained with 2% uranyl acetate and examined with a Philips EM 410 microscope.

Growth assays. A modified MRS (26) broth was made lacking carbohydrates and with the addition of bromocresol purple as a pH indicator. Carbohydrates were added at a final concentration of 20mg/ml and a final total broth volume of 10ml for each experiment. The carbohydrates tested were: glucose, lactose, mannose, fructose, maltose, and porcine stomach mucin type III (Sigma: M1778) which has a glycoprotein content similar to vaginal mucin (27, 28). Growth was compared to the standard MRS medium used to grow the strain (Difco™ Lactobacilli MRS Broth). Colony forming units (CFU) were enumerated by drop plating serial dilutions of the broth culture onto Columbia blood agar plates every 24h after inoculation. Optical density (OD) readings at 600 nm were also recorded.

References

1. Sommer D, Delcher A, Salzberg S & Pop M (2007) Minimus: A fast, lightweight genome assembler. *BMC Bioinformatics* 8: 64.
2. Zerbino DR & Birney E (2008) Velvet: Algorithms for de novo short read assembly using de bruijn graphs. *Genome Research* 18: 821-829.
3. Altschul SF, Gish W, Miller W, Myers EW & Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215: 403-410.
4. Waterhouse AM, Procter JB, Martin DMA, Clamp M & Barton GJ (2009) Jalview version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 25: 1189-1191.
5. Darling ACE, Mau B, Blattner FR & Perna NT (2004) Mauve: Multiple alignment of conserved genomic sequence with rearrangements. *Genome Research* 14: 1394-1403.
6. Zhao F, Zhao F, Li T & Bryant DA (2008) A new pheromone trail-based genetic algorithm for comparative genome assembly. *Nucl.Acids Res.* 36: 3455-3462.

7. de Vries MC, *et al* (2006) Comparative and functional analysis of the rRNA-operons and their tRNA gene complement in different lactic acid bacteria. *Syst Appl Microbiol* 29: 358-367.
8. R Development Core Team (2010) R: A language and environment for statistical computing
9. Isono K, McIninch JD & Borodovsky M (1994) Characteristic features of the nucleotide sequences of yeast mitochondrial ribosomal protein genes as analyzed by computer program GeneMark. *DNA Res* 1: 263-269.
10. Salzberg S, Delcher A, Kasif S & White O (1998) Microbial gene identification using interpolated markov models. *Nucl Acids Res* 26: 544-548.
11. Tatusov R, *et al* (2003) The COG database: An updated version includes eukaryotes. *BMC Bioinformatics* 4: 41.
12. Makarova K, *et al* (2006) Comparative genomics of the lactic acid bacteria. *Proceedings of the National Academy of Sciences* 103: 15611-15616.
13. Aziz R, *et al* (2008) The RAST server: Rapid annotations using subsystems technology. *BMC Genomics* 9: 75.
14. Moriya Y, Itoh M, Okuda S, Yoshizawa AC & Kanehisa M (2007) KAAS: An automatic genome annotation and pathway reconstruction server. *Nucl Acids Res* 35: W182-185.
15. Boekhorst J, de Been MWHJ, Kleerebezem M & Siezen RJ (2005) Genome-wide detection and analysis of cell wall-bound proteins with LPxTG-like sorting motifs. *J Bacteriol* 187: 4928-4934.
16. Finn RD, *et al* (2008) The pfam protein families database. *Nucl Acids Res* 36: D281-288.
17. Kankainen M, *et al* (2009) Comparative genomic analysis of *Lactobacillus rhamnosus* GG reveals pili containing a human- mucus binding protein. *Proceedings of the National Academy of Sciences* 106: 17193-17198.
18. Giombini E, Orsini M, Carrabino D & Tramontano A (2010) An automatic method for identifying surface proteins in bacteria: SLEP. *BMC Bioinformatics* 11: 39.
19. Lowe TM & Eddy SR (1997) tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res* 25: 955-964.
20. Sharp PM & Li W (1987) The codon adaptation index-a measure of directional synonymous codon usage bias, and its potential applications. *Nucl Acids Res* 15: 1281-1295.
21. Rice P, Longden I & Bleasby A (2000) EMBOSS: The European molecular biology open software suite. *Trends in Genetics* 16: 276-277.
22. O'Brien KP, Remm M & Sonnhammer ELL (2005) Inparanoid: A comprehensive database of eukaryotic orthologs. *Nucl Acids Res* 33: D476-480.
23. Ciccarelli FD, *et al* (2006) Toward automatic reconstruction of a highly resolved tree of life. *Science* 311: 1283-1287.
24. Dereeper A, *et al* (2008) Phylogeny.fr: Robust phylogenetic analysis for the non-specialist. *Nucl Acids Res* 36: W465-469.
25. Cole JN, Djordjevic SP & Walker MJ (2008) Isolation and solubilization of gram-positive bacterial cell wall-associated proteins. *2D PAGE: Sample Preparation and Fractionation* 295-311.
26. Man JCD, Rogosa M & Sharpe ME (1960) A medium for the cultivation of lactobacilli. *J Appl Microbiol* 23: 130-135.
27. Hill RRR (1986) Digestion of mucin polysaccharides in vitro by bacteria isolated from the rabbit cecum. *Curr Microbiol* 14: 117-120.
28. Wiggins R, Hicks SJ, Soothill PW, Millar MR & Corfield AP (2001) Mucinases and sialidases: Their role in the pathogenesis of sexually transmitted infections in the female genital tract. *Sexually Transmitted Infections* 77: 402-408.
29. Grant JR & Stothard P (2008) The CGView server: A comparative genomics tool for circular genomes. *Nucleic Acids Res* 36: W181-4.

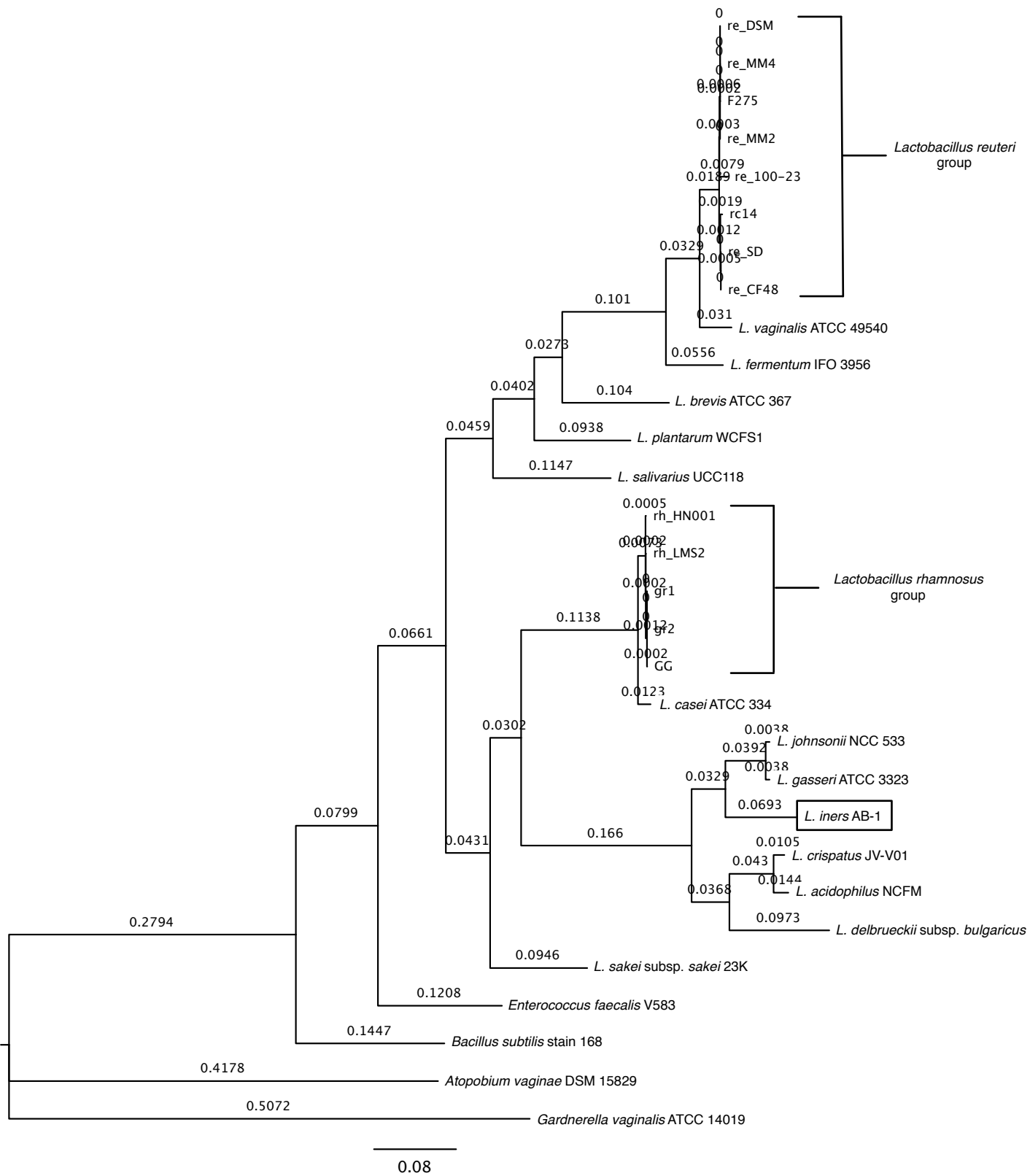


Figure S1. Phylogenetic tree constructed using concatenated ribosomal subunit protein sequences obtained from the GenBank database with the addition of sequence from *L. iners* AB-1. Amino acid substitution distances are given for each branch.

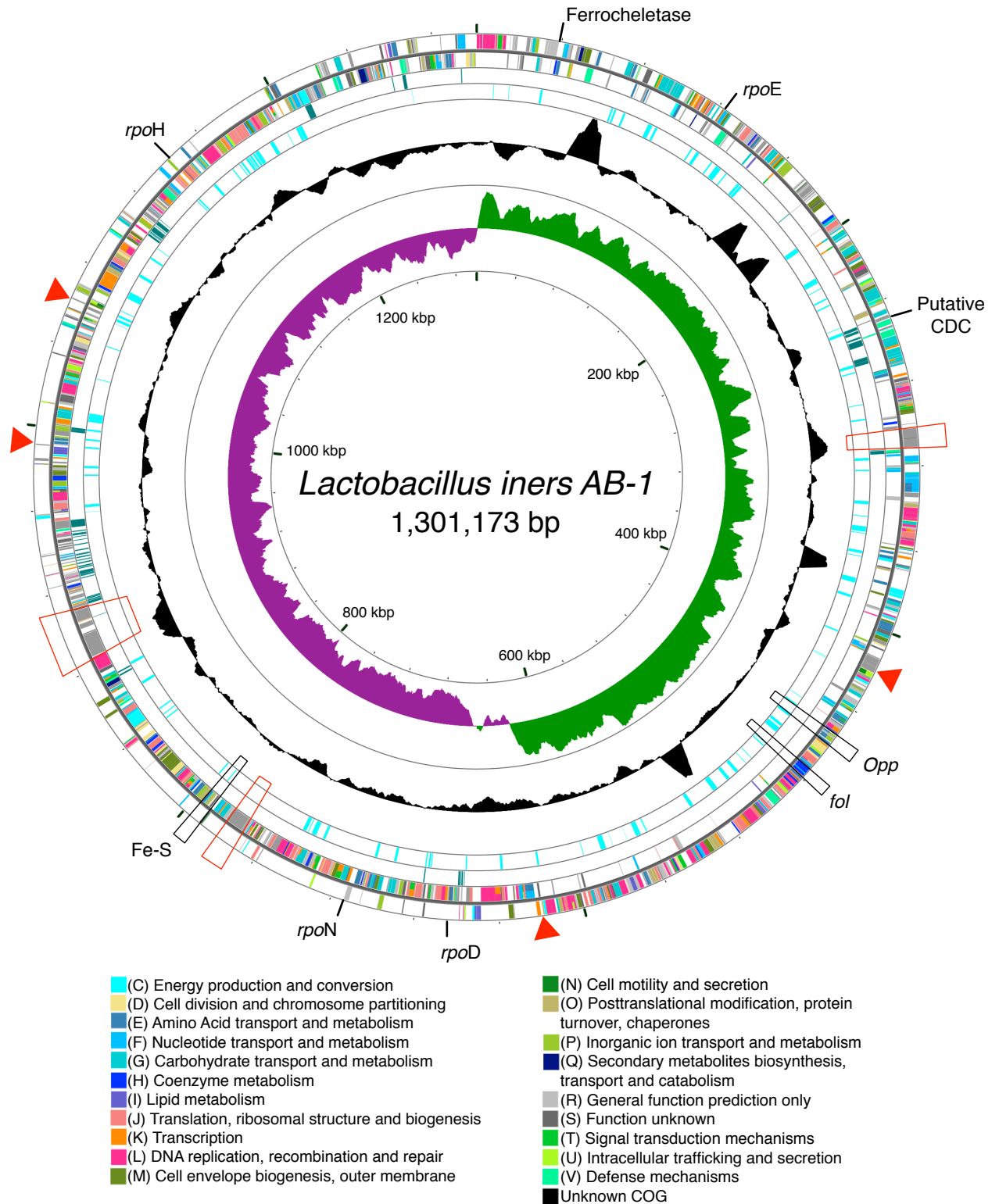


Figure S2. Genomic atlas of *L. iners* AB-1. From the outer circle inward: coding regions are marked on the first two rings: outside the dividing line if encoded on the positive strand, and inside the dividing line if encoded on the negative strand. The ORFs are color-coded by COG functional predictions. The third ring (dark green) marks ORFs predicted to be horizontally acquired. The fourth ring (cyan) shows ORFs predicted to be among the top 10% most highly expressed based on CAI. The fifth ring shows local CG content measured in a sliding window as a black plot. The innermost graph shows the CG-skew with sharp changes in skew occurring at the origin and terminus of replication. Genes of interest, as described in text, are marked on the outside of the atlas. Adhesins are marked in red. The atlas was constructed using the CGView Server (29).

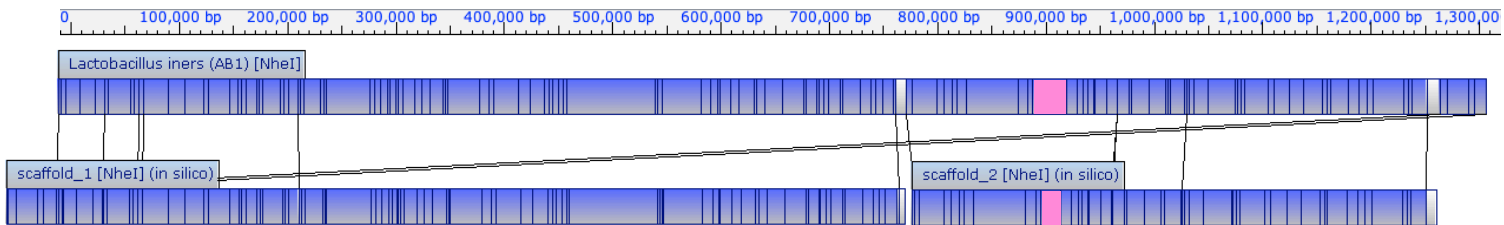


Figure S3. Alignment of the *L. iners* genome (2 scaffolds) to the optical map generated from chromosomal DNA cut with NheI. The scaffold sequence was cut in silico for comparison. The optical map confirms the assembled genome size of 1.3 mb. Two breaks of ~10 kb occur in the scaffold assembly due to tandemly repeated sequence (present in the 454 read set, and confirmed by PCR). Highlighted in pink is a region in the genome assembly which is approx. 12,000 bp shorter than the predicted optical map segment. The source of this 12 kb difference is unknown.

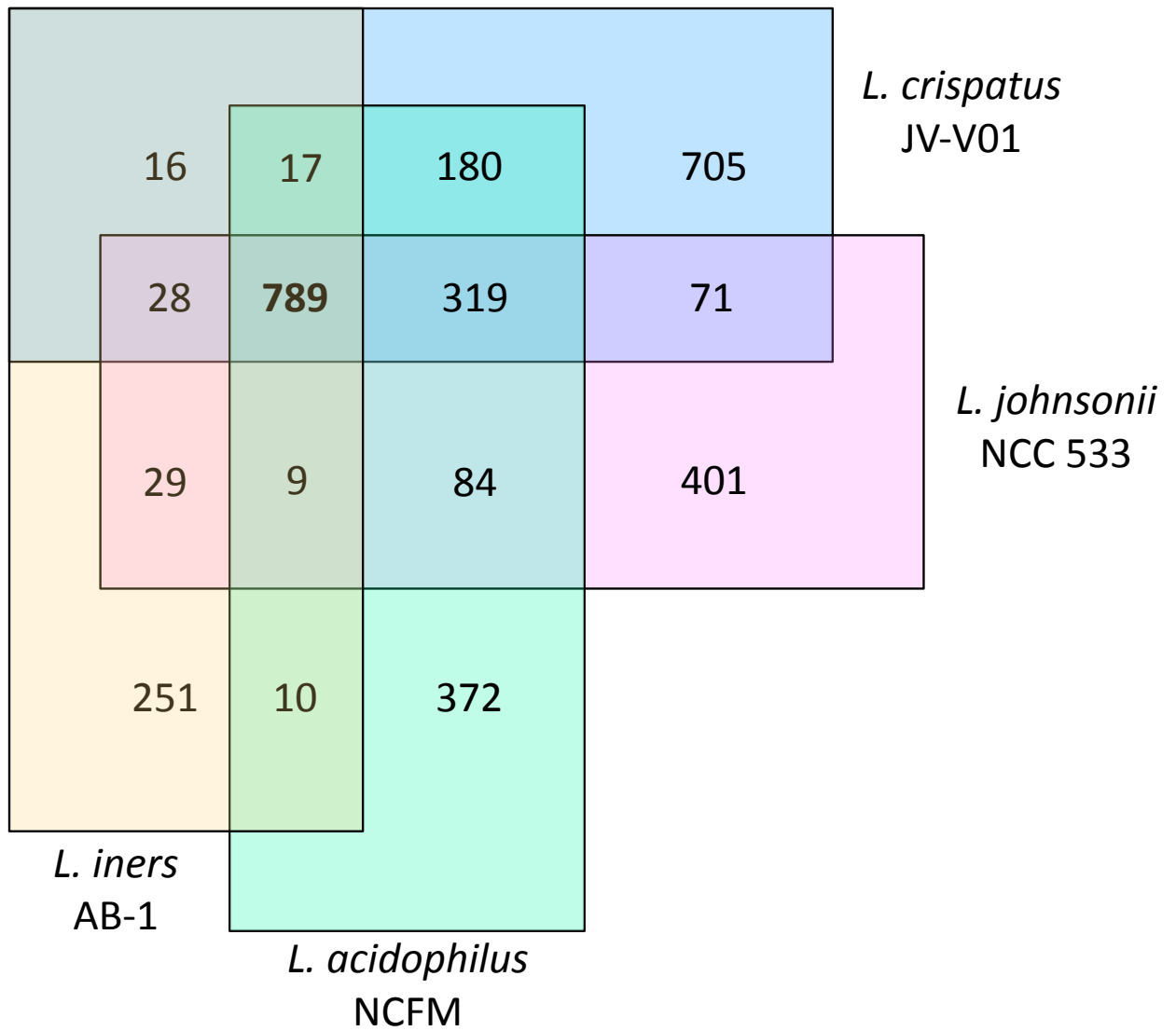


Figure S4. Venn diagram representing orthologous proteins between select species of the *L. iners* clade. Values are the number of orthologous genes between overlapping species as predicted by InParanoid.

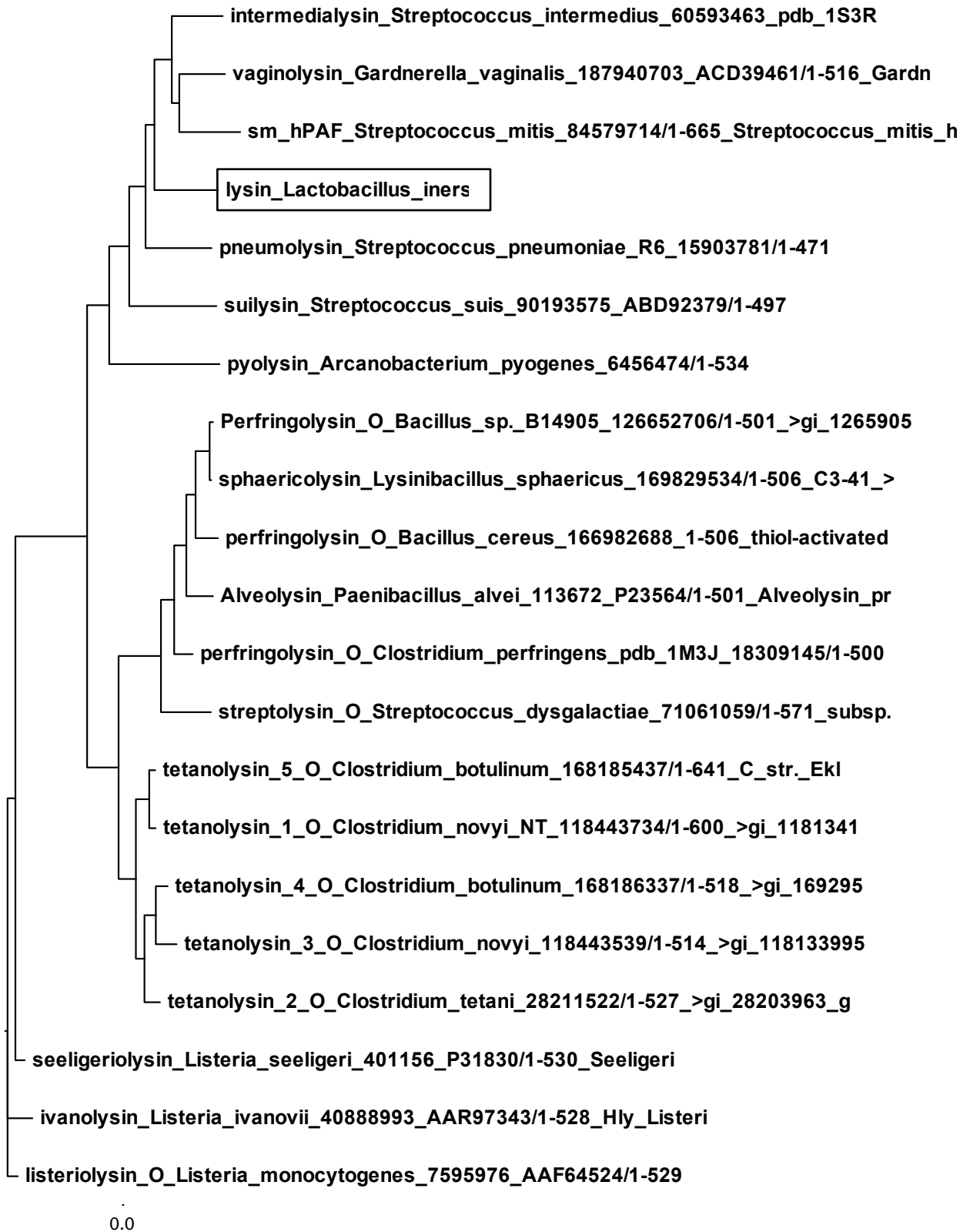


Figure S5. Phylogenetic tree constructed of various cholesterol-dependent cytolysins sequence from the GenBank database with the addition of LINAB1_0216 from *L. iners* AB-1. LINAB1_0216 is most closely related to the CD59-dependent intermedilysin family, however the binding domain 4 is more closely related to the CD59-independent tetanolysin (Figure S6).

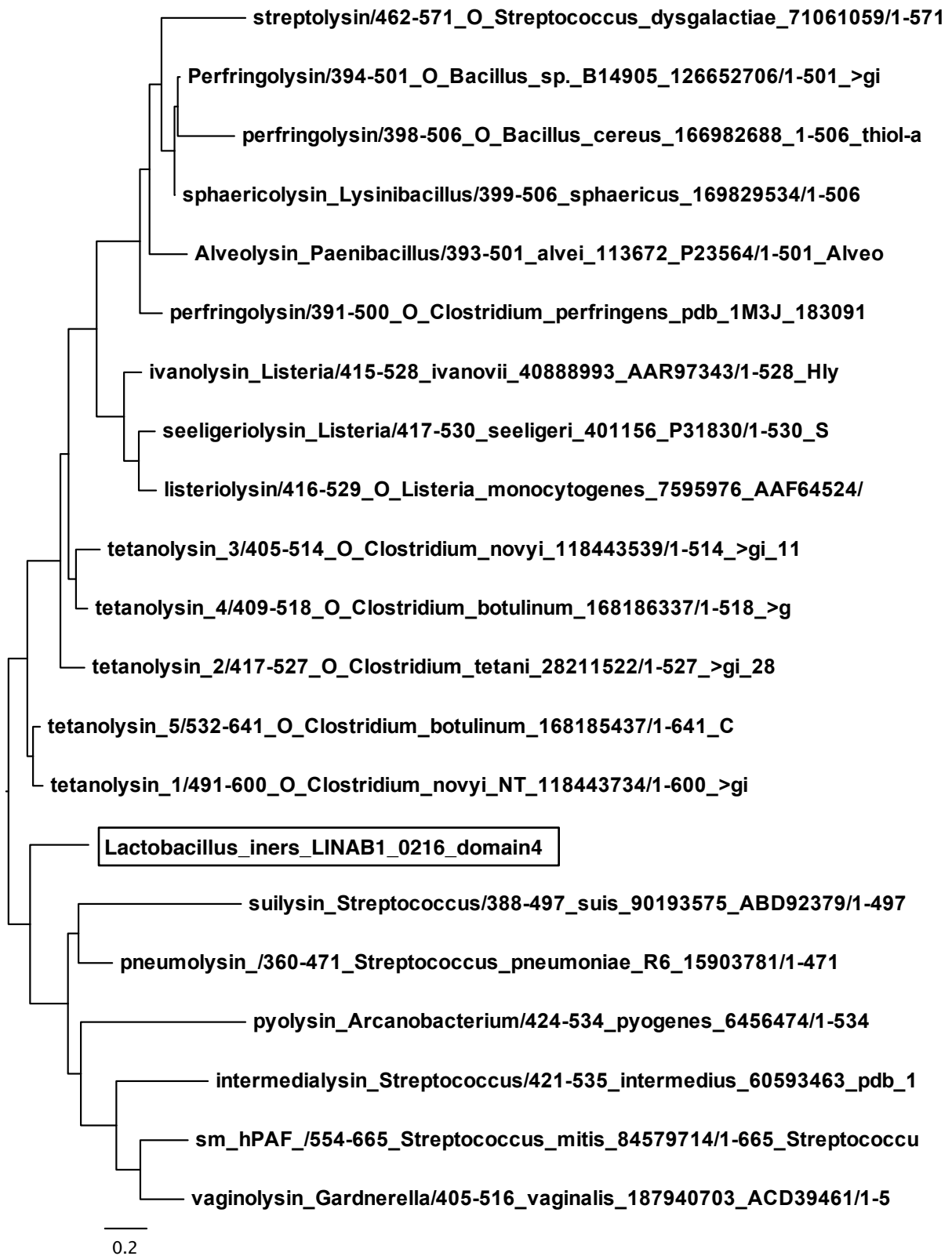


Figure S6. Phylogenetic tree constructed of domain 4 of various cholesterol- dependent cytolysins sequence from the GenBank database with the addition of LINAB1_0216 from *L. iners* AB-1. Domain 4 of LINAB1_0216 is most closely related to CD59- independent tetanolysin.

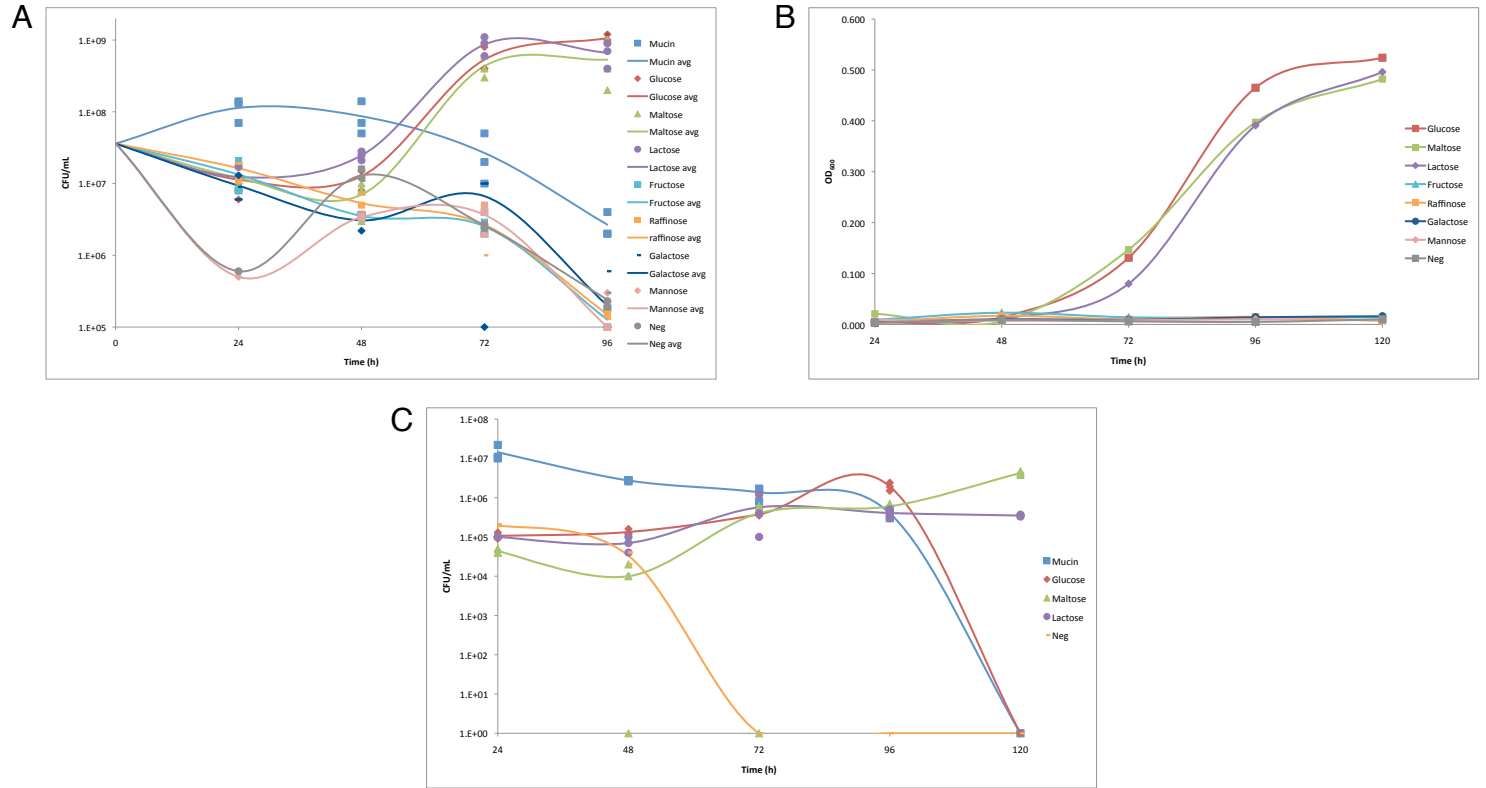
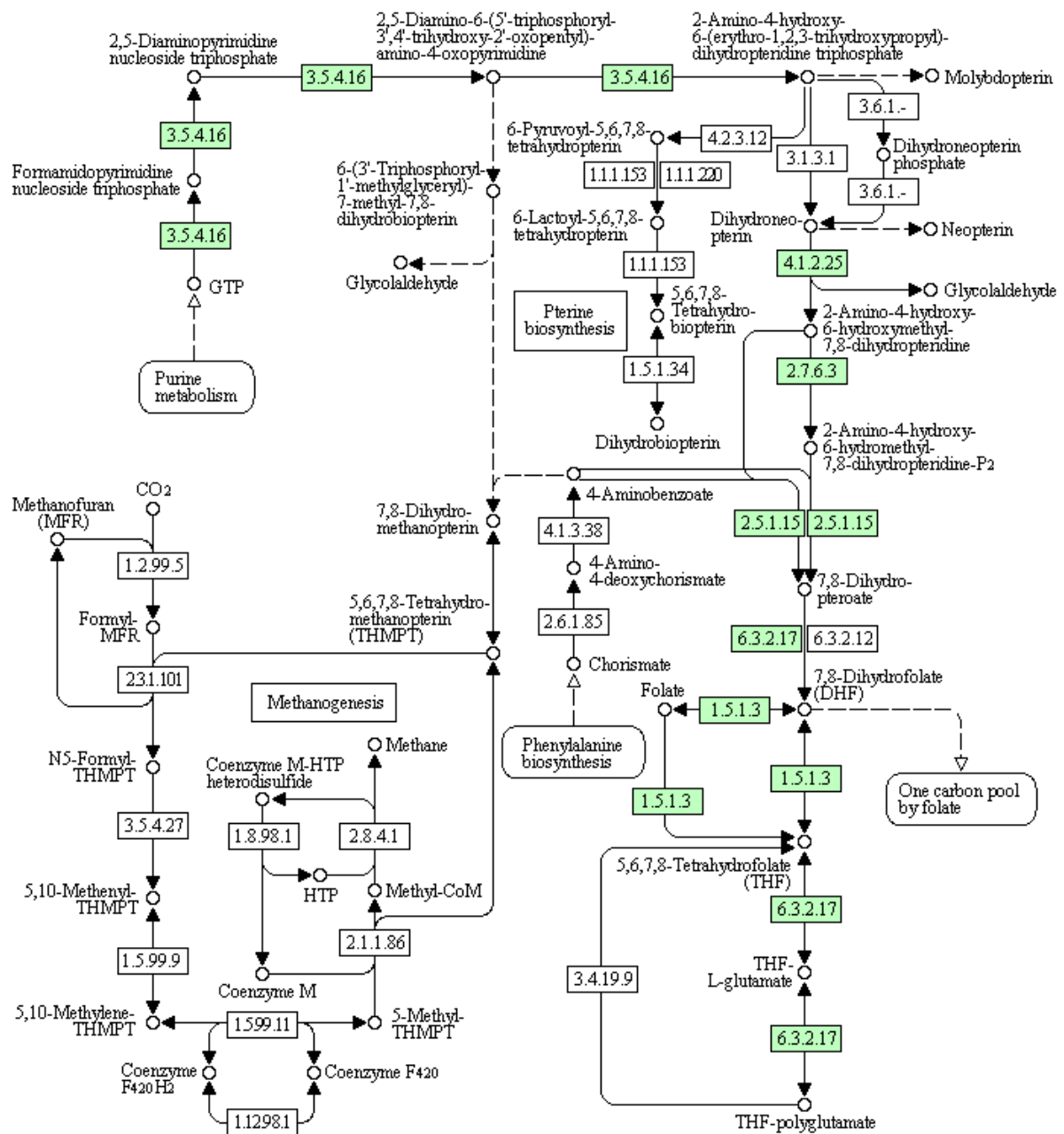


Figure S7. Representative growth of *L. iners* AB-1 in modified MRS substituted with various carbohydrate sources. Drop plates of serial dilutions of the broth cultures were performed every 24h after initial inoculation, and colony forming units (CFU) were enumerated after 72h incubation on Columbia blood agar (A). Additional to the CFU analysis, optical density (OD) readings at 600 nm were recorded every 24h (B). Repeating the experiment demonstrated similar growth patterns (C). Growth in mucin peaks shortly after initial inoculation at 24h while glucose, maltose, and lactose demonstrate peak growth at 96h. No growth was observed in fructose, raffinose, galactose, or mannose.

FOLATE BIOSYNTHESIS



00790 9/1/09
 (c) Kanehisa Laboratories

Figure S8. Folate synthesis pathway as predicted by KAAS. The genes present in *L. iners* AB-1 are shaded in green.

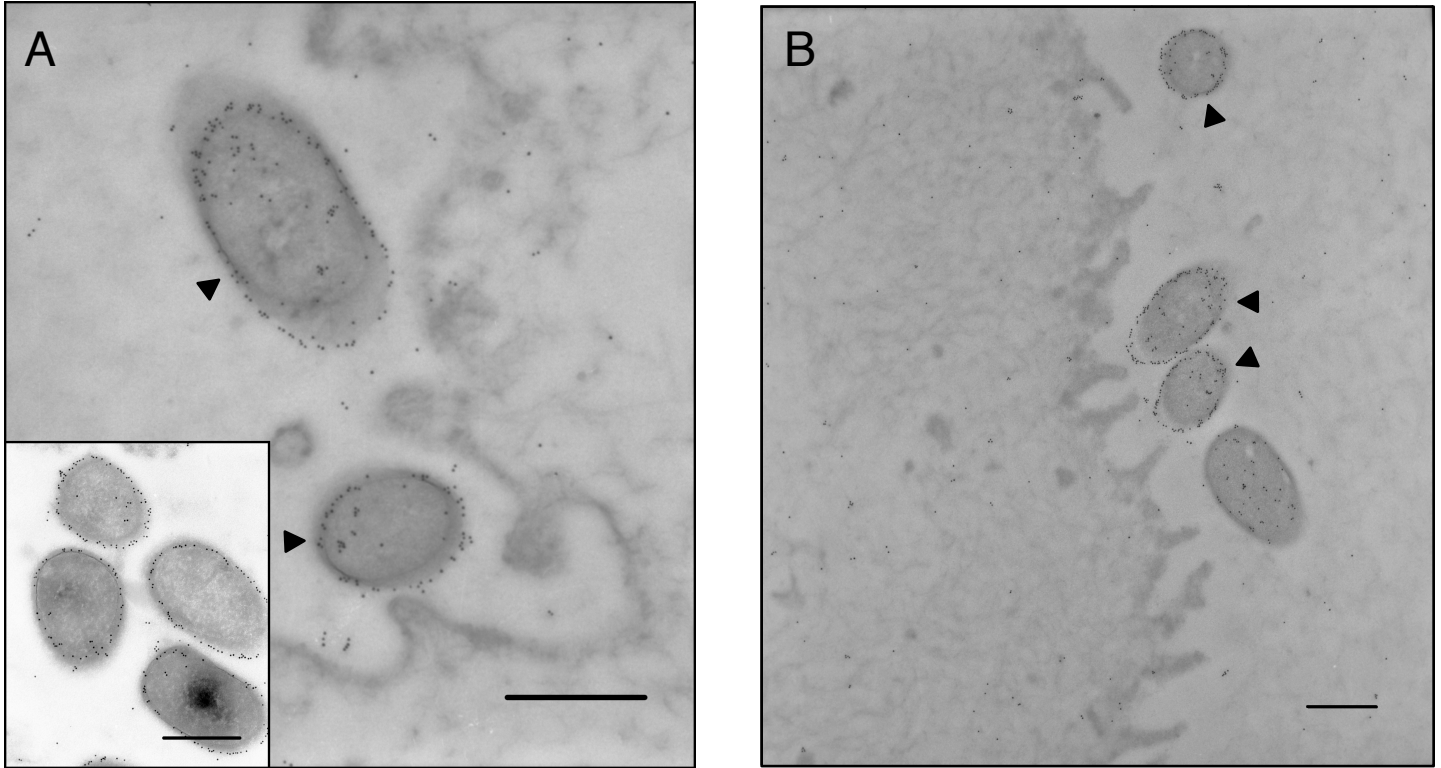


Figure S9. Transmission electron micrograph of immunogold labeled *L. iners*. Polyclonal antisera was raised in rabbits against formalin-fixed whole cells of *Lactobacillus iners* AB-1. The gold-labeled (10 nm) antibody appears specific for cell surface structures on the *L. iners* cell as shown by the inset micrograph of labeled *L. iners* monoculture (A). Vaginal swab samples collected from healthy women show well-labeled *L. iners* cells (black arrows) in association with the vaginal epithelial cells (A main panel and B). The scale bars represent 500 nm.

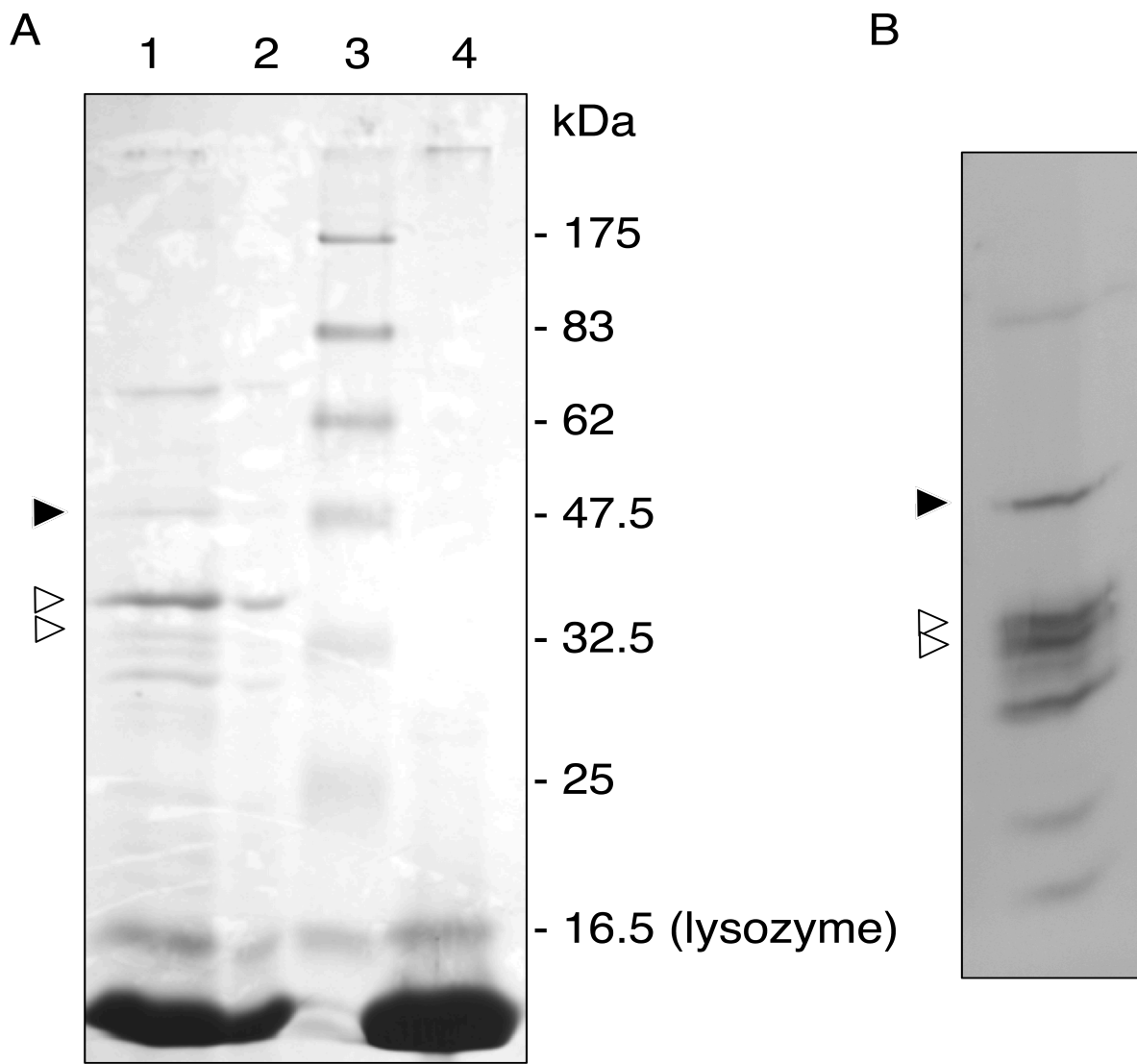


Figure S10. Coomassie stained SDS-PAGE (A) of cell-wall protein extract from *L. iners* AB-1 (lanes 1 and 2), molecular weight marker (lane 3), and lysozyme (lane 4). Molecular marker sizes in kDa are indicated. Bands confirmed to bind the *L. iners* polyclonal antisera by Western Blot (B) were extracted for mass spectrometry. The band indicated with the black arrow was identified as a unique cell surface protein containing a YSIRK signal peptide (LINAB1_0950). Two bands annotated with the white arrows were identified as an ABC-type transporter component (LINAB1_0341, 47219 kDa and 48% protein coverage with 12 masses) and L-lactate dehydrogenase (LINAB1_0978, 35025 kDa, 24% protein coverage with 9 peptide masses).

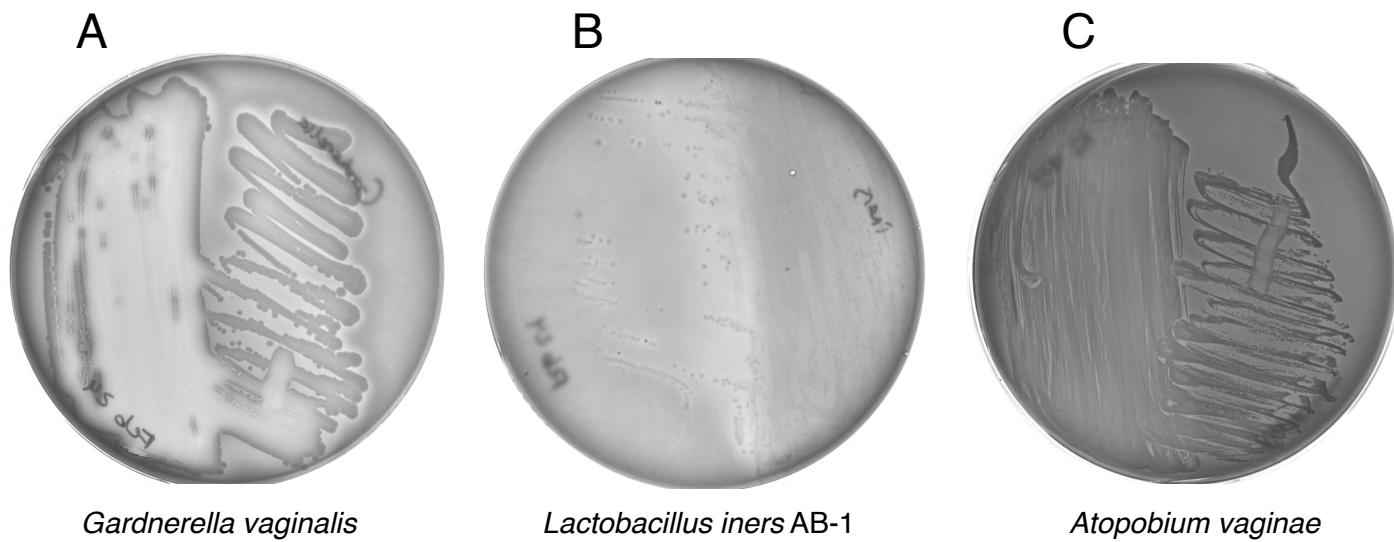


Figure S11. Vaginally-associated bacterial species grown on sheep's blood agar and photographed with transmitted white light. Hemolytic activity is observed as clearing of the blood agar by *Gardnerella vaginalis*, and to a lesser extent, *L. iners* AB-1. Single colonies display a zone of red blood cell clearing. In contrast, *Atopobium vaginae*, which is not predicted to contain a hemolysin, does not exhibit hemolytic activity.

Figure S12. Coverage map of Illumina reads mapped to the *L. iners* AB-1 assembled genome using Novopaired (figure available at http://www.biochem.uwo.ca/data/Coverage_ill.pdf). Reads mapped perfectly to the sequence with mate pairs the expected distance apart are represented by the blue plot with the height of the plot the measure of coverage. Yellow regions are areas of mate pairs too far apart, while cyan mate pairs are too close together. Note that there is uniform coverage across the vast majority of the genome with the exception of the ribosomal operons (regions of no coverage). Regions with low or high coverage, and regions of compressed or expanded mate pairs have been confirmed by either or all of: a) read walking, b) 454 read coverage (Figure S13), or c) by PCR.

Figure S13. Coverage map of 454 reads mapped to the *L. iners* AB-1 assembled genome using BLAST (figure available at http://www.biochem.uwo.ca/data/Coverage_400.pdf). Reads mapped perfectly to the sequence are represented by the blue plot with the height of the plot the measure of coverage. Red regions are areas with low coverage. Note that there is uniform coverage across the vast majority of the genome with the exception of the ribosomal operons (regions of no coverage). Regions with low or high coverage have been confirmed by either or all of: a) read walking, b) Illumina read coverage (Figure S12), or c) by PCR.

Table S1. General genome features of *L. iners* AB-1

Genome Size	1.3mb
GC Content, %	32.6
% coding	87.6
Total number of predicted ORFs	1190
ORFs with BLAST $\leq 1e-10$	1093 (91.8%)
ORFs with COG functional assignment	956 (80%)
rRNA operons	6
Plasmids	None

Table S2. Predicted pseudogenes of *L. iners* AB-1

ORF	Annotation	Length (AA)	HGT	Best hit to NCBI nrdb	Annotation of NCBI hit	Length of hit (AA)	e-val	%ID	ORF length/hit length	COG number	COG class
LINAB1_0276	hypothetical protein	70	*	gi 210623096 ref ZP_03293583.1	hypothetical protein CLOHIR_01533 [Clostridium hiranonis DSM 13275]	1077	6.00E-25	81.69	0.065	-	-
LINAB1_0277	hypothetical protein	76	*	gi 221076936 ref ZP_03548214.1	hypothetical protein BLAHAN_02002 [Blautia hansenii DSM 20583]	1035	3.00E-23	73.24	0.073	-	-
LINAB1_0025	hypothetical protein	38		gi 58336381 ref YP_192966.1	putative integrase [Lactobacillus acidophilus NCFM]	411	3.00E-14	92.31	0.092	-	-
LINAB1_0026	possible integrase	68		gi 58336381 ref YP_192966.1	putative integrase [Lactobacillus acidophilus NCFM]	411	7.00E-22	97.06	0.165	-	-
LINAB1_0845	hypothetical protein, phage-related	65		gi 77411267 ref ZP_00787617.1	phage portal protein, HK97 family [Streptococcus agalactiae CJB111]	358	3.00E-12	68	0.182	4695	S
LINAB1_0800	Putative cell-surface protein containing YSIRK and Gram-pos anchor ABC-type multidrug/protein/lipid transport system, SunT-containing, ATPase component	1804		gi 223043087 ref ZP_03613134.1	conserved hypothetical protein [Staphylococcus capitis SK14] ABC superfamily ATP binding cassette transporter ABC protein [Lactobacillus gasseri JV-V03]	9801	1.00E-97	30.7	0.184	-	-
LINAB1_0063	ATPase component	106		gi 227521926 ref ZP_03951975.1	terminase, large subunit, putative [Dehalococcoides ethenogenes 195]	537	5.00E-40	92.31	0.197	2274	V
LINAB1_0847	Phage terminase-like protein, large subunit	107	*	gi 57234156 ref YP_181803.1	SPP1 gp7 family phage head morphogenesis protein [Lactobacillus ruminis ATCC 25644]	535	1.00E-46	84.11	0.200	4626	R
LINAB1_0149	hypothetical protein	88		gi 227527121 ref ZP_03957170.1	ABC transporter ATPase component [Lactobacillus helveticus DPC 4571]	436	2.00E-16	58.21	0.202	-	-
LINAB1_0064	ABC transporter ATPase component	110		gi 161507527 ref YP_001577481.1	hypothetical protein FMG_0943 [Finegoldia magna ATCC 29328]	537	3.00E-49	89.09	0.205	1121	P
LINAB1_0994	conserved hypothetical protein	98	*	gi 169824640 ref YP_001692251.1	hypothetical protein ANACAC_00739 [Anaerostipes caccae DSM 14662]	446	4.00E-36	89.53	0.220	-	-
LINAB1_0844	phage portal family protein	101	*	gi 167746044 ref ZP_02418171.1	possible mucus binding protein precursor Mub [Lactobacillus jensenii JV-V16]	448	2.00E-37	74.26	0.225	4695	S
LINAB1_0795	Putative cell-surface protein containing YSIRK and Gram-pos anchor	2258		gi 227526342 ref ZP_03956391.1	50S ribosomal protein L13 [Lactobacillus acidophilus NCFM]	9775	7.00E-12	64.71	0.231	419	L
LINAB1_1046	LSU ribosomal protein L13	39		gi 58336661 ref YP_193246.1	conserved hypothetical protein [Anaerococcus lactolyticus ATCC 51172]	147	5.00E-14	94.87	0.265	102	J
LINAB1_0835	hypothetical protein	75	*	gi 227486554 ref ZP_03916870.1	hypothetical protein CLONEX_03262 [Clostridium nexile DSM 1787]	271	8.00E-25	77.33	0.277	-	-
LINAB1_0274	hypothetical nicotinate-nucleotide pyrophosphorylase (carboxylating)	86	*	gi 210616336 ref ZP_03291041.1	conserved hypothetical protein [Gardnerella vaginalis ATCC 14019]	299	1.00E-20	62.96	0.288	157	H
LINAB1_0369	hypothetical protein	904		gi 227507189 ref ZP_03937238.1	cytidine deaminase [Mobiluncus curtisii ATCC 43063]	3104	3.00E-32	29.32	0.291	-	-
LINAB1_0833	putative cysteine deaminase (truncated)	53		gi 227494012 ref ZP_03924328.1	integral membrane protein [Lactobacillus jensenii JV-V16]	181	6.00E-15	98.11	0.293	295	F
LINAB1_0329	integral membrane protein	63		gi 227525479 ref ZP_03955528.1	phage head morphogenesis protein, SPP1 gp7 family [Streptococcus suis 89/1591]	201	2.00E-13	64.41	0.313	4478	S
LINAB1_0181	conserved hypothetical protein	157		gi 223933186 ref ZP_03625177.1	Gmk, guanylate kinase [Lactobacillus jensenii 1153]	489	1.00E-33	48.25	0.321	-	-
LINAB1_0115	Guanylate kinase, truncated	61		gi 226817517 ref ZP_03806785.1	acyl-CoA thioesterase [Anaerococcus tetradius ATCC 35098]	185	2.00E-12	59.32	0.330	-	-
LINAB1_0820	possible acyl-CoA thioesterase	117	*	gi 227500902 ref ZP_03930951.1	possible autolysin [Lactobacillus jensenii JV-V16]	341	3.00E-50	82.05	0.343	-	-
LINAB1_0710	conserved hypothetical protein	212		gi 227525759 ref ZP_03955808.1	phage head morphogenesis protein, SPP1 gp7 family [Lactobacillus reuteri 100-23]	606	3.00E-17	33.94	0.350	-	-
LINAB1_0150	phage head morphogenesis protein partial ABC superfamily ATP binding cassette transporter, permease component	169		gi 194466762 ref ZP_03072749.1	ABC transporter, permease protein [Lactobacillus delbrueckii subsp. bulgaricus ATCC 11842]	481	4.00E-37	51.32	0.351	-	-
LINAB1_1120	putative integrase	90		gi 104773361 ref YP_618341.1	putative integrase [Finegoldia magna ATCC 29328]	248	2.00E-18	51.65	0.363	4587	R
LINAB1_0219	phage integrase	122	*	gi 169825072 ref YP_001692683.1	integrase [Lactobacillus reuteri CF48-3A]	327	3.00E-57	86.07	0.373	4974	L
LINAB1_1004	permease of the major facilitator superfamily	152		gi 227545543 ref ZP_03975592.1	MFS family major facilitator transporter [Enterococcus faecium TX1330]	405	2.00E-31	43.15	0.375	4974	L
LINAB1_0268	conserved hypothetical protein	182		gi 227552344 ref ZP_03982393.1	conserved hypothetical protein [Gardnerella vaginalis ATCC 14019]	457	2.00E-11	35	0.398	-	-
LINAB1_0368	conserved hypothetical protein	1273		gi 227507189 ref ZP_03937238.1	conserved hypothetical protein [Anaerococcus tetradius ATCC 35098]	3104	5.00E-69	31.67	0.410	-	-
LINAB1_0180	conserved hypothetical protein	122		gi 227500279 ref ZP_03930348.1	replication associated protein [Lactobacillus gasseri JV-V03]	285	4.00E-15	35.4	0.428	-	-
LINAB1_0113	conserved hypothetical protein	53		gi 227522066 ref ZP_03952115.1	hypothetical protein CLOBAR_01930 [Clostridium bartlettii DSM 16795]	123	5.00E-11	80.95	0.431	-	-
LINAB1_0220	type I restriction-modification system, subunit S	179	*	gi 164688285 ref ZP_02212313.1	transcriptional regulator [Lactobacillus jensenii JV-V16]	405	4.00E-55	60	0.442	732	V
LINAB1_0023	transcriptional regulator	126		gi 227525202 ref ZP_03955251.1	hypothetical protein ApreDRAFT_14040 [Anaerococcus prevotii DSM	282	3.00E-16	53.26	0.447	-	-
LINAB1_0988	conserved hypothetical protein	101	*	gi 228003647 ref ZP_04050634.1		218	3.00E-32	76.04	0.463	-	-

LINAB1_0841	conserved hypothetical protein	162	*	gi 126699471 ref YP_001088368.1	20548] hypothetical protein CD1860 [Clostridium difficile 630]	348	4.00E-81	92.99	0.466	4804	S
LINAB1_1108	Alpha-glycerophosphate oxidase, putative	286	*	gi 125718626 ref YP_001035759.1	Alpha-glycerophosphate oxidase, putative [Streptococcus sanguinis SK36]	608	1.00E-102	60.56	0.470	578	C
LINAB1_1159	di- and tricarboxylate transporter	174		gi 227889084 ref ZP_04006889.1	di- and tricarboxylate transporter [Lactobacillus johnsonii ATCC 33200]	368	1.00E-63	77.78	0.473	1055	P
LINAB1_0275	type I site-specific deoxyribonuclease	493	*	gi 221076936 ref ZP_03548214.1	hypothetical protein BLAHAN_02002 [Blautia hansenii DSM 20583]	1035	0	78.4	0.476	610	V
LINAB1_0273	Putative cell-surface protein containing Gram-pos anchor	1457		gi 227516270 ref ZP_03946319.1	possible biofilm-associated protein [Atopobium vaginae DSM 15829]	2902	1.00E-37	26.38	0.502	-	-
LINAB1_0502	competence protein ComEC	382		gi 227889787 ref ZP_04007592.1	competence protein ComEC [Lactobacillus johnsonii ATCC 33200]	760	7.00E-105	55.26	0.503	2333	R
LINAB1_0626	thermostable pullulanase	623		gi 58337968 ref YP_194553.1	thermostable pullulanase [Lactobacillus acidophilus NCFM]	1185	0	55.22	0.526	1523	G
LINAB1_1109	alpha-glycerophosphate oxidase	320	*	gi 148991959 ref ZP_01821733.1	alpha-glycerophosphate oxidase [Streptococcus pneumoniae SP9-BS68]	608	2.00E-135	75.62	0.526	578	C
LINAB1_1149	Transposase	140	*	gi 220927666 ref YP_002504575.1	Integrase catalytic region [Clostridium cellulolyticum H10]	266	7.00E-43	67.19	0.526	2801	L
LINAB1_1180	possible 5'-nucleotidase	143		gi 227360014 ref ZP_03844285.1	possible 5'-nucleotidase [Lactobacillus sakei subsp. carnosus DSM 15831]	270	2.00E-47	59.15	0.530	1011	R
LINAB1_0272	Putative cell-surface protein containing YSIRK motif	1559		gi 227516270 ref ZP_03946319.1	possible biofilm-associated protein [Atopobium vaginae DSM 15829]	2902	2.00E-35	33.88	0.537	-	-
LINAB1_0850	Resolvase domain-containing protein	170		gi 167757916 ref ZP_02430043.1	hypothetical protein CLOSCI_00247 [Clostridium scindens ATCC 35704]	308	6.00E-43	51.25	0.552	1961	L
LINAB1_0631	thermostable pullulanase	662		gi 58337968 ref YP_194553.1	thermostable pullulanase [Lactobacillus acidophilus NCFM]	1185	0	48.66	0.559	1523	G
LINAB1_0221	type I restriction-modification system, subunit S	234		gi 229878271 ref ZP_04497771.1	restriction endonuclease S subunit [Slackia heliotrinireducens DSM 20476]	416	8.00E-68	56.9	0.563	732	V
LINAB1_0512	conserved hypothetical protein	97		gi 58337168 ref YP_193753.1	hypothetical protein LBA0863 [Lactobacillus acidophilus NCFM]	171	5.00E-17	55.43	0.567	-	-
LINAB1_0069	amino acid transporter	263		gi 116630337 ref YP_815602.1	amino acid transporter [Lactobacillus gasseri ATCC 33323]	462	9.00E-116	77.78	0.569	531	E
LINAB1_0711	hypothetical protein	352		gi 227525759 ref ZP_03955808.1	possible autolysin [Lactobacillus jensenii JV-V16]	606	2.00E-23	31.52	0.581	-	-
LINAB1_0243	Cell wall-associated hydrolase	199		gi 199597375 ref ZP_03210805.1	Cell wall-associated hydrolase [Lactobacillus rhamnosus HN001]	332	3.00E-37	77.53	0.599	-	-
LINAB1_0824	putative toxin-antitoxin system DNA-damage-inducible protein J, RelB	56		gi 221194396 ref ZP_03567453.1	putative DNA-damage-inducible protein J [Atopobium rimae ATCC 49626]	93	3.00E-13	69.23	0.602	-	-
LINAB1_0057	transposase protein	255		gi 227878933 ref ZP_03996836.1	transposase protein [Lactobacillus crispatus JV-V01]	411	2.00E-50	48.71	0.620	-	-
LINAB1_0318	nucleoside-triphosphatase	131		gi 227888696 ref ZP_04006501.1	nucleoside-triphosphatase [Lactobacillus johnsonii ATCC 33200]	207	2.00E-56	79.69	0.633	127	F
LINAB1_0182	conserved hypothetical protein	42		gi 223933185 ref ZP_03625176.1	conserved hypothetical protein [Streptococcus suis 89/1591]	66	4.00E-12	85.37	0.636	-	-
LINAB1_0798	Putative fibrinogen-binding protein	777		gi 227499830 ref ZP_03929923.1	possible surface protein [Anaerococcus tetradius ATCC 35098]	1214	6.00E-72	35.88	0.640	-	-
LINAB1_0763	hypothetical hemolytic protein	61		gi 227526460 ref ZP_03956509.1	alpha-hemolysin family protein [Lactobacillus jensenii JV-V16]	95	8.00E-21	77.05	0.642	759	S
LINAB1_0838	transcriptional regulator	183	*	gi 168494490 ref ZP_02718633.1	transcriptional regulator [Streptococcus pneumoniae CDC3059-06]	284	2.00E-68	76.27	0.644	-	-
LINAB1_0825	toxin-antitoxin system DNA-damage-inducible protein J, RelB	66	*	gi 227507093 ref ZP_03937142.1	conserved hypothetical protein [Gardnerella vaginalis ATCC 14019]	101	1.00E-25	84.62	0.653	3077	L
LINAB1_0830	conserved hypothetical protein	44		gi 227516518 ref ZP_03946567.1	transcriptional regulator [Atopobium vaginae DSM 15829]	67	1.00E-16	95.45	0.657	1476	K
LINAB1_0118	conserved hypothetical membrane spanning protein	149		gi 227485471 ref ZP_03915787.1	hypothetical membrane spanning protein [Anaerococcus lactolyticus ATCC 51172]	222	7.00E-20	39.2	0.671	-	-
LINAB1_0674	translation initiation factor IF-3	117		gi 42519518 ref NP_965448.1	translation initiation factor IF-3 [Lactobacillus johnsonii NCC 533]	172	1.00E-39	83.64	0.680	290	J
LINAB1_1001	putative Zinc-dependent metalloproteinase	120		gi 194289064 ref YP_002004971.1	putative Zinc-dependent metalloproteinase [Cupriavidus taiwanensis]	175	2.00E-18	43.48	0.686	2856	E
LINAB1_0953	muramidase (lysozyme)	434		gi 227521754 ref ZP_03951803.1	muramidase [Lactobacillus gasseri JV-V03]	627	2.00E-72	58.96	0.692	1705	NU
LINAB1_0472	FAD synthase	216		gi 116629474 ref YP_814646.1	FAD synthase [Lactobacillus gasseri ATCC 33323]	312	1.00E-62	52.8	0.692	196	H
LINAB1_1139	transcriptional regulator	197		gi 227525202 ref ZP_03955251.1	transcriptional regulator [Lactobacillus jensenii JV-V16]	282	6.00E-41	41.46	0.699	-	-
LINAB1_1003	phage integrase family protein	191		gi 42518322 ref NP_964252.1	hypothetical protein LJ0236 [Lactobacillus johnsonii NCC 533]	273	1.00E-80	79.47	0.700	4974	L
LINAB1_0787	phosphomannose isomerase	229		gi 116629895 ref YP_815067.1	phosphomannose isomerase [Lactobacillus gasseri ATCC 33323]	321	2.00E-91	68.42	0.713	1482	G
LINAB1_0755	cysteine desulfurase	278		gi 227520661 ref ZP_03950710.1	cysteine desulfurase [Lactobacillus gasseri JV-V03]	384	3.00E-111	74.28	0.724	1104	E
LINAB1_1164	ABC transporter permease component	158		gi 42518158 ref NP_964088.1	ABC transporter permease component [Lactobacillus johnsonii NCC	217	3.00E-54	55.28	0.728	619	P

LINAB1_0577	putative primosome component related protein	158		gi 116629668 ref YP_814840.1	533] putative primosome component related protein [Lactobacillus gasseri ATCC 33323]	214	1.00E-51	61.01	0.738	3935	L
LINAB1_0990	conserved hypothetical protein	82	*	gi 227500141 ref ZP_03930211.1	conserved hypothetical protein [Anaerococcus tetradius ATCC 35098]	110	9.00E-29	93.9	0.745	4938	S
LINAB1_0524	LSU ribosomal protein L32p	47		gi 42519001 ref NP_964931.1	50S ribosomal protein L32 [Lactobacillus johnsonii NCC 533]	63	3.00E-18	93.62	0.746	333	J
LINAB1_0210	conserved hypothetical protein	92		gi 42518739 ref NP_964669.1	hypothetical protein LJ0814 [Lactobacillus johnsonii NCC 533]	123	2.00E-35	73.91	0.748	-	-
LINAB1_0951	d-ala,d-ala ligase	220		gi 227522105 ref ZP_03952154.1	d-ala,d-ala ligase [Lactobacillus gasseri JV-V03]	290	2.00E-52	47.17	0.759	-	-
LINAB1_1069	SSU ribosomal protein S17p (S11e)	67		gi 42518438 ref NP_964368.1	30S ribosomal protein S17 [Lactobacillus johnsonii NCC 533]	88	6.00E-20	94.03	0.761	186	J
LINAB1_0598	putative transcriptional regulator	93		gi 227521061 ref ZP_03951110.1	possible transcriptional regulator [Lactobacillus gasseri JV-V03]	122	1.00E-16	48.24	0.762	1725	K
LINAB1_0803	conserved hypothetical protein (DUF1814)	214	*	gi 228003637 ref ZP_04050625.1	protein of unknown function (DUF1814) [Anaerococcus prevotii DSM 20548]	278	3.00E-78	84.83	0.770	-	-
LINAB1_0408	ABC superfamily oligopeptide transport ATP-binding protein oppD	261		gi 227526056 ref ZP_03956105.1	oligopeptide ABC superfamily ATP binding cassette transporter ABC protein [Lactobacillus jensenii JV-V16]	338	1.00E-128	89.26	0.772	444	EP
LINAB1_0572	cell division initiation protein	113		gi 227520954 ref ZP_03951003.1	cell division initiation protein [Lactobacillus gasseri JV-V03]	145	5.00E-23	48.59	0.779	3599	D
LINAB1_0151	hypothetical protein	62		gi 227362953 ref ZP_03847100.1	hypothetical protein HMPREF0543_4180 [Lactobacillus sakei subsp. carnosus DSM 15831]	78	2.00E-14	67.74	0.795	1724	N
LINAB1_0956	conserved hypothetical protein	111		gi 227521758 ref ZP_03951807.1	conserved hypothetical protein [Lactobacillus gasseri JV-V03]	139	7.00E-38	61.26	0.799	-	-
LINAB1_0982	DEAD box ATP-dependent RNA helicase SrmB	388		gi 227888879 ref ZP_04006684.1	DEAD box ATP-dependent RNA helicase SrmB [Lactobacillus johnsonii ATCC 33200]	484	0	79.38	0.802	513	LKJ
LINAB1_0415	signal recognition particle docking protein FtsY	345		gi 227361857 ref ZP_03846062.1	signal recognition particle docking protein FtsY [Lactobacillus sakei subsp. carnosus DSM 15831]	430	1.00E-103	72.79	0.802	552	U
LINAB1_0305	preprotein translocase subunit YajC	121		gi 227521493 ref ZP_03951542.1	preprotein translocase subunit YajC [Lactobacillus gasseri JV-V03]	150	1.00E-30	61.95	0.807	1862	U
LINAB1_0498	possible rRNA (guanine-N(2)-)-methyltransferase	147		gi 227889783 ref ZP_04007588.1	possible rRNA (guanine-N(2)-)-methyltransferase [Lactobacillus johnsonii ATCC 33200]	182	2.00E-53	67.81	0.808	742	L
LINAB1_1167	beta-lactamase superfamily I metal-dependent hydrolase	215		gi 227521856 ref ZP_03951905.1	beta-lactamase superfamily I metal-dependent hydrolase [Lactobacillus gasseri JV-V03]	265	9.00E-85	68.84	0.811	1235	R
LINAB1_1040	CDF family cation diffusion facilitator	244		gi 227360777 ref ZP_03845025.1	CDF family cation diffusion facilitator [Lactobacillus sakei subsp. carnosus DSM 15831]	299	5.00E-69	61.14	0.816	1230	P
LINAB1_0585	phosphomevalonate kinase	311		gi 116629676 ref YP_814848.1	phosphomevalonate kinase [Lactobacillus gasseri ATCC 33323]	380	5.00E-115	68.39	0.818	1577	I
LINAB1_0634	major facilitator superfamily permease	382		gi 227516468 ref ZP_03946517.1	major facilitator superfamily permease [Atopobium vaginae DSM 15829]	466	3.00E-123	60.91	0.820	2814	G
LINAB1_0130	conserved hypothetical protein	280		gi 227522249 ref ZP_03952298.1	conserved hypothetical protein [Lactobacillus gasseri JV-V03]	341	6.00E-38	38.31	0.821	-	-
LINAB1_0995	conserved hypothetical protein	626	*	gi 227485418 ref ZP_03915734.1	conserved hypothetical protein [Anaerococcus lactolyticus ATCC 51172]	757	0	84.95	0.827	-	-
LINAB1_1189	ribonuclease P protein component	101		gi 42519905 ref NP_965835.1	ribonuclease P [Lactobacillus johnsonii NCC 533]	122	2.00E-27	61	0.828	594	J
LINAB1_0886	putative aggregation promoting protein	186		gi 227876979 ref ZP_03995072.1	peptidoglycan-binding protein [Lactobacillus crispatus JV-V01]	223	7.00E-37	40.89	0.834	-	-
LINAB1_1141	conserved hypothetical protein	234		gi 116628783 ref YP_813955.1	hypothetical protein LGAS_0105 [Lactobacillus gasseri ATCC 33323]	279	3.00E-36	45.34	0.839	-	-
LINAB1_0998	conserved hypothetical protein	110		gi 227485397 ref ZP_03915713.1	conserved hypothetical protein [Anaerococcus lactolyticus ATCC 51172]	131	1.00E-28	58.18	0.840	-	-
LINAB1_0241	oligoendopeptidase F	504		gi 161508023 ref YP_001577990.1	endopeptidase F [Lactobacillus helveticus DPC 4571]	598	0	76.34	0.843	1164	E
LINAB1_1000	conserved hypothetical protein	190		gi 225550631 ref ZP_03771580.1	conserved hypothetical protein [Ureaplasma urealyticum serovar 2 str. ATCC 27814]	225	8.00E-34	42.31	0.844	-	-
LINAB1_1150	ribonuclease HI	216		gi 227525285 ref ZP_03955334.1	ribonuclease [Lactobacillus jensenii JV-V16]	255	3.00E-52	45.49	0.847	328	L
LINAB1_0728	SUN protein	372		gi 42519411 ref NP_965341.1	SUN protein [Lactobacillus johnsonii NCC 533]	438	1.00E-129	62.03	0.849	144	J
LINAB1_0058	PTS system, mannose/fructose/N-acetylgalactosamine-specific IIA/IIB component	282		gi 116630348 ref YP_815624.1	phosphotransferase system, mannose/fructose-specific component IIA [Lactobacillus gasseri ATCC 33323]	332	6.00E-134	81.05	0.849	3444	G
LINAB1_0133	Acetyltransferase, including N-acetylase of ribosomal protein	165		gi 227889409 ref ZP_04007214.1	possible acetyltransferase [Lactobacillus johnsonii ATCC 33200]	193	4.00E-36	47.27	0.855	1670	J

LINAB1_0913	thioredoxin reductase	267		gi 42518777 ref NP_964707.1	thioredoxin reductase [Lactobacillus johnsonii NCC 533]	311	3.00E-104	67.92	0.859	492	O
LINAB1_0217	type I restriction-modification system DNA methyltransferase, subunit M	432	*	gi 169825074 ref YP_001692685.1	type I restriction-modification system DNA methylase [Finegoldia magna ATCC 29328]	502	0	84.95	0.861	286	V
LINAB1_0345	uracil-DNA glycosylase	231		gi 116629926 ref YP_815098.1	uracil-DNA glycosylase [Lactobacillus gasseri ATCC 33323]	268	1.00E-104	77.92	0.862	692	L
LINAB1_0156	phage head morphogenesis protein	294		gi 227520598 ref ZP_03950647.1	SPP1 gp7 family phage head morphogenesis protein [Lactobacillus gasseri JV-V03]	341	9.00E-64	46.24	0.862	-	-
LINAB1_0829	conserved hypothetical protein	144	*	gi 227507099 ref ZP_03937148.1	hypothetical protein HMPREF0421_0692 [Gardnerella vaginalis ATCC 14019]	167	1.00E-60	88.89	0.862	-	-
LINAB1_0136	phosphoenolpyruvate-dependent sugar phosphotransferase system EIIBC, arbutin specific	545		gi 42519689 ref NP_965619.1	phosphoenolpyruvate-dependent sugar phosphotransferase system EIIBC, arbutin specific [Lactobacillus johnsonii NCC 533]	628	0	91.01	0.868	1263	G
LINAB1_0923	hypothetical protein	192		gi 116629963 ref YP_815135.1	hypothetical protein LGAS_1335 [Lactobacillus gasseri ATCC 33323]	221	1.00E-65	61.98	0.869	1739	S
LINAB1_0727	phosphoprotein phosphatase	218		gi 227521178 ref ZP_03951227.1	phosphoprotein phosphatase [Lactobacillus gasseri JV-V03]	250	8.00E-90	73.27	0.872	631	T
LINAB1_0311	conserved hypothetical protein	90		gi 227521487 ref ZP_03951536.1	protein of hypothetical function DUF965 [Lactobacillus gasseri JV-V03]	103	1.00E-41	97.65	0.874	4472	S
LINAB1_0116	cell wall-associated hydrolase	229		gi 58338000 ref YP_194585.1	cell wall-associated hydrolase [Lactobacillus acidophilus NCFM]	262	2.00E-54	44.75	0.874	791	M
LINAB1_0008	single-stranded DNA-binding protein	151		gi 227889150 ref ZP_04006955.1	single-stranded DNA-binding protein [Lactobacillus johnsonii ATCC 33200]	172	5.00E-53	67.44	0.878	629	L
LINAB1_0926	recombination protein A	319		gi 227889603 ref ZP_04007408.1	recombination protein A [Lactobacillus johnsonii ATCC 33200]	363	4.00E-153	88.89	0.879	468	L
LINAB1_0049	D-alanyl transfer protein	358		gi 116630358 ref YP_815642.1	D-alanyl transfer protein [Lactobacillus gasseri ATCC 33323]	407	8.00E-163	81.79	0.880	1696	M
LINAB1_0525	conserved hypothetical protein	68		gi 227889856 ref ZP_04007661.1	conserved hypothetical protein [Lactobacillus johnsonii ATCC 33200]	77	4.00E-13	54.55	0.883	-	-
LINAB1_0942	dithiol-disulfide isomerase	205		gi 227889584 ref ZP_04007389.1	dithiol-disulfide isomerase [Lactobacillus johnsonii ATCC 33200]	232	7.00E-36	41.71	0.884	-	-
LINAB1_0694	UDP-N-acetylglucosamine 2-epimerase	384		gi 116629355 ref YP_814527.1	UDP-N-acetylglucosamine 2-epimerase [Lactobacillus gasseri ATCC 33323]	432	0	83.81	0.889	381	M
LINAB1_0430	50S ribosomal protein L21	139		gi 116629402 ref YP_814574.1	ribosomal protein L21 [Lactobacillus gasseri ATCC 33323]	156	5.00E-34	77.78	0.891	261	J
LINAB1_0466	transcription elongation factor NusA	362		gi 226820417 ref ZP_03809536.1	transcription elongation factor NusA [Lactobacillus gasseri MV-22]	406	1.00E-155	84.53	0.892	195	K
LINAB1_0863	Asp-tRNAAsn/Glu-tRNA ^{Gln} amidotransferase C subunit	100		gi 116630134 ref YP_815306.1	Asp-tRNAAsn/Glu-tRNA ^{Gln} amidotransferase C subunit [Lactobacillus gasseri ATCC 33323]	112	1.00E-38	80.81	0.893	721	J
LINAB1_0989	conserved hypothetical protein	67	*	gi 227500142 ref ZP_03930212.1	conserved hypothetical protein [Anaerococcus tetradius ATCC 35098]	75	3.00E-28	93.94	0.893	3655	K
LINAB1_0894	Zinc ABC transporter, periplasmic-binding protein ZnuA	303	*	gi 160947326 ref ZP_02094493.1	hypothetical protein PEPMIC_01260 [Peptostreptococcus micros ATCC 33270]	338	1.00E-119	68.11	0.896	803	P
LINAB1_1018	xanthine phosphoribosyltransferase	192		gi 227528033 ref ZP_03958082.1	xanthine phosphoribosyltransferase [Lactobacillus ruminis ATCC 25644]	214	2.00E-71	66.67	0.897	503	F

Table S3. Predicted foreign genes of *L. iners* AB-1 acquired by horizontal gene transfer (HGT)

Gene ID	Annotation	Best hit to NCBI nrdb	% ID	e-val	Best hit Annotation	Species	COG number	COG Class
LINAB1_0159	phage protein	gi 77414311 ref ZP_00790468.1	62.11	3.00E-28	hypothetical protein SAL_0616	Streptococcus agalactiae 515	-	-
LINAB1_0216	Putative cholesterol-dependent cytolysin	gi 187940703 gb ACD39461.1	55.14	4.00E-149	vaginolysin	Gardnerella vaginalis	-	-
LINAB1_0217	type I restriction-modification system DNA methyltransferase, subunit M	gi 169825074 ref YP_001692685.1	84.95	0	type I restriction-modification system DNA methylase	Fingoldia magna ATCC 29328	286	V
LINAB1_0219	putative integrase	gi 169825072 ref YP_001692683.1	86.07	3.00E-57	putative integrase	Fingoldia magna ATCC 29328	4974	L
LINAB1_0220	type I restriction-modification system, subunit S	gi 164688285 ref ZP_02212313.1	60	4.00E-55	hypothetical protein CLOBAR_01930	Clostridium bartlettii DSM 16795	732	V
LINAB1_0222	type I restriction-modification system, restriction subunit R	gi 169825069 ref YP_001692680.1	79.65	0	type I restriction-modification system restriction subunit	Fingoldia magna ATCC 29328	610	V
LINAB1_0229	PTS system, lactose-specific IIA component	gi 15675727 ref NP_269901.1	71.84	6.00E-37	PTS system, lactose-specific IIA component	Streptococcus pyogenes M1 GAS	1447	G
LINAB1_0230	PTS system, lactose-specific IIB/IIC component	gi 222152959 ref YP_002562136.1	82.77	0	lactose-specific phosphotransferase system (PTS), IIBC component 2	Streptococcus uberis 0140J	1455	G
LINAB1_0231	6-phospho-beta-galactosidase	gi 24379883 ref NP_721838.1	83.73	0	6-phospho-beta-galactosidase	Streptococcus mutans UA159	2723	G
LINAB1_0274	hypothetical nicotinate-nucleotide pyrophosphorylase (carboxylating)	gi 210616336 ref ZP_03291041.1	62.96	1.00E-20	hypothetical protein CLONEX_03262	Clostridium nexile DSM 1787	157	H
LINAB1_0275	type I site-specific deoxyribonuclease	gi 221076936 ref ZP_03548214.1	78.4	0	hypothetical protein BLAHAN_02002	Blautia hansenii DSM 20583	610	V
LINAB1_0276	hypothetical protein	gi 210623096 ref ZP_03293583.1	81.69	6.00E-25	hypothetical protein CLOHIR_01533	Clostridium hiranonis DSM 13275	-	-
LINAB1_0277	hypothetical protein	gi 221076936 ref ZP_03548214.1	73.24	3.00E-23	hypothetical protein BLAHAN_02002	Blautia hansenii DSM 20583	-	-
LINAB1_0592	toxin-antitoxin system DNA-damage-inducible protein J, RelB	gi 227873638 ref ZP_03991875.1	68.97	4.00E-28	DNA-damage-inducible protein J	Oribacterium sinus F0268	3077	L
LINAB1_0593	Putative predicted metal-dependent hydrolase	gi 150391747 ref YP_001321796.1	62.71	9.00E-90	hypothetical protein Amet_4055	Alkaliphilus metalliredigens QYMF	1451	R
LINAB1_0594	toxin-antitoxin system DNA-damage-inducible protein J, RelB	gi 227873638 ref ZP_03991875.1	66.67	3.00E-27	DNA-damage-inducible protein J	Oribacterium sinus F0268	3077	L
LINAB1_0595	methionine sulfoxide reductase A	gi 227486043 ref ZP_03916359.1	80.87	1.00E-132	methionine sulfoxide reductase A	Anaerococcus lactolyticus ATCC 51172	-	-
LINAB1_0614	putative oxidoreductase	gi 227516147 ref ZP_03946196.1	66.31	5.00E-64	possible NAD(P)H dehydrogenase (quinone)	Atopobium vaginae DSM 15829	431	R
LINAB1_0790	ABC superfamily ATP binding cassette transporter, ABC protein	gi 228002662 ref ZP_04049656.1	66.8	2.00E-101	ABC-type antimicrobial peptide transport system, ATPase component	Anaerococcus prevotii DSM 20548	1136	V
LINAB1_0792	response regulator	gi 227873881 ref ZP_03992103.1	60.18	1.00E-61	response regulator	Oribacterium sinus F0268	745	TK
LINAB1_0799	DNA binding helix-turn helix protein	gi 212697019 ref ZP_03305147.1	84.13	1.00E-22	hypothetical protein ANHYDRO_01584	Anaerococcus hydrogenalis DSM 7454	1476	K
LINAB1_0802	conserved hypothetical protein	gi 227493773 ref ZP_03924089.1	67	2.00E-77	hypothetical protein HMPREF0573_1671	Mobiluncus curtisii ATCC 43063	-	-
LINAB1_0803	conserved hypothetical protein (DUF1814)	gi 228003637 ref ZP_04050625.1	84.83	3.00E-78	protein of unknown function (DUF1814)	Anaerococcus prevotii DSM 20548	-	-
LINAB1_0804	possible transcriptional regulator	gi 227486796 ref ZP_03917112.1	91.28	2.00E-89	possible transcriptional regulator	Anaerococcus lactolyticus ATCC 51172	5340	K
LINAB1_0805	alpha-ribazole-5-phosphate phosphatase	gi 227500905 ref ZP_03930954.1	74.72	2.00E-73	phosphoglycerate mutase	Anaerococcus tetradius ATCC 35098	406	G
LINAB1_0806	possible acetyltransferase	gi 227507094 ref ZP_03937143.1	88.69	2.00E-84	possible acetyltransferase	Gardnerella vaginalis ATCC 14019	3981	R
LINAB1_0818	conserved hypothetical protein	gi 227516504 ref ZP_03946553.1	81.54	6.00E-23	hypothetical protein HMPREF0091_0759	Atopobium vaginae DSM 15829	-	-
LINAB1_0820	possible acyl-CoA thioesterase	gi 227500902 ref ZP_03930951.1	82.05	3.00E-50	acyl-CoA thioesterase	Anaerococcus tetradius ATCC 35098	-	-
LINAB1_0821	Plasmid stabilization protein	gi 227516519 ref ZP_03946568.1	85.98	3.00E-47	conserved hypothetical protein	Atopobium vaginae DSM 15829	-	-
LINAB1_0822	toxin-antitoxin system DNA-damage-inducible protein J, RelB	gi 227507077 ref ZP_03937126.1	96	3.00E-50	conserved hypothetical protein	Gardnerella vaginalis ATCC 14019	3077	L
LINAB1_0823	ABC superfamily ATP binding cassette transporter, ABC protein	gi 227507072 ref ZP_03937121.1	92.07	0	ABC superfamily ATP binding cassette transporter, ABC protein	Gardnerella vaginalis ATCC 14019	488	R
LINAB1_0825	toxin-antitoxin system DNA-damage-inducible protein J, RelB	gi 227507093 ref ZP_03937142.1	84.62	1.00E-25	conserved hypothetical protein	Gardnerella vaginalis ATCC 14019	3077	L
LINAB1_0826	conserved hypothetical protein	gi 227507092 ref ZP_03937141.1	75.7	3.00E-43	conserved hypothetical protein	Gardnerella vaginalis ATCC 14019	-	-
LINAB1_0827	conserved hypothetical protein with DNA binding helix-turn helix domain	gi 227507085 ref ZP_03937134.1	96.19	4.00E-106	conserved hypothetical protein	Gardnerella vaginalis ATCC 14019	1396	K
LINAB1_0828	Cadmium resistance protein	gi 227507831 ref ZP_03937880.1	98.52	1.00E-87	possible cadmium binding protein	Gardnerella vaginalis ATCC 14019	4300	P

LINAB1_0829	conserved hypothetical protein	gi 227507099 ref ZP_03937148.1	88.89	1.00E-60	hypothetical protein HMPREF0421_0692	Gardnerella vaginalis ATCC 14019	-	-
LINAB1_0832	conserved hypothetical protein	gi 228003721 ref ZP_04050708.1	96.61	3.00E-86	hypothetical protein ApreDRAFT_14780	Anaerococcus prevotii DSM 20548	-	-
LINAB1_0834	NUDIX family hydrolase	gi 227507082 ref ZP_03937131.1	92.5	2.00E-83	nudix family hydrolase	Gardnerella vaginalis ATCC 14019	1051	F
LINAB1_0835	hypothetical protein	gi 227486554 ref ZP_03916870.1	77.33	8.00E-25	conserved hypothetical protein	Anaerococcus lactolyticus ATCC 51172	-	-
LINAB1_0838	transcriptional regulator	gi 168494490 ref ZP_02718633.1	76.27	2.00E-68	transcriptional regulator	Streptococcus pneumoniae CDC3059-06	-	-
LINAB1_0841	conserved hypothetical protein	gi 126699471 ref YP_001088368.1	92.99	4.00E-81	hypothetical protein CD1860	Clostridium difficile 630	4804	S
LINAB1_0844	phage portal family protein	gi 167746044 ref ZP_02418171.1	74.26	2.00E-37	hypothetical protein ANACAC_00739	Anaerostipes caccae DSM 14662	4695	S
LINAB1_0847	Phage terminase-like protein, large subunit	gi 57234156 ref YP_181803.1	84.11	1.00E-46	terminase, large subunit, putative	Dehalococcoides ethenogenes 195	4626	R
LINAB1_0848	conserved hypothetical protein	gi 213031123 emb CAR31460.1	82.71	6.00E-103	hypothetical protein	Streptococcus pneumoniae	-	-
LINAB1_0849	putative recombinase	gi 229828216 ref ZP_04454285.1	63.64	7.00E-34	hypothetical protein GCWU000342_00273	Shuttleworthia satelles DSM 14600	-	-
LINAB1_0851	NUDIX family hydrolase	gi 154484819 ref ZP_02027267.1	60.74	1.00E-55	hypothetical protein EUBVEN_02537	Eubacterium ventriosum ATCC 27560	494	LR
LINAB1_0853	conserved hypothetical protein	gi 225388653 ref ZP_03758377.1	76.41	0	hypothetical protein CLOSTASPAR_02389	Clostridium asparagiforme DSM 15981	-	-
LINAB1_0892	putative zinc ABC transporter permease protein	gi 227901992 ref ZP_04019797.1	73	3.00E-88	ABC superfamily ATP binding cassette transporter, membrane protein	Fingoldia magna ATCC 53516	1108	P
LINAB1_0893	putative zinc ABC transporter ATP-binding protein	gi 227486040 ref ZP_03916356.1	55.96	9.00E-59	ABC superfamily ATP binding cassette transporter, ABC protein	Anaerococcus lactolyticus ATCC 51172	1121	P
LINAB1_0894	Zinc ABC transporter, periplasmic-binding protein ZnuA	gi 160947326 ref ZP_02094493.1	68.11	1.00E-119	hypothetical protein PEPMIC_01260	Peptostreptococcus micros ATCC 33270	803	P
LINAB1_0987	DNA-methyltransferase Dcm	gi 229859874 ref ZP_04479531.1	71.95	1.00E-144	DNA-methyltransferase Dcm	Streptobacillus moniliformis DSM 12112	270	L
LINAB1_0988	conserved hypothetical protein	gi 228003647 ref ZP_04050634.1	76.04	3.00E-32	hypothetical protein ApreDRAFT_14040	Anaerococcus prevotii DSM 20548	-	-
LINAB1_0989	conserved hypothetical protein	gi 227500142 ref ZP_03930212.1	93.94	3.00E-28	conserved hypothetical protein	Anaerococcus tetradius ATCC 35098	3655	K
LINAB1_0990	conserved hypothetical protein	gi 227500141 ref ZP_03930211.1	93.9	9.00E-29	conserved hypothetical protein	Anaerococcus tetradius ATCC 35098	4938	S
LINAB1_0992	copper-exporting ATPase	gi 227500140 ref ZP_03930210.1	95.12	0	copper-exporting ATPase	Anaerococcus tetradius ATCC 35098	2217	P
LINAB1_0993	transcriptional repressor CopY	gi 227500139 ref ZP_03930209.1	95.3	1.00E-78	penicillinase repressor	Anaerococcus tetradius ATCC 35098	3682	K
LINAB1_0994	conserved hypothetical protein	gi 169824640 ref YP_001692251.1	89.53	4.00E-36	hypothetical protein FMG_0943	Fingoldia magna ATCC 29328	-	-
LINAB1_0995	conserved hypothetical protein	gi 227485418 ref ZP_03915734.1	84.95	0	conserved hypothetical protein	Anaerococcus lactolyticus ATCC 51172	-	-
LINAB1_0997	conserved hypothetical protein	gi 227485396 ref ZP_03915712.1	95.08	6.00E-24	hypothetical protein HMPREF0072_0799	Anaerococcus lactolyticus ATCC 51172	-	-
LINAB1_1108	Alpha-glycerophosphate oxidase, putative	gi 125718626 ref YP_001035759.1	60.56	1.00E-102	Alpha-glycerophosphate oxidase, putative	Streptococcus sanguinis SK36	578	C
LINAB1_1109	alpha-glycerophosphate oxidase	gi 148991959 ref YP_01821733.1	75.62	2.00E-135	alpha-glycerophosphate oxidase	Streptococcus pneumoniae SP9-BS68	578	C
LINAB1_1110	glycerol kinase	gi 146318376 ref YP_001198088.1	76.34	0	glycerol kinase	Streptococcus suis 05ZYH33	554	C
LINAB1_1149	Transposase	gi 220927666 ref YP_002504575.1	67.19	7.00E-43	Integrase catalytic region	Clostridium cellulolyticum H10	2801	L
LINAB1_1154	ABC superfamily ATP binding cassette transporter, ABC protein possible MDR	gi 227507118 ref ZP_03937167.1	72.06	0	ABC superfamily ATP binding cassette transporter, ABC protein	Gardnerella vaginalis ATCC 14019	1132	V
LINAB1_1184	Ribonucleotide reduction protein NrdI	gi 77411315 ref ZP_00787664.1	66.67	3.00E-44	nrdI protein	Streptococcus agalactiae CJB111	1780	F

Table S4. Predicted highly expressed genes of *L. iners* AB-1 based on an average codon adaptation index (CAI) of 0.7 or higher

Gene ID	Annotation	HGT	sm_cai	lg_cai	avg
LINAB1_1055	LSU ribosomal protein L36		0.866	0.882	0.874
LINAB1_1020	50S ribosomal protein L7/L12		0.825	0.841	0.833
LINAB1_0504	30S ribosomal protein S20		0.833	0.824	0.829
LINAB1_0627	phosphopyruvate hydratase		0.825	0.829	0.827
LINAB1_1046	LSU ribosomal protein L13		0.813	0.827	0.820
LINAB1_0444	30S ribosomal protein S16		0.811	0.808	0.810
LINAB1_0103	phosphoglycerate mutase		0.805	0.813	0.809
LINAB1_1072	30S ribosomal protein S3		0.802	0.803	0.803
LINAB1_0078	Maltose/maltodextrin ABC transporter, substrate binding periplasmic protein MalE		0.798	0.801	0.800
LINAB1_0007	30S ribosomal protein S6		0.799	0.798	0.799
LINAB1_0899	enolase		0.795	0.802	0.799
LINAB1_0901	phosphoglycerate kinase		0.79	0.802	0.796
LINAB1_0505	30S ribosomal protein S15P/S13E		0.795	0.79	0.793
LINAB1_0456	30S ribosomal protein S2		0.782	0.802	0.792
LINAB1_0509	trigger factor		0.785	0.791	0.788
LINAB1_1074	30S ribosomal protein S19		0.783	0.792	0.788
LINAB1_0121	30S ribosomal protein S7		0.783	0.79	0.787
LINAB1_1022	50S ribosomal protein L1		0.775	0.798	0.787
LINAB1_0524	LSU ribosomal protein L32p		0.792	0.78	0.786
LINAB1_0900	triosephosphate isomerase		0.782	0.787	0.785
LINAB1_0902	glyceraldehyde 3-phosphate dehydrogenase		0.775	0.793	0.784
LINAB1_0190	fructose-bisphosphate aldolase class-II		0.775	0.788	0.782
LINAB1_0984	50S ribosomal protein L31 type B		0.781	0.782	0.782
LINAB1_0120	30S ribosomal protein S12		0.782	0.779	0.781
LINAB1_0122	elongation factor G (fusA)		0.772	0.781	0.777
LINAB1_1059	ribosomal protein L15		0.768	0.785	0.777
LINAB1_0528	pyruvate kinase		0.768	0.782	0.775
LINAB1_0540	DNA-binding protein HU		0.776	0.773	0.775
LINAB1_1157	conserved hypothetical protein		0.768	0.778	0.773
LINAB1_0610	YqeY domain-containing protein		0.767	0.776	0.772
LINAB1_0956	conserved hypothetical protein		0.763	0.781	0.772
LINAB1_0672	50S ribosomal protein L20		0.763	0.779	0.771
LINAB1_0015	hypothetical protein		0.821	0.718	0.770
LINAB1_0578	asparaginyl-tRNA synthetase		0.761	0.767	0.764
LINAB1_0757	30S ribosomal protein S4		0.763	0.762	0.763
LINAB1_0451	conserved hypothetical protein		0.746	0.778	0.762
LINAB1_0479	molecular chaperone DnaK		0.76	0.761	0.761
LINAB1_0978	L-lactate dehydrogenase		0.752	0.769	0.761
LINAB1_1065	30S ribosomal protein S14		0.76	0.761	0.761

LINAB1_0199	hit family protein		0.748	0.772	0.760
LINAB1_0125	elongation factor P		0.75	0.768	0.759
LINAB1_1076	50S ribosomal protein L23		0.741	0.775	0.758
LINAB1_0431	50S ribosomal protein L27		0.744	0.768	0.756
LINAB1_0950	Putative cell-surface protein containing YSIRK motif		0.752	0.759	0.756
LINAB1_0886	putative aggregation promoting protein		0.741	0.768	0.755
LINAB1_0235	galactose-6-phosphate isomerase subunit LacB		0.735	0.753	0.744
LINAB1_0341	ABC-type sugar transport system, periplasmic component		0.741	0.742	0.742
LINAB1_0668	hypothetical protein		0.734	0.749	0.742
LINAB1_1067	LSU ribosomal protein L24p (L26e)		0.73	0.753	0.742
LINAB1_0517	hypothetical protein		0.758	0.719	0.739
LINAB1_0300	chaperonin GroEL		0.732	0.744	0.738
LINAB1_0060	PTS system, mannose/fructose/N-acetylgalactosamine-specific IID component		0.739	0.732	0.736
LINAB1_1075	50S ribosomal protein L2		0.73	0.741	0.736
LINAB1_0375	hypothetical protein		0.737	0.732	0.735
LINAB1_0412	ABC superfamily oligopeptide ABC transporter, periplasmic oligopeptide-binding protein oppA		0.734	0.736	0.735
LINAB1_0310	alanyl-tRNA synthetase		0.728	0.74	0.734
LINAB1_1077	50S ribosomal protein L4		0.726	0.741	0.734
LINAB1_0433	elongation factor P		0.73	0.736	0.733
LINAB1_0508	translation elongation factor 1A (EF-1A/EF-Tu)		0.717	0.747	0.732
LINAB1_1035	aminopeptidase C		0.725	0.739	0.732
LINAB1_0216	inolysin	*	0.724	0.737	0.731
LINAB1_0135	maltose-6'-phosphate glucosidase		0.725	0.734	0.730
LINAB1_0873	glycosyltransferase		0.726	0.734	0.730
LINAB1_0075	maltose phosphorylase		0.723	0.732	0.728
LINAB1_0116	cell wall-associated hydrolase		0.724	0.731	0.728
LINAB1_0457	elongation factor Ts		0.724	0.732	0.728
LINAB1_1063	50S ribosomal protein L6		0.708	0.747	0.728
LINAB1_0326	conserved hypothetical protein		0.719	0.734	0.727
LINAB1_0662	possible nucleic acid-binding protein		0.726	0.727	0.727
LINAB1_0032	ABC superfamily ATP binding cassette transporter, binding protein		0.72	0.731	0.726
LINAB1_0561	manganese-dependent inorganic pyrophosphatase		0.721	0.73	0.726
LINAB1_0765	hypothetical protein		0.758	0.694	0.726
LINAB1_0334	acetaldehyde dehydrogenase (acetylating)		0.719	0.73	0.725
LINAB1_0612	hypothetical protein		0.73	0.72	0.725
LINAB1_0614	putative oxidoreductase	*	0.716	0.733	0.725
LINAB1_0661	DNA-binding response regulator		0.715	0.734	0.725
LINAB1_0949	hypothetical protein		0.729	0.721	0.725
LINAB1_0189	arginyl-tRNA synthetase		0.725	0.723	0.724
LINAB1_0305	preprotein translocase subunit YajC		0.724	0.723	0.724
LINAB1_0313	conserved hypothetical protein		0.709	0.737	0.723
LINAB1_0324	M20 family peptidase PepV		0.714	0.731	0.723
LINAB1_0193	conserved hypothetical protein		0.714	0.729	0.722

LINAB1_0861	aspartyl/glutamyl-tRNA amidotransferase subunit B		0.716	0.728	0.722
LINAB1_0065	asparagine synthetase AsnA		0.716	0.726	0.721
LINAB1_0362	hypothetical protein		0.714	0.725	0.720
LINAB1_0675	threonyl-tRNA synthetase		0.712	0.722	0.717
LINAB1_0760	universal stress protein UspA-like nucleotide-binding protein		0.716	0.718	0.717
LINAB1_0894	Zinc ABC transporter, periplasmic-binding protein ZnuA	*	0.715	0.719	0.717
LINAB1_1053	30S ribosomal protein S11		0.715	0.719	0.717
LINAB1_0766	F0F1 ATP synthase subunit epsilon		0.711	0.721	0.716
LINAB1_0213	Phosphocarrier protein of PTS system HPr		0.706	0.723	0.715
LINAB1_1066	50S ribosomal protein L5		0.71	0.719	0.715
LINAB1_0198	hypothetical protein		0.707	0.721	0.714
LINAB1_0635	glutamate--ammonia ligase		0.717	0.71	0.714
LINAB1_1054	30S ribosomal protein S13		0.704	0.724	0.714
LINAB1_0783	glucose-6-phosphate isomerase		0.708	0.717	0.713
LINAB1_0932	conserved hypothetical protein		0.7	0.725	0.713
LINAB1_1061	30S ribosomal protein S5		0.698	0.727	0.713
LINAB1_0590	signal peptidase I		0.706	0.718	0.712
LINAB1_1060	50S ribosomal protein L30		0.723	0.7	0.712
LINAB1_0527	6-phosphofructokinase		0.705	0.717	0.711
LINAB1_0669	PTS family fructose porter components IIABC		0.713	0.706	0.710
LINAB1_0279	adenylosuccinate synthetase		0.702	0.715	0.709
LINAB1_0377	hypothetical protein		0.71	0.708	0.709
LINAB1_0418	isoleucyl-tRNA synthetase		0.709	0.708	0.709
LINAB1_0469	translation initiation factor IF-2		0.704	0.714	0.709
LINAB1_0793	sugar ABC superfamily ATP binding cassette transporter, periplasmic component		0.717	0.7	0.709
LINAB1_1071	50S ribosomal protein L16		0.701	0.716	0.709
LINAB1_1087	methionyl-tRNA synthetase		0.707	0.71	0.709
LINAB1_1101	bleomycin hydrolase		0.703	0.715	0.709
LINAB1_1102	neutral endopeptidase		0.701	0.714	0.708
LINAB1_0402	50S ribosomal protein L28		0.705	0.708	0.707
LINAB1_0406	fatty acid/phospholipid synthesis protein PlsX		0.702	0.711	0.707
LINAB1_0463	prolyl-tRNA synthetase		0.703	0.711	0.707
LINAB1_1051	50S ribosomal protein L17		0.694	0.72	0.707
LINAB1_0058	PTS system, mannose/fructose/N-acetylgalactosamine-specific IIA/IIB component		0.702	0.71	0.706
LINAB1_0134	seryl-tRNA synthetase		0.699	0.713	0.706
LINAB1_0887	hypothetical protein		0.706	0.705	0.706
LINAB1_0952	lysine		0.696	0.715	0.706
LINAB1_0266	aspartate racemase		0.696	0.714	0.705
LINAB1_0890	D-methionine ABC superfamily ATP binding cassette transporter, binding protein		0.697	0.712	0.705
LINAB1_1045	30S ribosomal protein S9		0.713	0.696	0.705
LINAB1_0257	DNA replication initiation control protein YabA		0.712	0.696	0.704
LINAB1_0204	phenylalanine--tRNA ligase beta subunit		0.697	0.709	0.703
LINAB1_0285	orotate phosphoribosyltransferase		0.695	0.71	0.703

LINAB1_0712	non-heme iron-containing ferritin		0.708	0.698	0.703
LINAB1_0169	leucyl-tRNA synthetase		0.698	0.706	0.702
LINAB1_0856	hypothetical protein		0.729	0.674	0.702
LINAB1_0963	DNA-directed RNA polymerase subunit beta'		0.698	0.705	0.702
LINAB1_0231	6-phospho-beta-galactosidase	*	0.694	0.707	0.701
LINAB1_0416	dipeptidase		0.699	0.702	0.701
LINAB1_0718	possible phosphate transport regulator		0.697	0.704	0.701
LINAB1_0917	hypothetical protein		0.707	0.695	0.701
LINAB1_0969	lysyl-tRNA synthetase		0.696	0.706	0.701
LINAB1_0107	CTP synthetase		0.69	0.71	0.700
LINAB1_0877	teichoic acid biosynthesis protein		0.704	0.696	0.700
LINAB1_1094	conserved hypothetical protein		0.692	0.707	0.700

Table S5. Predicted transport protein of *L. iners* AB-1. Substrate/Function predictions from TransAAP.

Gene ID	FID	Subfamily	Substrate/Function	Family TC	Annotation	HGT	Possible pseudogene	avg CAI	COG number	COG Class
LINAB1_1152	ABC	membrane	amino acid (glutamine/glutamate/aspartate)	3.A.1	amino acid ABC superfamily ATP binding cassette transporter, membrane protein			0.608	765	E
LINAB1_1153	ABC	ABC	amino acid (glutamine/glutamate/aspartate)	3.A.1	ABC-type polar amino acid transport system, ATPase component			0.649	1126	E
LINAB1_1154	ABC	membrane	multidrug	3.A.1	ABC superfamily ATP binding cassette transporter, ABC protein possible MDR	*		0.568	1132	V
LINAB1_1159	DASS		sodium ion:anion symporter	2.A.47	di- and tricarboxylate transporter			0.583	1055	P
LINAB1_1163	ABC	ABC	cobalt ion	3.A.1	ABC-type cobalt transport system, ATPase component			0.601	1123	R
LINAB1_1172	NCS2		xanthine/uracil	2.A.40	NCS2 family nucleobase:cation symporter-2			0.571	2233	F
LINAB1_1177	LIVCS		branched-chain amino acid	2.A.26	LIVCS family branched chain amino acid:cation symporter			0.606	1114	E
LINAB1_1178	APC		glutamate:GABA antiporter	2.A.3	glutamate gamma-aminobutyrate antiporter			0.597	531	E
LINAB1_1188	Oxa1		60 KD inner membrane protein OxaA homolog	2.A.9	stage III sporulation protein J			0.598	706	U
LINAB1_0028	CIC		chloride ion channel	2.A.49	voltage-gated chloride channel family protein			0.570	38	P
LINAB1_0029	MFS		multidrug efflux	2.A.1	major facilitator superfamily permease			0.582	2814	G
LINAB1_0030	NCS2		xanthine/uracil	2.A.40	NCS2 family nucleobase:cation symporter-2			0.586	2252	R
LINAB1_0031	ABC	binding protein	sugar	3.A.1	ABC superfamily ATP binding cassette transporter, binding protein			0.633	1744	R
LINAB1_0032	ABC	binding protein	sugar	3.A.1	ABC superfamily ATP binding cassette transporter, binding protein			0.726	1744	R
LINAB1_0033	ABC	ABC	sugar	3.A.1	possible monosaccharide ABC superfamily ATP binding cassette transporter, ABC protein			0.643	3845	R
LINAB1_0035	ABC	membrane	sugar	3.A.1	ABC superfamily ATP binding cassette transporter, membrane protein			0.540	4603	R
LINAB1_0036	ABC	membrane	sugar	3.A.1	ABC superfamily ATP binding cassette transporter, membrane protein			0.594	1079	R
LINAB1_0038	POT		proton:dipeptide/tripeptide symporter	2.A.17	dipeptide/tripeptide transport protein			0.583	3104	E
LINAB1_0042	NCS2		xanthine/uracil	2.A.40	NCS2 family nucleobase:cation symporter-2			0.609	2252	R
LINAB1_0044	CPA2		potassium/sodium ion:proton antiporter	2.A.37	Na-H antiporter			0.604	475	P
LINAB1_0052	ABC	membrane	efflux (antimicrobial peptide)	3.A.1	ABC superfamily ATP binding cassette transporter, membrane protein, possible antimicrobial efflux			0.579	577	V
LINAB1_0054	ABC	ABC	efflux (antimicrobial peptide)	3.A.1	ABC superfamily ATP binding cassette transporter, ABC protein, possible antimicrobial efflux			0.562	1136	V
LINAB1_0055	APC		amino acid	2.A.3	amino acid transporter			0.584	531	E
LINAB1_0935	SSPTS	EnzymeIIA	glucitol/sorbitol	4.A	glucitol/sorbitol PTS, EIIA			0.685	3731	G
LINAB1_0934	AI-2E		Autoinducer-2 export	2.A.86	putative permease			0.613	628	R
LINAB1_0931	S-DNA-T		cell division/septum DNA translocation	3.A.12	FtsK/SpoIIIE family DNA translocase			0.584	1674	D
LINAB1_0919	IISP		preprotein translocase SecA	3.A.5	preprotein translocase subunit SecA			0.663	653	U
LINAB1_0898	IISP		preprotein translocase SecG	3.A.5	Preprotein translocase subunit SecG			0.516	1314	U
LINAB1_0894	ABC	binding protein	manganese/zinc ion	3.A.1	Zinc ABC transporter, periplasmic-binding protein ZnuA	*	*	0.717	803	P
LINAB1_0893	ABC	ABC	manganese/zinc ion	3.A.1	putative zinc ABC transporter ATP-binding protein			0.636	1121	P
LINAB1_0892	ABC	membrane	manganese/zinc ion	3.A.1	putative zinc ABC transporter permease protein	*		0.607	1108	P
LINAB1_0364	NCS1		cytosine/purines/uracil/thiamine/allantoin	2.A.39	cytosine permease			0.569	1457	F
LINAB1_0361	ABC	binding protein	manganese/zinc ion	3.A.1	ABC-type metal ion transport system, periplasmic component/surface adhesin			0.677	803	P
LINAB1_0353	ABC	binding protein	spermidine/putrescine	3.A.1	spermidine/putrescine ABC superfamily ATP binding cassette transporter, binding protein			0.671	687	E
LINAB1_0352	ABC	membrane	spermidine/putrescine	3.A.1	spermidine/putrescine ABC transporter permease			0.618	1177	E

LINAB1_0351	ABC	membrane	spermidine/putrescine	3.A.1	spermidine/putrescine ABC superfamily ATP binding cassette transporter, membrane protein		0.647	1176	E
LINAB1_0350	ABC	ABC	spermidine/putrescine	3.A.1	spermidine/putrescine ABC transporter ATPase		0.633	3842	E
	ABC	binding protein	sugar	3.A.1	ABC-type sugar transport system, periplasmic component		0.742	1653	G
LINAB1_0341									
LINAB1_0339	ABC	membrane	sugar	3.A.1	ABC-type sugar transport system, permease component		0.603	395	G
LINAB1_0338	ABC	membrane	sugar	3.A.1	ABC-type sugar transport system, permease component		0.608	1175	G
LINAB1_0337	ABC	ABC	spermidine/putrescine	3.A.1	ABC transporter ATPase component		0.622	3842	E
LINAB1_0333	APC		amino acid	2.A.3	APC family amino acid transporter		0.551	1113	E
LINAB1_0880	MFS		tetracycline efflux	2.A.1	major facilitator superfamily permease		0.613	-	-
LINAB1_0876	MOP	PST	polysaccharide export	2.A.66	PST family polysaccharide transporter		0.619	2244	R
LINAB1_0870	P-ATPase		calcium ion	3.A.3	cation-transporting ATPase		0.643	474	P
LINAB1_0857	NhaC		sodium ion:proton antiporter	2.A.35	Na ⁺ -H ⁺ -exchanging protein		0.571	1757	C
LINAB1_0713	ABC	membrane	unknown substrate	3.A.1	FeS assembly protein SufB		0.667	719	O
LINAB1_0716	ABC	membrane	unknown substrate	3.A.1	FeS assembly protein SufD		0.641	719	O
LINAB1_0717	ABC	ABC	unknown substrate	3.A.1	Iron-sulfur cluster assembly ATPase protein SufC		0.652	396	O
LINAB1_0719	PiT		phosphate	2.A.20	PiT family inorganic phosphate transporter		0.635	306	P
LINAB1_0720	ABC	ABC	amino acid (glutamine/glutamate/aspartate)	3.A.1	amino acid ABC superfamily ATP binding cassette transporter, ATP-binding		0.610	1126	E
LINAB1_0721	ABC	membrane	amino acid (glutamine/glutamate/aspartate)	3.A.1	amino acid ABC superfamily ATP binding cassette transporter, membrane protein		0.649	765	E
LINAB1_0766	F-ATPase		protons	3.A.2	F0F1 ATP synthase subunit epsilon		0.716	355	C
LINAB1_0767	F-ATPase		protons	3.A.2	F0F1 ATP synthase subunit beta		0.642	55	C
LINAB1_0768	F-ATPase		protons	3.A.2	F0F1 ATP synthase subunit gamma		0.615	224	C
LINAB1_0769	F-ATPase		protons	3.A.2	H(+)-transporting two-sector ATPase		0.652	56	C
LINAB1_0770	F-ATPase		protons	3.A.2	ATP synthase subunit delta		0.568	712	C
LINAB1_0771	F-ATPase		protons	3.A.2	ATP synthase subunit B		0.638	711	C
LINAB1_0772	F-ATPase		protons	3.A.2	ATP synthase subunit C		0.683	636	C
LINAB1_0773	F-ATPase		protons	3.A.2	ATP synthase subunit A		0.554	356	C
LINAB1_0781	ABC	membrane	sodium ion efflux	3.A.1	ABC transporter permease protein, possible sodium efflux		0.656	1668	CP
LINAB1_0782	ABC	ABC	sodium ion efflux	3.A.1	ABC transporter ATPase component		0.565	4152	R
LINAB1_0281	ABC	ABC	multidrug	3.A.1	ABC transporter, ATP-binding protein, possible MDR		0.608	1131	V
LINAB1_0283	FeoB		ferrous ion	9.A.8	GTP-binding protein HflX		0.588	2262	R
LINAB1_0287	MFS		metabolite (benzoate)	2.A.1	major facilitator superfamily permease		0.600	2814	G
LINAB1_0297	ABC	ABC	unknown substrate	3.A.1	ABC transporter ATPase component		0.623	488	R
LINAB1_0305	IISP		preprotein translocase YajC	3.A.5	preprotein translocase subunit YajC	*	0.724	1862	U
	MscS		small-conductance mechanosensitive ion channel	1.A.23	small-conductance mechanosensitive channel		0.530	668	M
LINAB1_0319									
LINAB1_0960	ABC	ABC	manganese/zinc ion	3.A.1	zinc/ion ABC superfamily ATP binding cassette transporter, ABC protein		0.602	1121	P
LINAB1_0959	ABC	membrane	manganese/zinc ion	3.A.1	metal cation ABC superfamily ATP binding cassette transporter, membrane protein		0.597	1108	P
LINAB1_0957	KUP		potassium ion uptake	2.A.72	KUP family potassium (K ⁺) uptake permease		0.620	3158	P
LINAB1_1040	CDF		cation efflux	2.A.4	CDF family cation diffusion facilitator		0.642	1230	P
LINAB1_1027	MFS		cyanate	2.A.1	major facilitator superfamily permease		0.554	2807	P
LINAB1_1019	NCS2		xanthine/uracil	2.A.40	xanthine/uracil permease		0.628	2233	F
LINAB1_1008	S-DNA-T		cell division/septum DNA translocation	3.A.12	FtsK family DNA segregation ATPase		0.566	1674	D
LINAB1_0992	P-ATPase		copper ion	3.A.3	copper-exporting ATPase	*	0.547	2217	P

LINAB1_0985	MFS		sugar	2.A.1	major facilitator sugar transporter		0.619	2211	G
LINAB1_0385	ABC	membrane	cobalt ion	3.A.1	ABC-type transport system, permease component		0.627	619	P
LINAB1_0386	ABC	ABC	dipeptide/oligopeptide	3.A.1	ATPase component of various ABC-type transport systems with duplicated ATPase domain		0.660	1123	R
	ABC	binding protein	unknown substrate	3.A.1	ABC transporter substrate-binding protein		0.669	2984	R
LINAB1_0389									
LINAB1_0390	ABC	membrane	unknown substrate	3.A.1	ABC transporter permease component		0.553	4120	R
LINAB1_0391	ABC	ABC	unknown substrate	3.A.1	ABC superfamily ATP binding cassette transporter, ABC protein		0.657	1101	R
LINAB1_0392	MFS		multidrug efflux	2.A.1	major facilitator superfamily permease		0.614	2814	G
LINAB1_0393	ABC	ABC	manganese/zinc ion	3.A.1	ABC-type metallic cation transport system, ATPase component		0.604	1121	P
LINAB1_0394	ABC	membrane	polysaccharide export	3.A.1	ABC-2 transporter, permease protein, putative		0.620	1682	GM
LINAB1_0395	NCS2		xanthine/uracil	2.A.40	NCS family uracil:cation symporter		0.542	2233	F
LINAB1_0396	MFS		multidrug efflux	2.A.1	MFS family major facilitator transporter		0.637	2814	G
LINAB1_0398	ABC	ABC	amino acid (glutamine/glutamate/aspartate)	3.A.1	ABC-type polar amino acid transport system, ATPase component		0.633	1126	E
	ABC	binding protein	amino acid (glutamine/glutamate/aspartate)	3.A.1	amino acid ABC superfamily ATP binding cassette transporter, binding protein		0.602	834	ET
LINAB1_0399									
LINAB1_0400	ABC	membrane	amino acid (glutamine/glutamate/aspartate)	3.A.1	ABC-type amino acid transport system, permease component		0.595	765	E
LINAB1_0401	ABC	membrane	amino acid (glutamine/glutamate/aspartate)	3.A.1	glutamine ABC transporter membrane protein		0.585	765	E
LINAB1_0408	ABC	ABC	dipeptide/oligopeptide	3.A.1	ABC superfamily oligopeptide transport ATP-binding protein oppD		0.657	444	EP
LINAB1_0409	ABC	ABC	oligopeptide	3.A.1	ABC superfamily oligopeptide transport ATP-binding protein oppF		0.679	1123	R
LINAB1_0410	ABC	membrane	dipeptide/oligopeptide	3.A.1	ABC superfamily oligopeptide transport system permease protein oppB		0.671	601	EP
LINAB1_0411	ABC	membrane	dipeptide/oligopeptide	3.A.1	ABC superfamily oligopeptide transport system permease protein oppC		0.674	1173	EP
	ABC	binding protein	dipeptide/oligopeptide	3.A.1	ABC superfamily oligopeptide ABC transporter, periplasmic oligopeptide-binding protein oppA		0.735	747	E
LINAB1_0412									
LINAB1_0415	IISP		signal recognition particle-docking protein FtsY	3.A.5	signal recognition particle docking protein FtsY	*	0.639	552	U
LINAB1_0443	IISP		signal recognition particle (SRP54)	3.A.5	Signal recognition particle GTPase		0.682	541	U
LINAB1_0452	ABC	membrane	multidrug	3.A.1	ABC-type multidrug transport system, ATPase		0.647	1132	V
LINAB1_0453	ABC	membrane	multidrug	3.A.1	ABC-type multidrug transport system, ATPase		0.649	1132	V
LINAB1_0698	HCC		hemolysin C (HlyC) homolog	9.A.40	hemolysin		0.655	1253	R
LINAB1_0684	HCC		hemolysin C (HlyC) homolog	9.A.40	hemolysin family protein		0.599	1253	R
LINAB1_0669	SSPTS	EnzymeIIABC	fructose	4.A	PTS family fructose/mannitol porter component IIABC		0.710	1299	G
LINAB1_0659	Oxa1		60 KD inner membrane protein OxaA homolog	2.A.9	preprotein translocase subunit YidC		0.628	706	U
LINAB1_0650	SSPTS	EnzymeIIA	mannose/fructose	4.A	PTS system, mannose/fructose family IIA component		0.658	2893	G
LINAB1_0649	SSPTS	EnzymeIIB	mannose/fructose	4.A	PTS system, mannose/fructose family IIB component		0.674	3444	G
LINAB1_0648	SSPTS	EnzymeIIC	mannose/fructose	4.A	PTS system, mannose/fructose family IIC component		0.691	3715	G
LINAB1_0647	SSPTS	EnzymeIID	mannose/fructose	4.A	PTS system, mannose/fructose family IID component		0.692	3716	G
LINAB1_0646	-	-	-		MgtC family protein		0.621	1285	S
LINAB1_0634	MFS		multidrug efflux	2.A.1	major facilitator superfamily permease	*	0.583	2814	G
LINAB1_0633	ABC	ABC	glycine betaine/L-proline/carnitine/choline	3.A.1	ABC-type proline/glycine betaine transport system, ATPase component		0.683	1125	E
	ABC	binding protein	glycine betaine/L-proline/carnitine/choline	3.A.1	ABC superfamily ATP binding cassette transporter, membrane protein		0.632	1732	M
LINAB1_0632									
LINAB1_0629	P-ATPase		cation	3.A.3	cation-transporting ATPase		0.619	474	P
LINAB1_0625	ABC	ABC	unknown substrate	3.A.1	ABC superfamily ATP binding cassette transporter, ABC protein		0.674	488	R
LINAB1_0624	ABC	ABC	D-methionine	3.A.1	D-Methionine ABC transporter ATP-binding protein		0.668	1135	P
LINAB1_0623	ABC	membrane	D-methionine	3.A.1	D-Methionine ABC transporter permease protein		0.643	2011	P

LINAB1_0622	ThrE		unknown substrate	2.A.79	integral membrane protein		0.618	2966	S
LINAB1_0597	ABC	ABC	multidrug	3.A.1	ABC superfamily ATP binding cassette transporter ABC protein		0.513	1131	V
LINAB1_1048	ABC	membrane	cobalt ion	3.A.1	cobalt/nickel ABC superfamily ATP binding cassette transporter, membrane protein		0.587	619	P
LINAB1_1049	ABC	ABC	cobalt ion	3.A.1	cobalt/nickel ABC superfamily ATP binding cassette transporter, ABC protein		0.632	1122	P
LINAB1_1050	ABC	ABC	cobalt ion	3.A.1	cobalt/nickel ABC superfamily ATP binding cassette transporter, ABC protein		0.638	1122	P
LINAB1_1058	IISP		preprotein translocase Sec61 alpha subunit	3.A.5	preprotein translocase subunit SecY		0.588	201	U
LINAB1_1088	P-ATPase		magnesium ion	3.A.3	magnesium-translocating P-type ATPase		0.597	474	P
LINAB1_1095	ABC	ABC	manganese/zinc ion	3.A.1	ABC transporter, ATP-binding protein		0.601	1121	P
LINAB1_1096	ABC	membrane	polysaccharide export	3.A.1	ABC-type transport system, permease component		0.655	1682	GM
LINAB1_1103	MFS		multidrug efflux	2.A.1	major facilitator superfamily transporter permease		0.627	2271	G
LINAB1_1107	MIP		glycerol uptake	1.A.8	glycerol uptake facilitator related permease (major Intrinsic protein family)		0.597	580	G
LINAB1_1119	ABC	ABC	multidrug	3.A.1	ABC transporter, ATP-binding protein		0.636	4586	R
LINAB1_1120	ABC	membrane	multidrug	3.A.1	partial ABC superfamily ATP binding cassette transporter, permease component		0.540	4587	R
LINAB1_1121	ABC	membrane	multidrug	3.A.1	ABC transporter, permease protein		0.627	3694	R
LINAB1_1125	ABC	membrane	multidrug	3.A.1	ABC superfamily ATP binding cassette transporter		0.591	1132	V
LINAB1_1126	SSPTS	EnzymeIIA	mannose/fructose	4.A	phosphotransferase system, mannose/fructose-specific component IIA		0.591	2893	G
LINAB1_0148	ABC	ABC	phosphate	3.A.1	phosphate ABC transporter, ATP-binding protein		0.614	1117	P
LINAB1_0147	ABC	ABC	phosphate	3.A.1	phosphate ABC transporter ATP-binding protein		0.563	1117	P
LINAB1_0146	ABC	membrane	phosphate	3.A.1	phosphate ABC superfamily ATP binding cassette transporter, membrane protein		0.610	581	P
LINAB1_0145	ABC	membrane	phosphate	3.A.1	phosphate ABC superfamily ATP binding cassette transporter transmembrane protein		0.567	573	P
LINAB1_0144	ABC	binding protein	phosphate	3.A.1	phosphate ABC superfamily ATP binding cassette transporter, binding protein		0.592	226	P
LINAB1_0142	ABC	ABC	efflux (antimicrobial peptide)	3.A.1	ABC transporter, ATP-binding protein		0.584	1136	V
LINAB1_0141	ABC	membrane	lipoprotein releasing	3.A.1	ABC superfamily ATP binding cassette transporter, membrane protein		0.618	4591	M
LINAB1_0139	SSPTS	EnzymeIIA	glucose	4.A	phosphoenolpyruvate-dependent sugar phosphotransferase system EIIA, glucose specific		0.618	2190	G
LINAB1_0136	SSPTS	EnzymeIIBC	glucose/maltose/N-acetylglucosamine	4.A	phosphoenolpyruvate-dependent sugar phosphotransferase system EIIBC, arbutin specific	*	0.647	1263	G
LINAB1_0111	ABC	membrane	multidrug	3.A.1	ABC superfamily ATP binding cassette transporter, predicted lactococcin A/multidrug		0.598	1132	V
LINAB1_0110	ABC	membrane	multidrug	3.A.1	ABC transporter ATPase and permease components		0.557	1132	V
LINAB1_0109	MIP		glycerol uptake	1.A.8	glycerol uptake facilitator protein		0.683	580	G
LINAB1_0080	ABC	membrane	sugar (maltose)	3.A.1	Maltose/maltodextrin ABC transporter, permease protein MalG		0.611	3833	G
LINAB1_0079	ABC	membrane	sugar	3.A.1	Maltose/maltodextrin ABC transporter, permease protein MalF		0.585	1175	G
LINAB1_0078	ABC	binding protein	sugar (maltose)	3.A.1	Maltose/maltodextrin ABC transporter, substrate binding periplasmic protein MalE		0.800	2182	G
LINAB1_0077	ABC	ABC	sugar (maltose)	3.A.1	ABC-type sugar transport system, ATPase component		0.697	3839	G
LINAB1_0069	APC		amino acid	2.A.3	amino acid transporter		0.563	531	E
LINAB1_0064	ABC	ABC	manganese/zinc ion	3.A.1	ABC transporter ATPase component	*	0.586	1121	P
LINAB1_0063	ABC	membrane	toxin secretion	3.A.1	ABC-type multidrug/protein/lipid transport system, SunT-containing, ATPase component	*	0.617	2274	V
LINAB1_0062	APC		amino acid	2.A.3	APC family amino acid-polyamine-organocation transporter		0.626	1113	E
LINAB1_0060	SSPTS	EnzymeIID	mannose/fructose	4.A	PTS system, mannose-specific IID component		0.736	3716	G
LINAB1_0059	SSPTS	EnzymeIIC	mannose/fructose	4.A	PTS system, mannose-specific IIC component		0.679	3715	G
LINAB1_0058	SSPTS	EnzymeIIAB	mannose/fructose	4.A	PTS system, mannose-specific IIA/IIB component		0.706	3444	G
LINAB1_0245	ThrE		unknown substrate	2.A.79	Integral membrane protein		0.603	2966	S
LINAB1_0230	SSPTS	EnzymeIIBC	cellobiose	4.A	PTS system, lactose-specific IIB/IIC component	*	0.639	1455	G

LINAB1_0226	SSPTS	EnzymeIIA	mannose/fructose	4.A	PTS system, mannose-specific IIA component		0.596	2893	G
LINAB1_0225	SSPTS	EnzymeIID	mannose/fructose	4.A	PTS system, mannose-specific IID component		0.624	3716	G
LINAB1_0224	SSPTS	EnzymeIIC	mannose/fructose	4.A	PTS system, mannose-specific IIC component		0.559	3715	G
LINAB1_0223	SSPTS	EnzymeIIB	mannose/fructose	4.A	PTS system, mannose-specific IIB component		0.645	3444	G
LINAB1_0214	GPTS	EnzymeI	unknown substrate	4.A	phosphoenolpyruvate-protein phosphotransferase (enzyme I)		0.699	1080	G
LINAB1_0213	GPTS	HPr	unknown substrate	4.A	Phosphocarrier protein of PTS system HPr		0.715	1925	G
LINAB1_0208	ABC	membrane	sodium ion efflux	3.A.1	putative ABC transporter permease component		0.573	1668	CP
LINAB1_0207	ABC	ABC	multidrug	3.A.1	ABC transporter, ATP-binding subunit, predicted bacitracin/multidrug transport		0.573	1131	V
LINAB1_0888	LIVCS		branched-chain amino acid	2.A.26	LIVCS family branched chain amino acid:cation symporter		0.674	1114	E
	ABC	binding protein	D-methionine	3.A.1	D-methionine ABC superfamily ATP binding cassette transporter, binding protein		0.705	1464	P
LINAB1_0890									
LINAB1_0789	ABC	membrane	lipoprotein releasing	3.A.1	ABC superfamily ATP binding cassette transporter, permease protein		0.602	3127	Q
LINAB1_0790	ABC	ABC	efflux (antimicrobial peptide)	3.A.1	ABC superfamily ATP binding cassette transporter, ABC protein	*	0.594	1136	V
	ABC	binding protein	sugar	3.A.1	sugar ABC superfamily ATP binding cassette transporter, periplasmic component		0.709	1653	G
LINAB1_0793									
LINAB1_0201	ABC	membrane	multidrug/exoprotein (EcsBA homolog)	3.A.1	ABC transporter, permease protein		0.552	4473	U
LINAB1_0200	ABC	ABC	multidrug	3.A.1	ABC-type multidrug transport system, ATPase component		0.670	1131	V
LINAB1_0175	SSPTS	EnzymeIIC	galactitol	4.A	PTS system, galactitol-specific IIC component		0.644	3775	G
LINAB1_0174	SSPTS	EnzymeIIB	galactitol	4.A	PTS system, galactitol-specific IIB component		0.647	3414	G
LINAB1_0173	SSPTS	EnzymeIIA	fructose	4.A	PTS system, galactitol-specific IIA component		0.689	1762	GT
LINAB1_0170	MOP	PST	polysaccharide export	2.A.66	polysaccharide transporter		0.582	2244	R
LINAB1_0168	MFS		multidrug efflux	2.A.1	major facilitator transporter, putative multidrug transport		0.607	2814	G
LINAB1_0268	MFS		multidrug efflux	2.A.1	permease of the major facilitator superfamily		0.429	-	-
LINAB1_0828	CadD		cadmium ion	2.A.77	Cadmium resistance protein	*	0.483	4300	P
LINAB1_0823	ABC	ABC	unknown substrate	3.A.1	ABC superfamily ATP binding cassette transporter, ABC protein	*	0.538	488	R
LINAB1_0813	-	-	-	-	Major facilitator superfamily protein		0.534	-	-
LINAB1_0024	-	-	-	-	major facilitator superfamily permease		0.571	-	-
LINAB1_1164	-	-	-	-	ABC transporter permease component		0.557	619	P
LINAB1_1138	-	-	-	-	ABC transporter membrane-spanning permease - macrolide efflux		0.571	-	-
LINAB1_1043	-	-	-	-	Multidrug resistance transport protein		0.613	2076	P
LINAB1_1025	-	-	-	-	preprotein translocase subunit SecE		0.679	690	U

Table S6. *L. iners* ORFs with significant hits to conserved adhesion/host interaction domains adapted from Kankainen, et al. 2009.

Type	Pfam ID	Gene ID	e-value
General adhesion	PF00028	-	-
	PF01468	-	-
	PF01473	-	-
	PF01856	-	-
	PF02216	-	-
	PF03351	-	-
	PF03642	-	-
	PF03895	-	-
	PF04122	-	-
	PF04829	-	-
	PF04830	-	-
	PF05062	-	-
	PF05321	-	-
	PF05345	LINAB1_0795	0.0003
	PF05594	-	-
	PF05860	-	-
	PF06435	-	-
	PF06443	-	-
	PF07501	-	-
	PF07675	-	-
PF07705	-	-	
PF08374	-	-	
PF08964	-	-	
Collagen	PF00039	-	-
	PF00040	-	-
	PF00041	-	-
	PF05737	-	-
	PF05738	-	-
	PF01391	-	-
	PF02352	-	-
Epithelium	PF03212	-	-
	PF05594	-	-
	PF05860	-	-
	PF06744	-	-
	PF06761	-	-
Fibrin	PF09131	-	-
	PF08017	-	-

Fibronectin	PF02412	-	-	
	PF02707	-	-	
	PF02722	-	-	
	PF02986	-	-	
	PF05670	LINAB1_0564	1.10E-21	
	PF05833	LINAB1_0564	1.40E-119	
	PF07012	-	-	
	PF07174	-	-	
	PF07299	-	-	
	PF07554	-	-	
	PF07564	-	-	
	Flagella	PF02465	-	-
		PF07195	-	-
PF05211		-	-	
Ig	PF00047	-	-	
	PF02246	-	-	
	PF02494	-	-	
	PF02368	-	-	
	PF02369	-	-	
	PF07523	-	-	
	PF07532	-	-	
	PF07679	-	-	
	PF07686	-	-	
	Intimin	PF02369	-	-
PF02368		-	-	
PF03549		-	-	
PF07489		-	-	
PF07490		-	-	
PF07979		-	-	
Invasin	PF05658	-	-	
	PF05662	-	-	
	PF05775	-	-	
	PF09134	-	-	
Lamin	PF02210	-	-	
	PF00054	-	-	
Lectin	PF00059	-	-	
	PF00139	-	-	
	PF00652	-	-	
	PF01419	-	-	
	PF01453	-	-	
	PF07472	-	-	
	PF07828	-	-	
PF07938	-	-		

	PF07979	-	-
	PF09160	-	-
	PF09222	-	-
	PF09112	-	-
	PF09113	-	-
	PF09264	-	-
	PF09458	-	-
Mucus	PF02465	-	-
	PF04225	-	-
	PF06255	-	-
	PF06458	-	-
	PF07195	-	-
	PF08525	-	-
	PF09085	-	-
Peptidoglycan	PF01471	LINAB1_0548	4.80E-14
	PF01476	-	-
	PF03562	-	-
	PF08823	-	-
Pili	PF09374	-	-
	PF00114	-	-
	PF00419	-	-
	PF02432	-	-
	PF03627	-	-
	PF04449	-	-
	PF04619	-	-
	PF04956	-	-
	PF04964	-	-
	PF05211	-	-
	PF05307	-	-
	PF05567	-	-
	PF05946	-	-
	PF06321	-	-
	PF06864	-	-
	PF07419	-	-
	PF07963	LINAB1_0374	0.0004
		LINAB1_0373	4.30E-07
	PF08805	-	-
	PF09160	-	-
	PF09222	-	-
Other	PF10425	LINAB1_0798	1.90E-14
	PF08428	LINAB1_0273	5.80E-14
		LINAB1_0273	5.40E-06
		LINAB1_0273	1.60E-08

	LINAB1_0796	0.0002
	LINAB1_0796	0.00049
	LINAB1_0796	0.00031
	LINAB1_0796	0.00061
	LINAB1_0796	0.00083
	LINAB1_0796	2.30E-07
	LINAB1_0796	5.70E-05
	LINAB1_0796	3.00E-09
	LINAB1_0796	7.50E-06
	LINAB1_0796	5.40E-08
	LINAB1_0796	4.80E-06
	LINAB1_0796	4.50E-08
PF00746	LINAB1_0795	5.20E-06
	LINAB1_0796	9.70E-05
	LINAB1_0800	4.30E-07
	LINAB1_0370	1.00E-05
	LINAB1_0273	3.60E-05
	LINAB1_0701	3.60E-05
PF04650	LINAB1_0272	5.90E-07
	LINAB1_0701	2.20E-07
	LINAB1_0795	2.10E-06
	LINAB1_0796	8.80E-07
	LINAB1_0368	6.40E-10
	LINAB1_0800	6.80E-09
	LINAB1_0950	2.00E-11