Title	Identification of putative adhesins and carbohydrate ligands of Lactobacillus paracasei using a combinatorial in silico and glycomics microarray profiling approach
Author(s)	Houeix, Benoit; Synowsky, Silvia; Cairns, Michael T.; Kane, Marian; Kilcoyne, Michelle; Joshi, Lokesh
Publication Date	2019-11-11
Publication Information	Houeix, Benoit, Synowsky, Silvia, Cairns, Michael T, Kane, Marian, Kilcoyne, Michelle, & Joshi, Lokesh. (2019). Identification of putative adhesins and carbohydrate ligands of Lactobacillus paracasei using a combinatorial in silico and glycomics microarray profiling approach. Integrative Biology, 11(7), 315-329. doi: 10.1093/intbio/zyz026
Publisher	Oxford University Press (OUP)
Link to publisher's version	https://doi.org/10.1093/intbio/zyz026
Item record	http://hdl.handle.net/10379/15666
DOI	http://dx.doi.org/10.1093/intbio/zyz026

Identification of putative adhesins and carbohydrate ligands of *Lactobacillus paracasei* using a combinatorial *in silico* and glycomics microarray profiling approach

Benoit Houeix,[a,b] Silvia Synowsky,[c] Michael T. Cairns,[a,b] Marian Kane,[a,b] Michelle Kilcoyne,[b,d,#] Lokesh Joshi[a,b,#]

[a]Glycoscience Group, National Centre for Biomedical Engineering Science, National University of Ireland Galway, Galway, Ireland

[b]Advanced Glycoscience Research Cluster, National Centre for Biomedical Engineering Science, National University of Ireland Galway, Galway, Ireland

[c] Biomedical Sciences Research Complex, University of St. Andrews, St. Andrews, KY16 9ST, UK

[d]Carbohydrate Signalling Group, Discipline of Microbiology, School of Natural Sciences, National University of Ireland Galway, Galway, Ireland

Running Title: *Lactobacillus paracasei* adhesins and carbohydrate ligands

[#]Address correspondence to Lokesh Joshi, lokesh.joshi@nuigalway.ie

Michelle Kilcoyne, michelle.kilcoyne@nuigalway.ie

† Electronic supplementary information available.

Insight Box

Commensal bacteria colonise host gastrointestinal tract mucus *via* surface lectin-like adhesins to impart health benefits but little is known about their adhesins and associated carbohydrate ligands. Novel adhesins from the important human commensal *Lactobacillus paracasei* subsp. *paracasei* ATCC 25302 were identified using *in silico* and *in vitro* expression analysis. Glycomics microarray profiling indicated common and specific ecological niches for *L. paracasei*, *L. rhamnosus* GG and *L. johnsonii* and revealed that *L. paracasei* differentially recognised the blood group A and B and Lewis antigens compared to the other two species. This multidisciplinary approach is promising for furthering knowledge of bacterial-host cross-talk.

ABSTRACT

Commensal bacteria must colonise host mucosal surfaces to exert health-promoting properties, and bind to gastrointestinal tract (GIT) mucins *via* their cell surface adhesins. Considerable effort has been directed towards discovery of pathogen adhesins and their ligands to develop anti-infective strategies, however, little is known about the lectin-like adhesins and associated carbohydrate ligands in commensals. In this study an *in silico* approach was used to detect surface exposed adhesins in the human commensal *Lactobacillus paracasei* subsp. *paracasei*, a promising probiotic commonly used in dairy product fermentation that presents anti-microbial activity. Of the 13 adhesin candidates, 3 sortase-dependent pili clusters were identified in this strain and expression of the adhesin candidate genes was confirmed *in vitro*. Mass spectrometry analysis confirmed the presence of surface adhesins elongation factor Tu and the chaperonin GroEL, but not pili expression. Whole cells were subsequently incubated on microarrays featuring a panel of GIT mucins from nine different species and two human-derived cell lines and a library of carbohydrate structures. Binding profiles were compared to those of two known pili-producing lactobacilli, *L. johnsonii* and *L. rhamnosus* and displayed overlapping but distinct signatures which may indicate different abilities for regiospecific GIT colonisation. In addition, *L. paracasei* whole cells favoured binding to α-(2→3)-linked sialic acid and α-(1→2)-linked fucose-containing carbohydrate structures including blood groups A, B and O and Lewis antigens x, y and b. This study furthers our understanding of host-commensal cross-talk by identifying potential adhesins, and specific GIT mucin and carbohydrate ligands and provides insight into the selection of colonisation sites by commensals in the GIT.

1. Introduction

Disruptions to the normal balance between the gut microbiota and the host have been linked with various diseases such as inflammatory bowel disease (IBD), metabolic syndromes, colonic cancers and neurological disorders.[1,2] The gastrointestinal tract (GIT) forms a dynamic ecosystem which includes a mucus layer populated by commensal microbes, nutrients and opportunistic microbes.[3] The mucus layer is mainly comprised of heavily glycosylated proteins, or mucins, and the mucus layer is especially thick in the colon which contains the highest populations of commensal bacteria. Mucin glycosylation varies throughout the human GIT, decreasing in fucose content in a gradient but increasing in sialic acid content from stomach to colon.[4] Specific carbohydrate structures are associated with particular GIT niches, species, tissues or health status of the host,[5] and altered mucin glycosylation has been associated with disease development, including ulcerative colitis.[6] Carbohydrate structures present on mucins are likely ligands for commensal bacterial binding, although there is limited literature on commensal adhesins with confirmed carbohydrate specificity, and mucin glycosylation is thought to be important in influencing gut microbiome community distributions.[6,7] Commensal bacteria have also been shown to trigger glycosylation changes in the intestinal mucosa by interfering with the expression of host glycosyltransferases and glycosidases to produce structures which favour bacterial colonisation or metabolism.[6,8] Considering the interdependent relationship of the host and gut microbiome, the use of normally commensal bacteria are increasingly being explored as potential therapeutics for restoring or maintaining gut health and homeostasis and combating pathogens by competition or inhibition.[9-11]

Lactic acid bacteria are some of the most widely used commensals for their probiotic effects. *Lactobacillus paracasei* subsp. *paracasei* is known to have beneficial effects for the host including prevention of pathogen-associated intestinal injury, favourable immunomodulating

effects such as enhanced antibody production and lymphocyte proliferation, and modification of intestinal microbiota population towards a more beneficial distribution.[9-11] It also has a broad inhibitory activity against pathogens including *Staphylococcus aureus*, *Escherichia coli*, *Salmonella enterica*, and *Klebsiella pneumoniae*.[9] *L. paracasei* subsp. *paracasei* is closely related to the most studied human probiotic strain *L. rhamnosus* GG, which has a human mucus-binding protein (MUB, also known as MubBP) on pili with a role in adherence to intestinal epithelial cells, biofilm formation and immunomodulation.[12] However, the presence or identity of potential adhesins or pili with carbohydrate-binding (lectin) functionality in *L. paracasei* subsp. *paracasei* is currently unknown. Greater localisation of *L. paracasei* subsp. *paracasei* in the colon of gnotobiotic mice has been shown with regiospecific metabolic effects.[13] The *in vivo* localisation in the human or other mammalian GIT is not known, but it has been found in the newborn human infant gut[14] and from age one year and in the human adult mouth and large colon,[15] though specific localisation may be different between adults and infants as variations in the gut microbiome occur with weaning and maturation.[16] As binding to mucus is associated with persistence in the GIT,[17] identifying potential *L. paracasei* subsp. *paracasei* adhesins, their specificity of carbohydrate target ligand(s) and any potential mammalian GIT regioselectivity could lend insight into the successful colonisation strategies of particular regions of the GIT, persistence and mechanism(s) of any potential therapeutic effect.

Bacterial adhesion has been studied extensively in pathogens but very few commensal adhesins have been described to date. Bacteria usually attach to molecules on GIT surfaces, within the mucus layer or in the extracellular matrix (ECM) using fimbriae or pili, which contain adhesins and/or bacterial cell surface proteins, and many of these adhesins are lectins.[18,19] The structures of pili differ between Gram-negative and Gram-positive bacteria.[20] In Gram-positive bacteria, sortases catalyse the assembly of pilin subunits into pili

appendages and the anchoring of pili to the cell wall,[21] while cell wall proteins involved in adhesion include transmembrane proteins, lipoproteins and peptidoglycan anchored molecules. Recent reports have also implicated various commensal surface-exposed proteins (SEPs) in adhesion to epithelial cells,[22] ECM proteins[23] and mucins.[24] A number of MUBs with lectin function have been identified in lactobacilli,[25] including sortase-dependent pilin subunits, such as SpaC from *L. rhamnosus* GG,[26] and afimbrial large surface proteins with highly repetitive motifs involved in mucus adhesion, such as mucin-binding domains from *L. rhamnosus* GG[27] and *L. reuteri*.[7] Although pili have been found in *L rhamnosus* GG[26] and *L. ruminis* ATCC 25644, and pilus-like gene clusters identified in *L. johnsonii* NCC 533, *L. casei* and *L. paracasei*,[28] the identification of pili in commensal species is infrequent compared to pathogens. Despite its close relationship to *L. rhamnosus* GG, the presence of pili or pilus-like gene clusters in *L. paracasei* subsp. *paracasei* is currently not known.

The majority of strict anaerobes populating the microbiota are recalcitrant to conventional culture methods therefore many current approaches to identifying putative bacterial adhesion molecules rely on *in silico* methods, which assume that conserved distinct amino acid sequences share similar roles.[29-31] In addition, software such as SurfG+ has been developed to predict the cellular localisation of a given protein using signal motifs for Gram-positive bacteria, which was used to demonstrate that *L. johnsonii* and *L. gasseri* potentially encoded a larger variety of SEPs compared to the yoghurt bacterium *L. delbruekii* ssp. *bulgaricus*.[31] Predicted SEPs can then be further analysed using conserved domains searches or alignments to determine homologies with characterised adhesins.

Specific carbohydrate ligands for commensal adhesins or whole bacteria have only been rarely identified beyond a characterised general binding to epithelial cells, mucins or mucus.[17,26,27] For example, the mucin-feeding human GIT resident and potential probiotic *Akkermansia muciniphilia*[32] was shown to have increased binding to colonic mucin from

ulcerative colitis patients compared to healthy samples on a mucin microarray platform.[33] Within *Lactobacillus*, a cell surface mannose-specific adhesion (Msa) of *L. plantarum* was identified.[34] In addition, surface layer proteins extracted from eleven strains of lactobacilli bound to human blood group A and B trisaccharides, structures which are present in the intestinal mucosa, but this was done with a pool of proteins.[35] *L. johnsonii* NCC533 (La1) whole bacteria was shown to bind to asialo-GM1, a structure found on glycolipids,[36] and the elongation factor Tu (EF-Tu) bound to normal human colonic mucins at pH 5.0.[37] However, specific mucin or carbohydrate ligands for *L. paracasei* subsp. *paracasei* have not been identified to date.

In this interdisciplinary study, an *in silico* approach was taken to identify putative adhesins from *L. paracasei* subsp. *paracasei* ATCC 25302. Putative adhesin expression was evaluated by real-time quantitative polymerase chain reaction (RT-qPCR), transmission electron microscopy (TEM) and mass spectrometric protein identification. Finally, whole cells of *L. paracasei* subsp. *paracasei*, *L rhamnosus* GG and *L. johnsonii* were profiled on microarrays featuring mucins from the GIT of nine different mammalian species and two human derived cell lines and a library of carbohydrate structures to elucidate specific carbohydrate ligands and potential GIT regioselectivity.

2. Materials and methods

2.1. Materials

Bicinchoninic acid assay (BCA) kit, SYTO® 82 Orange Nucleic Acid stain, and SnakeSkin 3.5 kDa molecular weight cut off (MWCO) dialysis tubing were purchased from Thermo Fisher Scientific (Waltham, MA, U.S.A). SuperScript VILO Reverse Transcription Kit, 3-(N-

morpholino)propanesulfonic acid (MOPS) buffer and sodium dodecyl sulphate-(SDS-)polyacrylamide (PAGE) gels (NuPAGE® 4-12% Bis-Tris gels) were from Life Technologies (Carlsbad, CA, U.S.A.). RNeasy, RNA isolation kits, RNase-free DNase and QuantiTect SYBR® Green PCR kits were obtained from Qiagen (Manchester, U.K.). Nexterion® slide H microarray slides were from Schott AB (Mainz, Germany). Neoglycoconjugates were obtained from IsoSep AB (Tullinge, Sweden), Dextra Laboratories Ltd. (Reading, U.K.). Tetramethylrhodamine-(TRITC-)labelled lectins (Table 1) were from EY Labs, Inc. (San Mateo, CA, U.S.A). De Man, Rogosa and Sharpe agar (MRS) was from Oxoid (Basingstoke, U.K.). GIT mucus from various GIT locations and mammalian species and cell lines were obtained from the Veterinary Science Centre in University College Dublin licensed by the Department of Health and Children, Ireland, in accordance with the Cruelty to Animals Act (Ireland 1897) and the European Community Directive 86/609/EC, and were sanctioned by the Animals Research Ethics Committee, University College Dublin, Ireland. Mucins were purified from collected mucus as previously described.[38] All other reagents were from Sigma-Aldrich Co. (Dublin, Ireland) unless otherwise indicated and were of the highest grade available.

2.2. *In silico* analysis for surface-exposed protein (SEP) localisation

The proteome from *L. paracasei* subsp. *paracasei* ATCC® 25302 (GenBank accession NZ_ACGY00000000), abbreviated throughout as *L. paracasei*, was analysed using the SurfG+ programme essentially as previously described.[31] In brief, the protein in FASTA format was tested sequentially for the presence of a transmembrane helix predictor sequence (TMMOD), a secretion signal predictor sequence (SignalP), a lipoprotein signal predictor sequence (LipoP) and a sequence alignment for protein profiles (HMMER). The proteins were classified into their predicted cell compartments or whether secreted, cytoplasmic, cell wall localised or exposed on the bacterial surface (SEPs) (Fig. S1, ESI†). The predicted SEPs

(Table S1, ESI†) were further analysed using BLASTp (http://www.ncbi.nlm.nih.gov) to find any homologues to previously described adhesins or known motifs found on adhesins using the Conserved Domains Database (CDD, http://www.ncbi.nlm.nih.gov). A selection of candidate adhesins were compiled for expression studies (Table 2). In addition, the ExPASy bioinformatics portal was used to calculate the molecular mass and isoelectric point (pI) of the protein sequences (http://web.expasy.org/compute_pi) and the Phyre2 suite (http://www.sbg.bio.ic.ac.uk/phyre2) used to model the structures of selected pilin subunits.[39] PilB was successfully modelled based on template c3kptA_ (100% confidence, 30% coverage) corresponding to the protein gbs052 and built on the crystal structure of SpaB, the minor pilin in the Gram-positive pathogen *Streptococcus agalactiae.* SpaC_vWF was modelled on the adhesive tip pilin gbs104 from *S. agalactiae* (c3txaA_, 100% confidence, 24% coverage). Finally, SpaF was modelled on a collagen adhesion protein (template c2pz4A_, 99.9% confidence, 17% coverage), corresponding to the crystal structure of BcpA, the major pilin subunit of *Bacillus cereus.*

2.3. Bacterial culture and growth curves

Type strains of lactobacilli were obtained either from the Leibniz Institute DSMZ-German Collection of Microorganisms and Cell Cultures, Braunschweig, Germany (*L. paracasei* subsp. *paracasei* ATCC 25302/DSM 5622, hereafter referred to as *L. paracasei*) or from the BCCM/LMG Bacteria Collection, Ghent University, Belgium (*L. rhamnosus* GG ATCC 53103/LMG 18243 and *L. johnsonii* ATCC 33200/LMG 9436). Bacteria were grown anaerobically in sealed BD GasPak™ EZ Anaerobe Systems pouches (Becton, Dickinson and Co., Franklin Lakes, NJ, U.S.A.) on MRS agar plates for 48 h at 37 °C. Liquid cultures were grown in MRS broth using a modification of the roll tube method for anaerobic culturing[40] in sterile syringes supplemented with 0.05% (w/v) L-cysteine. For this purpose, a single colony was transferred to a 1 mL syringe containing MRS broth and incubated overnight at 37 ˚C. A

5% inoculum was subsequently transferred to a 20 mL syringe, incubated in similar conditions and transferred to larger volumes (50 mL syringe and 1 L glass bottle).

For growth curves, bacteria were grown as described above on agar plates and a single colony was transferred to a 1 mL syringe and incubated overnight at 37 ˚C. A 5% inoculum was transferred to MRS broth in a 20 mL syringe and incubated overnight at 37 ˚C. A 5% inoculum was then used for 50 mL syringes containing regular MRS media (MRS), MRS supplemented with 0.22 µm sterile-filtered 0.1% porcine gastric mucin (PGM) (MRS + PGM), MRS supplemented with 0.3 M NaCl (MRS + NaCl) or both 0.1% PGM and 0.3 M NaCl (MRS +PGM/NaCl). An aliquot of 2 mL of culture was removed every 2 h and absorbance was measured at 600 nm (Biophotometer, Eppendorf, Germany) (Fig. S2A, ESI†). pH was measured at the same time points (Symphony, VWR, Ireland) (Fig. S2B, ESI†).

2.4. Bacterial RNA isolation

The RNA extraction procedure was adapted from Karatzas, *et al*.[41] Three biological replicates of bacteria were cultured in broth in four different media, MRS, MRS + PGM, MRS + NaCl and MRS + PGM/NaCl, as described above. Bacteria were harvested in lag (3 h, MRS only) or log (8 h, MRS, MRS, MRS + PGM, MRS + NaCl and MRS + PGM/NaCl) growth phases by centrifugation to generate a total of fifteen samples. Bacteria were washed twice in phosphate buffered saline, pH 7.4 (PBS) at 4 ˚C for 10 min and pelleted by centrifugation at 5,000 x *g*. Bacterial pellets were subsequently incubated on ice in 5% phenol/ 95% ethanol (v/v), for 15 min and then centrifuged twice at 4 ˚C; first for 10 min at 5,000 x *g* followed by 3 min at 9,000 x *g*. Pellets were immediately stored at –80 ˚C until extraction.

For RNA extraction, the frozen cell pellets were mechanically disrupted using a 3 min homogenisation cycle in a Mini Beadbeater™ (Biospec Products, Bartlesville, Oklahoma,

U.S.A.) which was repeated twice, with the samples incubated on ice for 5 min between cycles. Total RNA isolations were carried out on the disrupted cells using an RNeasy mini kit following the manufacturer's instructions, including an on-column DNase I digestion step using the RNase-Free DNase Set. Total RNA from each biological replicate for the five conditions investigated was quantified and the quality estimated using the ratio of absorbance at 260/280 nm measured by NanoDrop spectrophotometer (Thermo Fisher Scientific, Wilmington, Delaware, U.S.A). The RNA purity was further assessed using the RNA integration number (R.I.N) determined using the RNA6000 Nano Kit on a 2100 Bioanalyser (Agilent Technologies, CA, U.S.A) following the manufacturer's protocol. All samples satisfied the threshold applied (R.I.N > 7, Table S2, ESI†).

2.5. Quantitative real time PCR (RT-qPCR)

Total RNA (2 µg) from all fifteen samples described above was reverse transcribed using the SuperScript VILO kit according to manufacturer's instructions. Oligonucleotide primers for 13 candidate genes were designed using Primer3Plus software (http://primer3plus.com/cgi-bin/dev/primer3plus.cgi ; Table 3). RT-qPCR was performed in triplicate on three independent biological samples with the 16S rRNA reference gene on each plate. Control reactions were carried out to confirm the absence of genomic DNA contamination (RT negative) and the similar level of the reference gene across conditions (Table S3, ESI†). The two step RT-qPCR method was carried out using the QuantiTect SYBR® Green PCR kit in a Stratagene Mx3000p RT-qPCR thermal cycler (Agilent Technologies). Amplification was carried out in a 20 µL total reaction volume containing 2X Quantitect SYBR green PCR master mix, 500 nM sense and anti-sense primers, and 5 µL of the cDNA template diluted 10 times. The thermal profile used was 95 ˚C for 15 min, followed by 40 cycles of 94 ˚C for 15 s, 60 ˚C for 30 s and 72 ˚C for 30 s. The specificity of the amplicon was confirmed by the presence of a single peak in DNA melting curves. Data were normalised to one of the most

commonly used reference gene (16S rRNA)[42] and fold changes between conditions calculated following the 'delta delta Ct' method for relative quantification of gene expression[43] with the lag phase used as the calibrator. Statistical tests were subsequently carried out in the form of analysis of variance (ANOVA) and *post hoc* Tukey test (SPSS statistics 20, IBM, Ireland) between the four conditions (MRS, MRS + PGM, MRS + NaCl and MRS + PGM/NaCl) for the whole set of adhesin candidates selected from the *in silico* analysis or between conditions for each putative adhesin.

2.6. Transmission electron microscopy (TEM)

Bacteria grown in MRS broth were harvested at mid-exponential phase (8 h, Fig. S2A, ESI†) by centrifugation, washed twice in sterile Tris buffered saline (TBS; 20 mM Tris-HCl, 100 mM NaCl, pH 7.2), pelleted by centrifugation (5,000 x *g*, 2 min, 4 °C) and adjusted to an absorbance at 600 nm of 1.0 in TBS (approximately 1 x 10^8 cfu/mL). A formvar carbon grid (Agar Scientific, U.K.) was placed shiny side towards a drop of 10 µl of cells and left for 1 min to mount. Excess liquid was subsequently removed using silk paper and the grid mounted on a drop of 10 µL negative staining solution phosphotungstic acid (1%). Staining was carried out for 1 min, excess liquid removed using silk paper and the grid air-dried prior to visualisation. Samples were viewed and photographed using a transmission electron microscope (Hitachi H-7500, Hitachi High-Technologies Corp., Tokyo, Japan).

2.7. Bacterial surface-exposed proteins (SEPs) enrichment and electrophoretic separation

Bacterial cell SEPs were extracted using a modification of the method of Rojas, *et al.*[44] Bacterial cells grown in 50 mL broth overnight were centrifuged at 12,000 × *g* for 4 min at 4 ˚C. The pellets were washed once with PBS and centrifuged again. Pellets were then resuspended in 10 mL of 1 M LiCl in PBS, incubated on ice with gentle rotation (4 rpm) for 4

h and centrifuged again. The supernatant was filtered through a 0.22 μm membrane and digested for 3 h at 37 ˚C with DNase I and RNase A, both at final concentrations of 0.1 mg/mL. The digested supernatant was then dialysed overnight at 4 ˚C against deionised water using a 3.5 kDa MWCO membrane, frozen and then lyophilised to dryness. Protein concentration was estimated using the BCA assay with a bovine serum albumin (BSA) standard and aliquots were stored at -20 °C until use.

SEPs (50 μg) were electrophoresed on a pre-cast NuPAGE® 4-12% Bis-Tris gels in reducing conditions in MOPS buffer at 150 V for 1 h. Gels were stained overnight with 0.25% Coomassie Brilliant Blue G-250 in 30% ethanol and 10% glacial acetic acid and destained with distilled water. Stained gels were then imaged on a gel documentation system using white light (Uvidoc HD6, UVITEC Ltd., Cambridge, U.K.) and images were stored as .tif files.

2.8. Protein identification by mass spectrometry

Selected gel bands and gel chunks were cut into 1 mm cubes and subjected to in-gel digestion using a ProGest Investigator in-gel digestion robot (Genomic Solutions, Ann Arbor, MI, U.S.A.) using standard protocols.[45] Briefly, gel cubes were destained by washing with 50% acetonitrile (ACN) and subjected to reduction and alkylation before digestion with trypsin at 37 °C. The peptides were extracted with 5% formic acid and concentrated to 20 μL by centrifugal evaporation (ThermoSavant SpeedVac, ThermoFisher Scientific). A portion of the resultant peptides were then injected on an Acclaim PepMap 100 C18 trap and an Acclaim PepMap RSLC C18 column (ThermoFisher Scientific), using a nanoLC Ultra 2D plus loading pump and nanoLC as-2 autosampler (Eksigent). For the gel bands (samples Fpr57, Lpp35, Lpp102), peptides were loaded onto the trap column for 5 min while for gel chunks (sample Lpp48 and LppW) samples were loaded for 10 min at a flow rate of 5 ul/min of

loading buffer (98% water/2% ACN/ 0.05% trifluoroacetic acid). Trap column was then switched in line with the analytical column and peptides were eluted with a gradient of increasing ACN, containing 0.1 % formic acid at 300 nl/min. Peptides from gel bands were eluted using the following gradient: 2-40% ACN in 6 min, 40-98% in a further 3 min, followed by 98% ACN to clean the column before re-equilibration to 2% ACN. For the gel chunks, peptides were eluted using the following gradient: 2-40% ACN in 45 min, then 40-98% in 4 min followed by 98% ACN to clean the analytical column before re-equilibration to 2% ACN. The eluate was sprayed into a TripleTOF 5600+ electrospray tandem mass spectrometer (AB Sciex Pte. Ltd., Foster City, CA, U.S.A.) and analysed in Information Dependent Acquisition (IDA) mode, performing 120 ms of MS followed by 80 ms MS/MS analyses on the 20 most intense peaks seen by MS.

The MS/MS data file generated *via* the 'Create mgf file' script in PeakView (Sciex) was analysed using the Mascot search algorithm (Matrix Science Inc., Boston, MA, U.S.A.), against the NCBInr database (accessed Aug 2016) all species (93,482,448 sequences), using trypsin as the cleavage enzyme and carbamidomethylation as a fixed modification of cysteine and methionine oxidation as a variable modifications. The peptide mass tolerance was set to 20 ppm and the MS/MS mass tolerance to ± 0.1 Da. Additionally MS/MS data was searched against an in-house database after adding the protein sequences of interest (total of 4,122 sequences were added including the 13 protein sequences in Table 2). A protein was accepted as identified if it had two or more peptides with Mascot Ion Scores above the Identity ($p<0.05$) Threshold and, for those proteins identified by only two peptides, the MS/MS spectral assignments had to match most of the peaks in the MS/MS spectra.

2.9. Fluorescence labelling of whole bacteria

Three biological replicates of *L. paracasei*, *L. johnsonii* and *L. rhamnosus* were cultured as described above in 20 mL broth MRS to mid-exponential phase (approximately 8 h). Bacterial pellets were stained with SYTO® 82 as previously described.[46] The fluorescence intensity from dye uptake was titrated from 5 to 100 µM (excitation 541 nm, emission 560 nm) in a 96 well black microtitre plate using a Spectramax M5e spectrophotometer (Molecular Devices, CA, U.S.A). The maximum fluorescence was obtained using 10 µM of dye each for *L. johnsonii* and *L. rhamnosus* and 20 µM for *L. paracasei* and these optimal concentrations were used throughout. Following staining and washing three times in TBS supplemented with Ca^{2+} and Mg^{2+} ions (TBSS; 20 mM Tris-HCl, 100 mM NaCl, 1 mM $CaCl_2$, 1 mM $MgCl_2$, pH 7.2), bacterial pellets were resuspended in TBSS supplemented with 0.05% Tween 20 (TBSS-T) and the absorbance at 600 nm was adjusted to 2.0 with TBSS-T prior to incubation on microarrays.

2.10. Carbohydrate and GIT mucin microarray construction and bacterial profiling

Carbohydrate and GIT mucin microarrays of eight replicate subarrays were constructed essentially as described previously.[38,47] The GIT mucin microarray contained 44 features, each of six replicates, consisting of GIT mucins from nine animal species, two human cell lines and reference glycoproteins (Table S4, ESI†). The carbohydrate microarray contained 76 neoglycoconjugates (NGCs) and reference glycoproteins, each of 6 replicates, divided between twinned A and B slides (Tables S5 and S6, ESI†) with 20 common features in the same position between slides to facilitate later data normalisation.[47] Fluorescently labelled bacterial cells were incubated on the microarray slides essentially as previously described.[46,47] Initial titrations for the labelled bacteria were carried out to determine optimal signal to noise ratio and the optimal dilution of 40 µL of labelled bacterial cells (of absorbance at 600 nm of 2.0 stock) in a final volume of 70 µL TBSS-T was used for all microarray experiments. Two separate controls (TRITC-labelled lectins at 1 µg/mL - AIA and MAA for the carbohydrate

microarrays, UEA-I and WFA for the mucin microarray) to verify printing were included in two subarrays of each slide. Samples were incubated on the microarray slides for 1 h at 23 °C with gentle rotation (4 rpm). Experiments were carried out in biological triplicates. Following incubation, slides were washed six times using an automatic microarray wash station (Implen, Advawash, Germany) in TBSS-T for 2 min each wash followed by one wash in TBSS. The microarray slides were centrifuged dry (350 × *g*, 5 min, 10 °C) and scanned immediately in an Agilent G2505B microarray scanner (Agilent Technologies, California, U.S.A) using the green channel (532 nm, 90% PMT, 5 μm resolution). Images were saved as *.tif files for data extraction.

2.11. Microarray data extraction and analysis

Signal intensities were extracted using GenePix v6.1.0.4 (Molecular Devices, Sunnyvale, CA, U.S.A) essentially as previously described.[38,46,47] For each feature, intensity was taken as the median of the six replicate intensities of each feature after background subtraction and was handled as a single data point for graphical and statistical analysis. For mucin microarrays, data intensities were normalised to the per-subarray (i.e. same sample) total intensity mean of the three biological replicate slides and binding intensity data were presented as bar charts of the mean with error bars of one standard deviation.[38] For carbohydrate microarrays, data from the twinned panels and three replicates were normalised across all six microarray slides (three replicates of twinned microarrays) to the per-subarray total intensity mean of the 20 overlapping probes.[47] Unsupervised clustering of microarray data was performed using Hierarchical Clustering Explorer v3.0 (HCE 3.0, University of Maryland, College Park, U.S.A, http://www.cs.umd.edu/hcil/hce/hce3.html). Normalised data were imported into HCE 3.0 for hierarchical clustering by Euclidean distances with complete linkage. Statistical tests

were carried out by analysis of variance (ANOVA) and *post hoc* Tukey test (SPSS statistics 20, IBM, Ireland) between the three species for all features.

3. Results

3.1. *In silico* prediction of adhesion candidates from *Lactobacillus paracasei* subsp. *paracasei*

The proteome of *L. paracasei* was analysed using SurfG+, a predictive software suite for SEP localisation, and proteins were classified according to the predicted cell compartment localisation (Fig. 1A). The majority of the proteins analysed were predicted to be either cytoplasmic (2,022 out of 2,789 sequences, 72.5% of the proteome) or membrane-bound (454 sequences, 16.3%). A small proportion was predicted to be secreted proteins (100, 3.6%). The predicted SEPs represented 7.6% (213 sequences, Table S1, ESI†) of the total proteome. The topology of the predicted SEPs was modelled (Fig. 1B) with the majority of SEPs predicted to have a protruding carboxy terminal (Ct, 41.8%). Other classifications included proteins exposing an amino terminal (Nt, 23.9%), surface-exposed loops (14.6%), lipoproteins (9.9%) or possessing cell wall retention signals (9.9%). The majority of predicted SEPs were 251-500 amino acids (42.3%), with some containing less (25.4%) and others 501-1,000 amino acids (29.1%) or more (3.5%) (Fig. 1C).

BLAST analysis was carried out on the *L. paracasei* proteome to search for homology to 16 commensal *Lactobacillus* sequences which are functionally characterised as adhesins (Table S7, ESI†). Four proteins were homologous to known lactobacilli adhesins, namely the translation factor EF_Tu, the chaperonin GroEL, the fibronectin-binding protein A (FbpA) and the collagen-binding protein (CnBP). The 213 SEP sequences (Table S1, ESI†) predicted

by SurfG+ were further investigated for the presence of conserved motifs found in known pilin subunits or in bacterial proteins that bind to mucus (MUB) or ECM components. Nine proteins were added to complete a list of 13 selected adhesin candidates from *L. paracasei*, five of which were homologous to proteins from the related strain in the *L. casei* group, *L. rhamnosus* GG (Table 2). The candidates included two large proteins of 1,069 amino acids (EEI68918) and 1,173 amino acids (EEI68919), which shared common sequences with two adhesion extracellular proteins from *L. rhamnosus* GG CAR88818 and BAI42720, respectively. Also among the 213 predicted SEPs, two sequences, EEI68684 and EEI68682, with collagen-binding domains were homologous to the major pilin SpaD (CAR88265) and the ancillary pilin SpaF (CAR88267) of *L. rhamnosus* GG. The genome of *L. rhamnosus* GG contains two sortase-dependent pili gene clusters, *SpaCBA* and *SpaFED*. SpaF was reported as the human MUB of the SpaFED pilus.[48] Furthermore, an LPXTG-motif cell wall anchor domain protein (EEI69168) presented high homology to SpaA (CAR86337), the major pilin encoded by the SpaCBA pilus cluster from *L. rhamnosus* GG.[26]

Three predicted *L. paracasei* SEPs (Table S1, ESI†) were homologous to known sortases. One sequence (EEI68552) was homologous to a sortase from *L. casei* (srtA, WP_003588431) and two (EEI69169 and EEI68454) showed homology to class C pilin-specific sortases. Because the predicted SEPs listed pilins and sortases, the presence of sortase-dependent pili clusters in *L. paracasei* seemed probable. Thus, based on conserved domains and sequence alignments of neighbouring genes of the predicted adhesins, three putative clusters encoding sortase-dependent pili were created. Each of these clusters contained one pilin-specific sortase (SrtC). Cluster A (Fig. 2A) included a protein termed PilB (EEI68452) which contained two repeats of collagen-binding domain Cna_B and was homologous to a collagen adhesion protein (EEV50121) found in *Enterococcus faecium*. A second pilin, PilA (EEI68453), contained a fimbrial domain found in streptococcal adhesins with lectin-like

properties.[49] This suggested the presence of a pilus made of two subunits of similar mass assembled by sortases, EEI68455, EEI68451 and/or EEI68454. The putative *SpaCBA* operon (cluster B) in *L. paracasei* (Fig. 2B) consists of four pilin subunits, two of which present homology with the tip adhesin SpaC from *L. rhamnosus* GG. The largest pilin, EEI69165 (SpaC_vWF) (Fig. 2B), contained adhesin-like conserved domains including the von Willebrand factor motif found in the *L. plantarum* msa and in *L. rhamnosus* GG SpaC, a collagen-binding domain, and a metal ion-dependent adhesion site (MIDAS), which suggests that this protein may possess lectin function. The predicted cluster also harbours two genes that might be involved in transcriptional regulation. The first gene, a DNA resolvase (Rase, EEI69164), belongs to a class of proteins that has been found previously directly upstream of pili clusters.[50] The second gene found upstream of the pili cluster is a regulatory protein (ArpU, EEI68878), which contains a sigma factor needed for initiation of RNA synthesis (domain accession [TIGR01637]). The proposed *SpaFED* operon (cluster C) contains three pilin subunits, EEI68684 (SpaD), EEI68683 (SpaE), and EEI68682 (SpaF), with EEI68682 representing the likely large adhesin (Fig. 2C). This cluster also contains several regulatory elements: (i) a transcriptional regulator such as the universal stress protein (USP, EEI68681), (ii) a motif required for DNA binding (DBHTP, EEI68680), which is involved in pili phase variation in *Neisseria* spp., (iii) a tyrosine phosphatase found in signalling pathways (TPase, EEI68679), and (iv) a protein involved in the control of gene expression such as the iron-sulphur cluster-binding protein (QueG, EEI68878), which is identical to a regulatory protein (EPC74700).

Sequences of putative adhesins from the sortase-dependent pili were further analysed using the Phyre2 suite that predicts protein structure and function. All three pilin subunits were successfully modelled and characterised as cell adhesion proteins (data not shown).

3.2. *In vitro* expression and identification of putative adhesins in *L. paracasei*

Total RNA was isolated from *L. paracasei* harvested in mid-log phase in regular MRS media (MRS), in broth supplemented with mucin (MRS+PGM), reported to trigger lactobacilli adhesion,[51] in broth supplemented with salt (MRS+NaCl) to mimic the osmolarity of the gut,[52] or in media supplemented with both mucin and salt (MRS+PGM/NaCl). Transcript levels of the 13 putative adhesion genes (Table 2) were determined by RT-qPCR (Table S8, ESI†). All adhesin candidates were expressed *in vitro*. Statistical analysis of fold changes by ANOVA identified significant up-regulation of all the candidate genes from lag to log phase (Table S8, ESI†). Tukey tests subsequently identified five of the 13 candidates that were significantly up-regulated ($P<0.05$) in MRS+PGM/NaCl which were the translation factor EF-Tu, which has been previously described as involved in adhesion in lactobacilli strains[37,53] and the four pilin candidates EEI68452, EEI69168, EEI68684 and EEI68682 (Table S8, ESI† and Fig. 2).

L. paracasei cells from MRS+PGM/NaCl was grown to mid-exponential phase, imaged by TEM and compared to two other *Lactobacillus* species, *L. johnsonii*, which has a pilus-like gene cluster, and *L. rhamnosus* GG, which has been previously shown to express pili *in vitro*.[26,54] Previously, genes encoding the sortase-dependent pilus *spaCBA* from *L. rhamnosus* GG were expressed in the exponential phase of growth,[55] so *L. johnsonii* and *L. rhamnosus* GG were also cultured to mid-exponential phase. Despite an indication of thin cilia-like structures surrounding the bacteria, the thicker appendages which extend substantially from the cell wall that are characteristic of pili[26,54] were not observed for any of the three species examined (Fig. 3).

To confirm the presence of adhesin candidate proteins from the *in silico* analysis, proteins from the three species were isolated using 1 M LiCl, a method which has been previously used for selective surface protein extraction to identify a mucus- and mucin-binding surface protein from *L. fermentum* 104R.[44] Under native conditions, SDS-PAGE analysis of the

isolated SEPs resulted in multiple bands ranging in size from 20 to 180 kDa (Fig. 4A). The pattern of SEP migration differed between the three species, which suggested differences in SEP composition. The proteins in several gel bands and a chunk were digested in gel and analysed by mass spectrometry (Fig 4B). *L. paracasei* EF-Tu, GroEL chaperonin and PII-type proteinase were identified (Table 4). Although there were bands visible between 65 and 80 kDa (Fig. 4A) which could correspond to PilA (73 kDa) or PilB (69 kDa) of cluster A, pili components were not identified by MS analysis.

3.3. GIT mucin specificity of *L. johnsonii, L. rhamnosus* and *L. paracasei*

Whole cells of *L. paracasei*, *L. johnsonii* and *L. rhamnosus* were internally fluorescently labelled and incubated on GIT mucin microarrays (Table S4, ESI†) to determine their mucin-binding specificity (Fig. 5A). All three species bound to GIT mucins in a similar pattern with a few differences in terms of their relative intensities of binding. After hierarchically clustering the binding intensity data, *L. rhamnosus* and *L. johnsonii* shared 81% similarity in terms of their GIT mucin specificity while *L. paracasei* shared only 52% similarity with the binding patterns of *L. rhamnosus* and *L. johnsonii* (Fig. 5A). *L. paracasei* mainly differed from the other two species in terms of the relative intensity of binding, with the only 'on/off' difference that *L. paracasei* bound to rat duodenum mucin and not rat cecum mucin while *L. rhamnosus* and *L. johnsonii* bound to mucin from rat cecum but not rat duodenum. In terms of overall GIT mucin localisation in presented species, all three *Lactobacillus* species bound well to small intestinal mucin (chicken proximal small intestine (monogastric), ovine jejunum (ruminant), equine small intestine (hind gut fermentor) and porcine jejunum (mongastric)), one cecal mucin (porcine ceca), one ruminant stomach mucin (bovine abomasum (lower stomach of multicompartment stomach)), some large intestinal mucins (ovine spiral colon (ruminant), bovine spiral colon (ruminant) and mouse large intestine (hind gut fermentor)) and the human colon carcinoma-derived LS174T cell line, which mainly contains MUC2.[38]

Interestingly none of the three *Lactobacillus* species bound to the E12 mucin which is from human goblet-like HT29-MTX-E12 cells which mainly contain MUC5AC,[38] which may indicate a preference for large intestine localisation in the human GIT. Additionally, *L. rhamnosus* and *L. johnsonii* favoured binding to stomach mucins (porcine stomach (monogastric), deer abomasum (multicompartment ruminant)), porcine spiral colon mucin (large intestine, monogastric), rat cecal mucin and equine colonic mucins (left and right ventral colon and dorsal colon (hind gut fermentor)) to which *L. paracasei* did not bind or only bound with low intensity (Fig. 5A). Taken together, these data may indicate a similar and more promiscuous *in vivo* GIT localisation for *L. rhamnosus* and *L. johnsonii* while *L. paracasei* may exhibit a more conserved GIT localisation preference, but these preferences are host species dependent.

The GIT mucin microarray was also incubated with the lectins AIA, MAA, WFA and UEA-I, which have varying binding affinities for specific carbohydrate structures (Table 1), to determine the presence of certain carbohydrate structure in the GIT mucins (Fig. 5A). The binding patterns of whole cells of *L. paracasei*, *L. rhamnosus* and *L. johnsonii* to mucins and control glycoproteins were most similar to that of MAA (56% similarity with *L. paracasei* binding and 52% similarity with *L. rhamnosus* and *L. johnsonii* binding). As MAA has a binding specificity for α-(2→3)-linked sialic acid, these data suggest that α-(2→3)-linked sialic acid may be an important ligand for binding of the three *Lactobacillus* species to GIT mucins.

3.4. Carbohydrate specificity of *L. johnsonii*, *L. rhamnosus* and *L. paracasei*

Whole cells were also incubated on carbohydrate microarrays (Tables S5 and S6, ESI†) to identify individual carbohydrate structures bound by the three *Lactobacillus* species. Although glycosaminoglycans are also components of the microenvironment, we have

focused on mucin-type glycosylation in this study to complement and further discriminate the mucin-binding preferences of the species. The binding profiles of all three species were similar, with the binding pattern of *L. johnsonii* and *L. paracasei* most similar to one another (71% similarity) (Fig. 5B). In general, all three *Lactobacillus* species favoured binding to fucosylated and/or sialylated structures but none bound to high mannose type N-linked structures as found on glycoproteins (see M3BSA and Inv, Fig. 5B).

In terms of the preference for α-(2→3)-linked sialic acid (Neu5Ac) suggested by the GIT mucin microarray profiling, all three species bound intensely to both α-(2→3)- and α-(2→6)-linked Neu5Ac on 3'- and 6'-sialyllactose (3'SLac and 6'SLac, respectively) (see 3SLacHSA and 6SLacHSA, Fig. 5B). However, only *L. rhamnosus* bound intensely to 3'-sialyl-*N*-acetyllactosamine (3'SLacNAc) (see 3SLNBSA, Fig. 5B) as is commonly found on N-linked oligosaccharides on glycoproteins, while *L. paracasei* bound moderately and *L. johnsonii* weakly to the same neoglycoconjugate. Interestingly, none of the species bound to human α_1-acid glycoprotein (AGP, Fig. 5B), which has complex type N-linked oligosaccharides with terminal Neu5Ac α-(2→3)-linked to *N*-acetyllactosamine (LacNAc) and terminal β-linked galactose (Gal) residues of LacNAc on short antennae. These structures can also be fucosylated producing sialyl Lewis x structures (SLex, Neu5Ac-α-(2→3)-Gal-β-(1→4)-[Fuc-α-(1→3)-]GlcNAc).[56] Additionally, neither *L. johnsonii* nor *L. rhamnosus* bound to SLex (see 3SLexBSA3, 3SLexBSA14 and 3SFL, Fig. 5B) while *L. paracasei* exhibited weak binding, which indicated that binding to α-(2→3)-linked Neu5Ac was inhibited or blocked in the presence of α-(1→3)-linked fucose (Fuc). However, when the Lex structure was on a different antenna from the terminal α-(2→3)-linked Neu5Ac, binding to *L. rhamnosus* was still inhibited while *L. paracasei* and *L. johnsonii* bound (see MMLNnHHSA, Fig. 5B and Tables S5 and S6, ESI†). However, this lack of binding of *L. rhamnosus* to terminal α-(2→3)-linked Neu5Ac in the MMLNnHHSA configuration may be due to the presentation of

the terminal Neu5Ac on an extended structure rather than on a shorter structure, as *L. rhamnosus* also did not bind to terminal α-(2→3)-linked Neu5Ac presented on the SLNnTHSA extended structure but both *L. paracasei* and *L. johnsonii* did (Fig. 5B and Tables S5 and S6, ESI†). All three species also bound to the ganglioside structure GM1 (Gal-β-(1→3)-GalNAc-β-(1→4)-[Neu5Ac-α-(2→3)-]Gal-β-(1→4)-Glc-APD-HSA) but *L rhamnosus* binding was abolished when the branched α-(2→3)-linked Neu5Ac was removed (asialoGM1) while *L. paracasei* and *L. johnsonii* still bound (see GM1HSA and aGM1HSA, Fig. 5B).

Differential binding between species was investigated by ANOVA and Tukey tests (Fig. S3, ESI†) and 30 neoglycoconjugates were identified as significant between the two casei group members *L. rhamnosus* and *L. paracasei*, 11 neoglycoconjugates between *L. johnsonii* and *L. paracasei* and 16 neoglycoconjugates between *L. rhamnosus* and *L. johnsonii*. Of the ten common neoglycoconjugates that differed between *L. paracasei* and the other two strains (Fig. 5C), nine structures were fucosylated, seven of which contained α-(1→2)-linked Fuc as found in the H2 antigen. In certain structures, such as blood group A and B (see BGABSA and BGB, Fig. 5B and 5C), *L. paracasei* bound while *L. rhamnosus* and *L. johnsonii* did not which indicated that *L. rhamnosus* and *L. johnsonii* had a binding preference for terminal α-(1→2)-linked Fuc, as is found in the H2 antigen (blood group O), while *L. paracasei* bound to both terminal and branching (non-terminal) α-(1→2)-linked Fuc. Binding to terminal α-(1→2)-linked Fuc for *L. rhamnosus* was inhibited in the presence of branching α-(2→6)-linked Neu5Ac on an extended structure but not *L. johnsonii* or *L. paracasei* (see SLNFVHSA, Fig. 5B and Tables S5 and S6, ESI†). Additionally, *L. paracasei* also bound to Lewis y (LeyHSA, Fig. 5B and 5C), which indicated a tolerance for α-(1→3)-linked Fuc in the presence of terminal α-(1→2)-linked Fuc (Lewis y). *L. paracasei* also bound to Lewis b structures (LNDHIBSA and LebBSA, Fig. 5B), which additionally incorporated α-(1→4)-

linked Fuc (Fig. 5C), and difucosyl-para-lacto-N-hexaose (DFPLH, Fig. 5B), which indicated a tolerance for α-(1→3)-linked Fuc in the presence of α-(1→4)-linked Fuc (Fig. 5C).

Overall taking these data together, all *Lactobacillus* species profiled favoured binding structures which contained terminal α-(2→3)-linked Neu5Ac or terminal α-(1→2)-linked Fuc. The presence of α-(1→3)-linked Fuc abolished *L. rhamnosus* and *L. johnsonii* binding while *L. paracasei* tolerated the presence of this motif and had an additional binding preference for branched α-(1→2)- and α-(1→4)-linked Fuc.

4. Discussion

Colonising the mucin layer is paramount to the host-commensal relationship in the GIT and commensal adhesins with lectin function are key to successful persistence in the GIT. However, despite their importance in understanding commensal colonisation, investigation of commensal adhesins has lagged behind those of pathogens. In this study we used an *in silico* approach to identify putative adhesins of the important human commensal *L. paracasei* subsp. *paracasei*, along with RT-qPCR, SDS-PAGE and MS protein identification to examine adhesin expression. We also investigated the potential GIT localisation and carbohydrate specificity of the surface bound *L. paracasei* adhesins using GIT mucin and carbohydrate microarrays.

Following *in silico* analysis, 13 of 213 surface-exposed proteins were selected based on conserved domains as potential *L. paracasei* adhesins, seven of which contained motifs from the collagen-binding domain superfamily that mediates bacterial adherence. The identified potential adhesins had a wide range of isoelectric points (pI 4.5 - 9.6), which suggested that attachment of *L. paracasei* to the host ECM or epithelial cell surface could occur in various

oro-gastrointestinal tract habitats differing in their redox potential. None of the selected adhesin candidates for *L. paracasei* strain ATCC 25302 contained the MUB domain found in various lactobacilli and which was recently described in *L. reuteri* ATTC 53608 MUB[57] and in *L. reuteri* DSM20016 protein Lar0958.[24] In addition, the largest lactobacilli known cell-surface protein, the lactobacilli aggregation promoting factor (AggLb, 319 kDa), which was recently identified in *L. paracasei* strain BGNJ1-64[58] and contains multiple repeats of collagen-binding domains, was not identified in *L. paracasei* subsp. *paracasei* ATCC 25302 following BLAST analysis. However, AggLb is plasmid-encoded in BGNJ1-64 and therefore would not necessarily be expected to be present in strain ATCC25302. However several candidates presented homology to pilin subunits or adhesion proteins from the related strain of the casei group of Lactobacilli *L. rhamnosus* GG that bind human mucus.[59]

All of the 13 adhesin candidate genes were expressed during growth *in vitro.* Five candidates were up-regulated during exponential phase when mucin and salt were present in the growth media. Translation factor EF-Tu, one of the up-regulated genes, has been previously described as involved in adhesion in lactobacilli strains.[37,53] Four other adhesion candidates corresponded to pilin subunits of three proposed clusters encoding sortase-dependent pili. A trimeric model of the mucus-/collagen-binding fimbriae encoded by the *SpaCBA* cluster very different from the *C. diphtheriae* prototype was previously suggested for *L. rhamnosus* GG, based on TEM of immunogold-labelled antibodies raised against the various pilins.[54] In this prototype a backbone is formed by the major subunit SpaA with minor pilin (SpaB) subunits positioned between the major subunits and decoration of the shaft by the adhesin SpaC.[59] In contrast, the SpaCBA pilus proposed here in *L. paracasei* includes four proteins, two minor pilin subunits homologous to SpaC, with the largest (SpaC_vWF) likely to be the adhesin due to the presence of a lectin domain. Phase variation of piliation involving transcriptional elements has been described in *E. coli* for type 1 and P fimbriae formation.[60,61] The molecular

'switch' of the *fim* operon, responsible for the production of type 1 fimbriae, consists of an invertible DNA element containing the promoter. Recombinases such as FimB and FimE act to invert that element and enable the transcription or the silencing of the operon upon temperature change.[62] In the predicted *SpaCBA* cluster from *L. paracasei*, a recombinase was found directly upstream of the sortase (resolvase, EEI69164) as well as a regulatory protein (ArpU, EEI68878) containing a possible sigma factor (TIGR01637). These factors have been associated with virulence in a number of species by acting to redirect transcription initiation.[63]

The proposed *SpaFED* operon in *L. paracasei* presented homology to the *L. rhamnosus* GG *SpaFED* operon, which has with two major subunits (SpaD and SpaE) and the minor pilin SpaF,[58] with SpaF from *L. rhamnosus* GG characterised as the likely adhesion.[48] The proposed *SpaFED* operon in *L. paracasei* harbours a universal stress protein (USP). Functions of USPs are largely unknown but they have been linked to oxidative stress defence and cell adhesion in *E. coli*.[64] Moreover, the expression of *USP* genes is upregulated by quorum sensing in some species of *Burkholderia*.[65] Three other proteins involved in gene regulation were also found directly upstream of the *SpaFED USP* gene. Thus, the *USP* gene in *L. paracasei* may be part of the regulatory system of the *SpaFED* operon and may function as a 'switch' for piliation. Similarly to the *Tad* pilus genes from *Bifidobacterium breve* UCC2003,[66] the *L. rhamnosus* GG *SpaFED* pilin genes are not transcribed *in vitro*, possibly due to the lack of an appropriate environmental stimuli. In our study, *SpaD* and *SpaF* transcripts were only expressed and upregulated in the presence of mucin and salt.

Previously, cell surface proteins containing EF-Tu were isolated from *L. johnsonii* NCC533 using 0.5 M LiCl,[37] total surface proteins containing aggregation-promoting factor (APF) were isolated from *L. johnsonii* and *L. gasseri* using 5 M LiCl[67] and a surface protein from *L. fermentum* was extracted using 1 M LiCl, which bound to partially purified porcine intestinal

mucus and PGM.[44] In this work, cell surface proteins extracted from *L. paracasei* with 1 M LiCl enabled the identification of *L. paracasei* EF-Tu as well as the chaperonin GroL, but pili subunit proteins were not found to be present. EF-Tu has been previously identified in many species including *L. johnsonii* NCC 533 where it was shown to bind to normal human colonic mucins.[37] The function of EF-Tu is broader than just helping the bacteria adhere to mucin or the cell surface. *L. plantarum* EF-Tu from mediated *L. plantarum* binding to Caco-2 cells and reduced pathogen adhesion to cells,[68] while EF-Tu and GroEL from *L. paracasei*, *L. plantarum* and *L. brevis* were characterised as binding to actin.[69] In addition, *L. johnsonii* NCC 533 GroEL was shown to mediate bacterial attachment to mucins and epithelial cells in a pH dependent manner and aggregation of the gastric pathogen *Helicobacter pylori*.[70]

If pili subunit proteins were indeed expressed under these conditions, the LiCl concentration used may have been too dilute to extract proteins that are more tightly associated with the bacterial surface such as pili. More potent methods such as total protein extracts combined with immunoaffinity chromatography using antibodies against the pili proteins[26] may be more effective options for future efforts to conclusively identify the expression of pili subunit proteins in *L. paracasei*. In addition, appendages on the whole bacteria were not visible by TEM. Bacteria sense and adapt to their surrounding environment *via* quorum sensing. Stress-related conditions such as temperature, iron deficiency, high salt concentration or acidity, such as are found in the GIT, can turn on expression of certain genes and change behaviour of microbes. For example, kefir-isolated *L. paracasei* strains demonstrated increased adhesion *in vitro* to Caco-2 cells and PGM after simulated gastrointestinal passage.[17] Some pili such as the *Tad* pilus genes from *Bifidobacterium breve* UCC2003 can only be expressed in the gut environment[66] and so *L. paracasei* pili may not have been expressed under our *in vitro* conditions. However, previously *L. rhamnosus* GG pili have only been visualised using gold labelled anti-pili subunit antibodies,[26,54] so it is likely that even if pili had been present under

our conditions, we would not be able to visualise them using TEM without gold-labelled antibodies.

Very few studies have identified specific carbohydrate ligands for commensal adhesins. However, GIT mucins are more reflective of the natural environment since they present hundreds of heterogeneous glycans in a multimeric manner and distinct binding signatures on the GIT mucin microarray may indicate the ability of the individual *Lactobacillus* species to colonise particular ecological niches of the host GIT. Even though *L. paracasei* and *L. rhamnosus* are phylogenetically close, *L. rhamnosus* and *L. johnsonii* displayed the most similar binding on the GIT mucin microarray which may indicate a different region-specificity for *L. paracasei* in certain host species. Previously in gnotobiotic mice, *L. paracasei* subsp. *paracasei* was found throughout the GIT with greater localisation in the colon,[13] while on the mucin microarray *L. paracasei* bound moderately to conventionally raised mouse large intestine mucin but not to mucins from elsewhere in the mouse GIT. More intense binding was observed with GIT mucins from other species, including rat, which may indicate that mouse is not the optimal species to study *L. paracasei* colonisation and metabolic effects. In humans however, all three *Lactobacillus* species may exhibit preference for the same GIT region, the large intestine. The binding affinity of all three *Lactobacillus* species for the α-(2→3)-linked Neu5Ac structural motif indicated by their mucin microarray binding signatures was confirmed by profiling on the carbohydrate microarray, which further revealed an additional affinity for the H2 antigen. This was supported by strong *L. paracasei* binding to bovine spiral colonic mucin, which showed the highest proportion of α-(1→2)-linked Fuc residues across various mammalian mucins.[38] However, *L. paracasei* was distinguished from the other two species by its ability to additionally bind α-(1→2)-linked Fuc in non-terminal positions and so could bind to blood group A and B in addition to type O (H2 antigen), while *L. rhamnosus* and *L. johnsonii* could only bind to the H2 antigen.

Additionally, *L. paracasei* also recognised the fucosylated Lewis antigens Lex, Ley and Leb, which may further indicate the ability of *L. paracasei* to adapt to specific niches in the GIT.

There is mounting evidence for the importance of the role of Fuc in commensal binding. The secretor phenotype, where a functional fucosyltransferase-2 (FUT2) adds α-(1→2)-linked Fuc to mucosal surfaces, has been recently related to the composition of the GIT microbiota, with an increased numbers of bifidobacteria and lactobacilli colonising the GIT of secretors.[71,72] ABO histo-blood group antigens are ligands for lactobacilli strains from the acidophilus group such as *L. brevis*.[35] Further studies involving recombinant expression or affinity purification of identified adhesins would be required to discriminate respective ligands for individual adhesins or pili.

Conclusion

This study provides new insights into adhesion of beneficial microbes to the human gastrointestinal tract. A combinatorial multidisciplinary approach using *in silico* tools and *in vitro* expression analysis was applied to identify novel adhesins from the GIT commensal *Lactobacillus paracasei* subsp. *paracasei* ATCC 25302. While pili were identified *in silico* and gene expression analysis, pili were not observed by mass spectrometry or TEM. However, EF-Tu and GroEL adhesins were identified by mass spectroscopy. Mucin microarray profiling indicated the potential of *L. paracasei*, *L. rhamnosus* GG and *L. johnsonii* to colonise common and specific ecological niches throughout the GIT while carbohydrate microarray profiling revealed that *L. paracasei* differentially recognised the blood group A and B and Lewis antigens compared to the other two *Lactobacillus* species. The approach demonstrated in this work is promising for the identification of specific adhesins and elucidation of potential GIT localisation of commensals *via* identification of

their GIT mucin and carbohydrate ligands, and can it further our knowledge of bacterial-host cross-talk.

Conflicts of interest

There are no conflicts of interest to declare.

Acknowledgements

This work was supported by Science Foundation Ireland Strategic Research Cluster programme in support of Alimentary Glycoscience Research Cluster (grant number 08/SRC/B1393). MK is grateful to the Royal Society of Chemistry Analytical Chemistry Trust Fund (ACTF) for the ACTF Fellowship Award 2018. The authors acknowledge the Centre for Microscopy & Imaging at the National University of Ireland, Galway (NUI, Galway) funded by NUI Galway and the Irish Government Programme for Research in Third Level Institutions, Cycles 4 and 5, National Development Plan 2007-2013. The authors thank Prof. Maarten Van de Guchte and Mr Valentin Loux, Unité Mathématique, Informatique et Génome, INRA, Jouy en Josas, France for help with SurfG+ analysis, Prof. Y. C. Lee, Department of Biology, Johns Hopkins University, Baltimore, Maryland, U.S.A., for the kind gift of pigeon egg white and Prof. Stephen D. Carrington, Dr. Mary E. Gallagher, Dr. Colm Reid and Dr. Marguerite Clyne of University College Dublin for the kind gift of mucins.

References

1 J. A. Hawrelak and S. P. Myers, The causes of intestinal dysbiosis: a review, *Altern. Med. Rev.*, 2004, **9**, 180-197.

2 E. A. Mayer, T. Savidge and R. J. Shulman, Brain-gut microbiome interactions and functional bowel disorders, *Gastroenterology*, 2014, **146**, 1500-1512.

3 R. E. Ley, P. J. Turnbaugh, S. Klein and J. I. Gordon JI, Microbial ecology - human gut microbes associated with obesity, *Nature*, 2006, **444**, 1022-1023.

4 C. Robbe, C. Capon, E. Maes, M. Rousset, A. Zweibaum, J. P. Zanetta and J. C. Michalski, Evidence of regio-specific glycosylation in human intestinal mucins - presence of an acidic gradient along the intestinal tract, *J. Biol. Chem.*, 2003, **278**, 46337-46348.

5 I. Sekirov, S. L. Russell, L. C. Antunes and B. B. Finlay, Gut microbiota in health and disease, *Physiol. Rev.*, 2010, **90**, 859-904.

6 K. S. B. Bergstrom and L. Xia, Mucin-type *O*-glycans and their roles in intestinal homeostasis, *Glycobiology*, 2013, **23**, 1026-1037.

7 S. Etzold and N. Juge, Structural insights into bacterial recognition of intestinal mucins, *Curr. Opin. Struct. Biol.*, 2014, **28**, 23-31.

8 M. E. V. Johansson, H. E. Jakobsson, J. Holmén-Larsson, A. Schütte, A. Ermund, A. M., Rodríguez-Piñeiro, L. Airke, C. Wising, F. Svensson, F. Bäckhed and G.C. Hansson, Normalisation of host intestinal mucus layers requires long-term microbial colonisation, *Cell Host Microbe*, 2015, **18**, 582-592.

9 F. Bendali, N. Madi and D. Sadoun, Beneficial effects of a strain of *Lactobacillus paracasei* subsp *paracasei* in *Staphylococcus aureus*-induced intestinal and colonic injury, *Int. J. Inf. Dis.*, 2011, **15**, E787-E794.

10 S.-S. Chiang and T.-M. Pan, Beneficial effects of *Lactobacillus paracasei* subsp. *paracasei* NTU 101 and its fermented products, *Appl. Microbiol. Biotechnol.*, 2012, **93**, 903-916.

11. M. L. Y. Wan, S. J. Forsythe and H. El-Nezami, Probiotics interaction with foodborne pathogens: a potential alternative to antibiotics and future challenges, *Crit. Rev. Food Sci. Nutr.*, 2018, **5**, 1-14.

12. S. Lebeer, I. Claes, H. L. P. Tytgat, T. L. A. Verhoeven, E. Marien, I. von Ossowski, J. Reunanen, A. Palva, W. M. de Vos, S. C. J. De Keersmaecker and J. Vanderleyden, Functional analysis of *Lactobacillus rhamnosus* GG pili in relation to adhesion and immunomodulatory interactions with intestinal epithelial cells, *Appl. Environ. Microbiol.*, 2012, **78**, 185-193.

13. F.-P. J. Martin, Y. Wang, N. Sprenger, E. Holmes, J. C. Lindon, S. Kochhar and J. K. Nicholson, Effects of probiotic *Lactobacillus paracasei* treatment on the host gut tissue metabolic profiles probed via magic-angle spinning NMR spectroscopy, *J. Prot. Res.*, 2007, **6**, 1471-1481.

14. M. Arici, B. Bilgin, O. Sagdic and C. Ozdemir, Some characteristics of *Lactobacillus* isolates from infant faeces. *Food Microbiol.*, 2004, **21**, 19-24.

15. J. Maukonen, J. Motto, M. L. Suihko and M. Saarela, Intra-individual diversity and similarity of salivary and faecal microbiota. *J. Med. Microbiol.*, 2008, **57**, 1560-1568.

16. J. L. Sonnenburg, L. T. Angenent and J. I. Gordon. "Getting a grip on things: how do communities of bacterial symbionts become established in our intestine?" *Nat. Immunol.*, 2004, **5**, 569-573.

17. A. A. Bengoa, *et al*., Simulated gastrointestinal conditions increase adhesion ability of *Lactobacillus paracasei* strains isolated from kefir to Caco-2 cells and mucin, *Food Res. Int*., 2018, **103**, 462-467.
18. J. M. Patti, B. L. Allen, M. J. McGavin and M. Hook, MSCRAMM-mediated adherence of microorganisms to host tissues. *Ann. Rev. Microbiol*., 1994, **48**, 585-617.
19. W. W. Navarre and O. Schneewind, Surface proteins of Gram-positive bacteria and mechanisms of their targeting to the cell wall envelope. *Microbiol. Mol. Biol. Rev.*, 1999, **63**, 174-229.
20. T. Proft T and E. N. Baker, EN. Pili in Gram-negative and Gram-positive bacteria - structure, assembly and their role in disease. *Cell. Mol. Life Sci.*, 2009, **66**, 613-635.
21. H. Ton-That and O. Schneewind, Assembly of pili in Gram-positive bacteria. *Trends Microbiol*., 2004, **12**, 228-234.
22. B. Wang, H. Wei, J. Yuan, Q. R. Li, Y. S. Li, N. Li and J. S. Li, Identification of a surface protein from *Lactobacillus reuteri* JCM1081 that adheres to porcine gastric mucin and human enterocyte-like HT-29 cells, *Curr. Microbiol.*, 2008, **57**, 33-38.
23. B. L. Buck, E. Altermann, T. Svingerud and T. R. Klaenhammer, Functional analysis of putative adhesion factors in *Lactobacillus acidophilus* NCFM. Appl Environ Microbiol., 2005, **71**, 8344-8351.
24. S. Etzold, D. A. MacKenzie, F. Jeffers, J. Walshaw, S. Roos, A. M. Hemmings and N. Juge, Structural and molecular insights into novel surface-exposed mucus adhesins from *Lactobacillus reuteri* human strains. *Mol. Microbiol*., 2014, **92**, 543-556.
25. N. Juge,Microbial adhesins to gastrointestinal mucus. *Trends Microbiol.*, 2012, **20**, 30-39.
26. J. Reunanen, I. von Ossowski, A. P. A. Hendrickx, A. Palva and W. M. de Vos, Characterization of the SpaCBA pilus fibers in the probiotic *Lactobacillus rhamnosus* GG. *Appl. Environ. Microbiol*., 2012, **78**, 2337-2344.
27. I. von Ossowski, R. Satokari, J. Reunanen, S. Lebeer, S. C. J. De Keersmaecker, J. Vanderleyden, W. M. de Vos and A. Palva A. 2011. Functional characterization of a mucus-specific LPXTG surface adhesin from probiotic *Lactobacillus rhamnosus* GG. *Appl. Environ. Microbiol.*, 2011, **77**, 4465-4472.
28. V. Krishnan, P. Chaurasia and A. Kant, Chapter 6 in 'Pili in probiotic bacteria, Probiotics and prebiotics in human nutrition and health'. Eds. V. Rao, IntechOpen, London, 115-133, 2016.
29. J. Boekhorst, B. M. H. J. de Been, M. Kleerebezem and R. J. Siezen, Genome-wide detection and analysis of cell wall-bound proteins with LPxTG-Like sorting motifs. *J. Bacteriol*., 2005, **187**, 4928-4934.
30. G. Sachdeva, K. Kumar, P. Jain and S. Ramachandran, SPAAN: a software program for prediction of adhesins and adhesin-like proteins using neural networks. *Bioinformatics*, 2005, **21**, 483-491.
31. A. Barinov, V. Loux, A. Hammani, P. Nicolas, P. Langella, D. Ehrlich, E. Maguin and M. van de Guchte, Prediction of surface exposed proteins in *Streptococcus pyogenes*, with a potential application to other Gram-positive bacteria. *Proteomics*, 2009, **9**, 61-73.
32. P. D. Cani and W. M. de Vos, Next-generation beneficial microbes: The case of *Akkermansia muciniphilia*. *Front. Microbiol*., 2017, **8**, 1765.
33. H. Earley, *et al*., A preliminary study examining the binding capacity of *Akkermansia muciniphilia* and *Desulfovibrio* spp. to colonic mucin and health and ulcerative colitis. *PLoS ONE*, 2015, **10**, e0135280.
34. G. Pretzer, *et al*., Biodiversity-based identification and functional characterization of the mannose-specific adhesin of *Lactobacillus plantarum*. *J. Bacteriol*., 2005, **187**, 6128-6136.

35. H. Uchida, *et al*., Lactobacilli binding human A-antigen expressed in intestinal mucosa. *Res. Microbiol*., 2006, **157**, 659-665.
36. J.-R. Neeser, D. Granato, M. Rouvet, A. Servin, S. Teneberg and K.-A. Karlsson, *Lactobacillus johnsonii* La1 shares carbohydrate-binding specificities with several enteropathogenic bacteria, *Glycobiology*, 2000, **10**, 1193-1199.
37. D. Granato, G. E. Bergonzelli, R. D. Pridmore, L. Marvin, M. Rouvet and I. E. Corthesy-Theulaz, Cell surface-associated elongation factor Tu mediates the attachment of *Lactobacillus johnsonii* NCC533 (La1) to human intestinal cells and mucins, *Infect. Immun*., 2004, **72**, 2160-2160.
38. M. Kilcoyne, J. Q. Gerlach, R. Gough, M. E. Gallagher, M. Kane, S. D. Carrington and L. Joshi, Construction of a natural mucin microarray and interrogation for biologically relevant glyco-epitopes. *Anal. Chem*., 2012, **84**, 3330-3338.
39. L. A. Kelley, S. Mezulis, C. M. Yates, M. N. Wass and M. J. E. Sternberg, The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.*, 2015, **10**, 845-858.
40. R. E. Hungate and J. Macy, The roll-tube method for cultivation of strict anaerobes. *Bull. Ecol. Res. Comm. (Stockholm)*, 1973, **17**, 123-126.
41. K. A. G. Karatzas, O. Brennan, S. Heavin, J. Morrissey and C. P. O'Byrne, Intracellular accumulation of high levels of γ-aminobutyrate by *Listeria monocytogenes* 10403S in response to low pH: uncoupling of γ-aminobutyrate synthesis from efflux in a chemically defined medium. *Appl. Environ. Microbiol*., 2010, **76**, 3529-3537.
42. D. J. P. Rocha, C. S. Santos and L. G. C. Pacheco, Bacterial reference genes for gene expression studies by RT-qPCR: survey and analysis, *Antonie Van Leeuwenhoek Int. J. Gen. Mol. Microbiol.*, 2015, **108**, 685-693.
43. K. J. Livak and T. D. Schmittgen, Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) method. *Methods*, 2001, **25**, 402-408.
44. M. Rojas, F. Ascensio and P. L. Conway, Purification and characterization of a surface protein from *Lactobacillus fermentum* 104R that binds to porcine small intestinal mucus and gastric mucin, *Appl. Environ. Microbiol.*, 2002, **68**, 2330-2336.
45. S. Shevchenko, M. Wilm, O. Vorm and M. Mann, Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. *Anal. Chem*., 1996, **68**, 850-858.
46. N. O'Boyle, B. Houeix, M. Kilcoyne, L. Joshi, A. Boyd, The MSHA pilus of *Vibrio parahaemolyticus* has lectin functionality and enables TTSS-mediated pathogenicity. *Int. J. Med. Microbiol*., 2013, **303**, 563-573.
47. M. Utratna, H. Annuk, J. Q. Gerlach, Y. C. Lee, M. Kane, M. Kilcoyne and L. Joshi, Rapid screening for specific glycosylation and pathogen interactions on a 78 species avian egg white glycoprotein microarray. *Sci. Rep*., 2017, **7**, 6477.
48. J. Rintahaka, X. Yu, R. Kant, A. Palva and I. von Ossowski, Phenotypical analysis of the *Lactobacillus rhamnosus* GG fimbrial spaFED operon: surface expression and functional characterization of recombinant SpaFED pili in *Lactococcus lactis*. *PLoS ONE*, 2014, **9**, e113922.
49. A. Marchler-Bauer, *et al*., CDD: a Conserved Domain Database for the functional annotation of proteins. *Nucleic Acids Res.*, 2011, **39**, D225-D229.
50. I. Sohel, J. L. Puente, S. W. Ramer, D. Bieber, C. Y. Wu and G. K. Schoolnik, Enteropathogenic *Escherichia coli*: Identification of a gene cluster coding for bundle-forming pilus morphogenesis. *J. Bacteriol*., 1996, **178**, 2613-2628.
51. H. Jonsson, E. Strom and S. Roos, Addition of mucin to the growth medium triggers mucus-binding activity in different strains of *Lactobacillus reuteri in vitro*. *FEMS Microbiol. Lett.*, 2001, **204**, 19-22.

52. S. D. Sleator, T. Clifford and C. Hill, Gut osmolarity: a key environmental cue initiating the gastrointestinal phase of *Listeria monocytogenes* infection? *Med. Hypotheses*, 2007, **69**, 1090-1092.
53. K. Ramiah, C. A. van Reenen and L. M. T. Dicks, Expression of the mucus adhesion genes Mub and MapA, adhesion-like factor EF-Tu and bacteriocin gene plaA of *Lactobacillus plantarum* 423, monitored with real-time PCR. *Int. J. Food Microbiol*., 2007, **116**, 405-409.
54. M. Kankainen, *et al.*, Comparative genomic analysis of *Lactobacillus rhamnosus* GG reveals pili containing a human-mucus binding protein. *Proc. Nat. Acad. Sci. (U.S.A.)*, 2009, **106**, 17193-17198.
55. K. Laakso, *et al*., Growth phase-associated changes in the proteome and transcriptome of *Lactobacillus rhamnosus* GG in industrial-type whey medium. *Microbial Biotechnol.*, 2011, **4**, 746-766.
56. C. L. Fernandes, R. Ligabue-Braun and H. Verli, Structural glycobiology of human a_1-acid glycoprotein and its implications for pharmacokinetics and inflammation. *Glycobiology*, 2015, **25**, 1125-1133.
57. D. A. MacKenzie, L. E. Tailford, A. M. Hemmings and N. Juge, Crystal structure of a mucus-binding protein repeat reveals an unexpected functional immunoglobulin binding activity. *J. Biol. Chem.*, 2009, **284**, 32444-32453.
58. M. Miljkovic, I. Strahinic, M. Tolinacki, M. Zivkovic, S. Kojic, N. Golic and M. Kojic, AggLb is the largest cell-aggregation factor from *Lactobacillus paracasei* subsp. *paracasei* BGNJ1-64, functions in collagen adhesion, and pathogen exclusion *in vitro*. *PLoS ONE*, 2015, **10**, e0126387.
59. I. von Ossowski, J. Reunanen, R. Satokari, S. Vesterlund, M. Kankainen, H. Huhtinen, S. Tynkkynen, S. Salminen, W. M. de Vos and A. Palva, Mucosal adhesion properties of the probiotic *Lactobacillus rhamnosus* GG SpaCBA and SpaFED pilin subunits. *Appl. Environ. Microbiol.*, 2010, **76**, 2049-57.
60. B. I. Eisenstein, Phase variation of type-1 fimbriae in *Escherichia coli* is under transcriptional control. *Science*, 1981, **214**, 337-339.
61. X. W. Nou, B. Skinner, B. Braaten, L. Blyn, D. Hirsch and D. Low, Regulation of pyelonephritis-associated pili phase-variation in *Escherichia coli* - binding of the Papl and the Lrp regulatory proteins is controlled by DNA methylation. *Mol. Microbiol*., 1993, **7**, 545-553.
62. D. M. Wolf and A. P. Arkin, Fifteen minutes of fim: control of type 1 pili expression in *E. coli*. *Omics*, 2002, **6**, 91-114.
63. M. J. Kazmierczak, M. Wiedmann and K. J. Boor, Alternative sigma factors and their roles in bacterial virulence. *Microbiol. Mol. Biol. Re*v., 2005, **69**, 527-543.
64. L. Nachin, U. Nannmark and T. Nystrom, Differential roles of the universal stress proteins of *Escherichia coli* in oxidative stress resistance, adhesion, and motility. *J. Bacteriol.*, 2005, **187**, 6265-6272.
65. H. Kim, E. Goo, Y. Kang, J. Kim and I. Hwang, Regulation of Universal Stress Protein genes by quorum sensing and RpoS in *Burkholderia glumae*. *J. Bacteriol*., 2012, **194**, 982-992.
66. M. O'Connell Motherway, *et al*., Functional genome analysis of *Bifidobacterium breve* UCC2003 reveals type IVb tight adherence (Tad) pili as an essential and conserved host-colonization factor. *Proc. Nat. Acad. Sci. (U.S.A.)*, 2011, **108**, 11217-11222.
67. M. Ventura, I. Jankovic, D. C. Walker, R. D. Pridmore and R. Zink, Identification and characterization of novel surface proteins in *Lactobacillus johnsonii* and *Lactobacillus gasseri*. *Appl. Environ. Microbiol*., 2002, **68**, 6172-6181.

68. A. S. Dhanani, S. B. Gaudana and T. Bagchi, The ability of *Lactobacillus* adhesin EF-Tu to interfere with pathogen adhesion. *Eur. Food Res. Technol.*, 2011, **232**, 777-785.
69. Z. Peng, R. F. Vogel, M. A. Ehrmann and T. Xiong, Idenetification and characterisation of adhesion proteins in lactobacilli targeting actin as receptor. *Mol. Cell. Probes*, 2018, **37**, 60-63.
70. G. E. Bergonzelli, D. Granato, R. D. Pridmore, L. F. Marvin-Guy, D. Donnicola and I. E. Corthèsy-Theulaz, GroEL of *Lactobacillus johnsonii* La1 (NCC 533) is cell surface associated: potential role in interactions with the host and the gastric pathogen *Helicobacter pylori*. *Infect. Immun.*, 2006, **74**, 425-434.
71. P. Wacklin, H. Makivuokko, N. Alakulppi, J. Nikkila, H. Tenkanen, J. Rabina, J. Partanen, K. Aranko and J. Matto, Secretor genotype (FUT2 gene) is strongly associated with the composition of Bifidobacteria in the human intestine. *PLoS ONE*, 2011, **6**, e20113.
72 P. Wacklin, *et al*., Faecal microbiota composition in adults is associated with the FUT2 gene determining the secretor status. *PLoS ONE*, 2014, **9**, e94863.

FIGURES

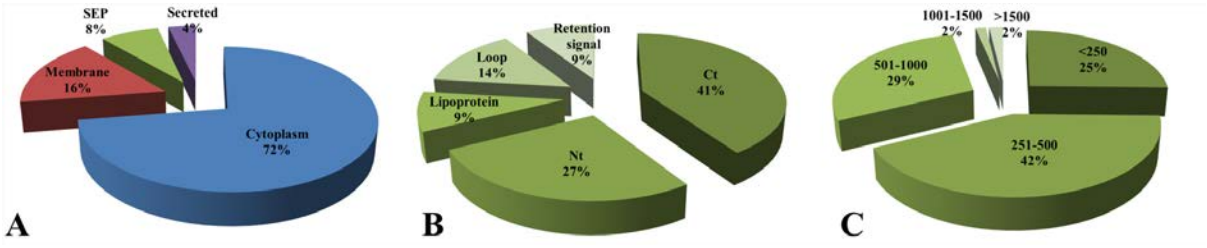

Fig. 1. SurfG+ analysis of the *Lactobacillus paracasei* subsp. *paracasei* ATCC 25302 proteome. (A) Protein localisation. (B) SEP topology; Nt and Ct refer to surface exposed N- or C-terminal of transmembrane helices. (C) SEP chain length (in amino acid residues).

Fig. 2. Putative pilus clusters and candidate pilin expression in *Lactobacillus paracasei* subsp. *paracasei* ATCC 25302. (A), (B) and (C) represent predicted cluster A-C of genes involved in pilus formation. Each coloured box corresponds to a gene and its corresponding protein accession number (NCBI database). Blue boxes correspond to sortase genes (srtA and srtC), purple boxes to pilin subunits (putative adhesin highlighted in red) and grey boxes to putative transcriptional regulators. USP, universal stress protein; DBHTP, DNA bind helix turn helix protein; TPase, tyrosine phosphatase; QueG, epoxyqueueosine reductase-like protein; Rase, resolvase; ArpU, ArpU regulatory protein. Estimated molecular mass of proteins are given in kDa.

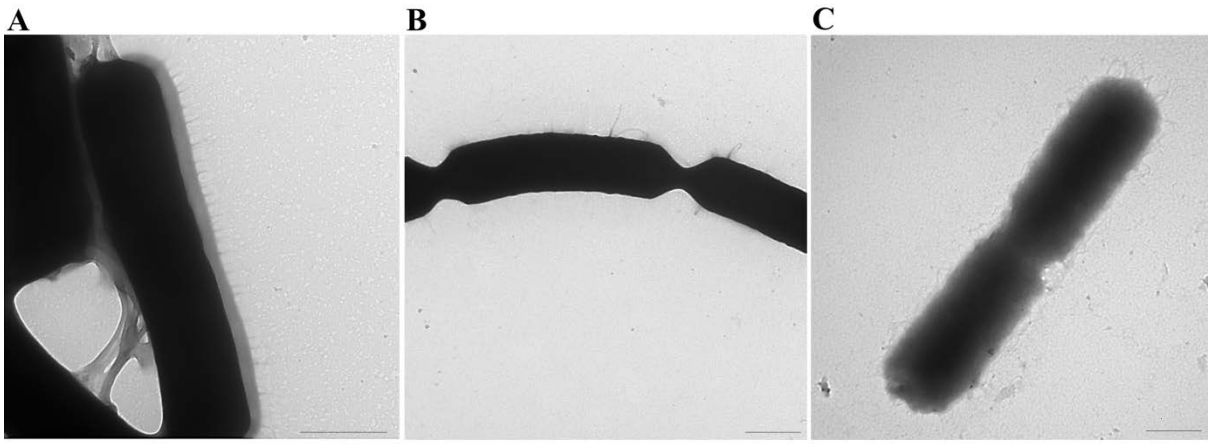

Fig. 3. TEM images of (A) *L. johnsonii*. (B) *L. rhamnosus*, (C) *L. paracasei*. Scale bar corresponds to 500 nm.

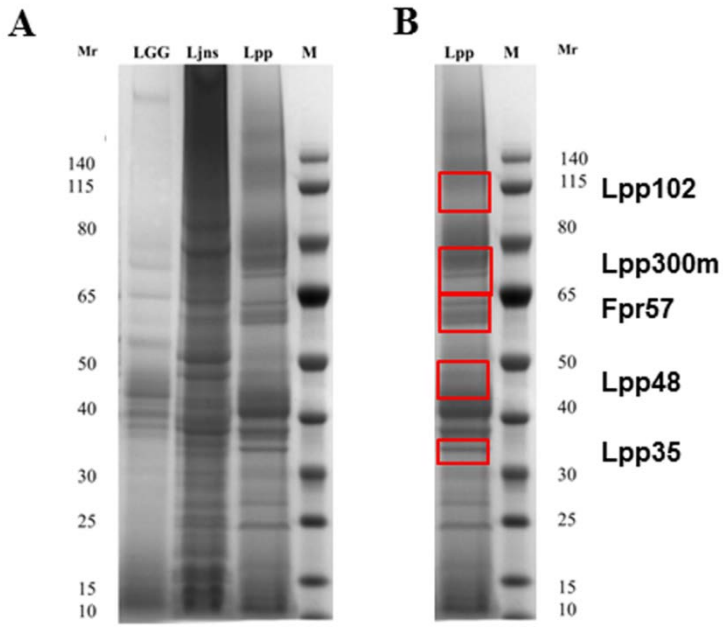

Fig. 4. SDS-PAGE analysis of 1 M LiCl-extracted protein from *L. paracasei* (Lpp), *L. rhamnosus* GG (LGG), and *L. johnsonii* (Ljns). (A) Coomassie-stained SDS-PAGE (4-12%) under reducing conditions loaded with 50 µg protein extracts. (B) Gel bands and chunk excised for MS analysis. M, Fermentas PageRuler; Mr, protein molecular mass.

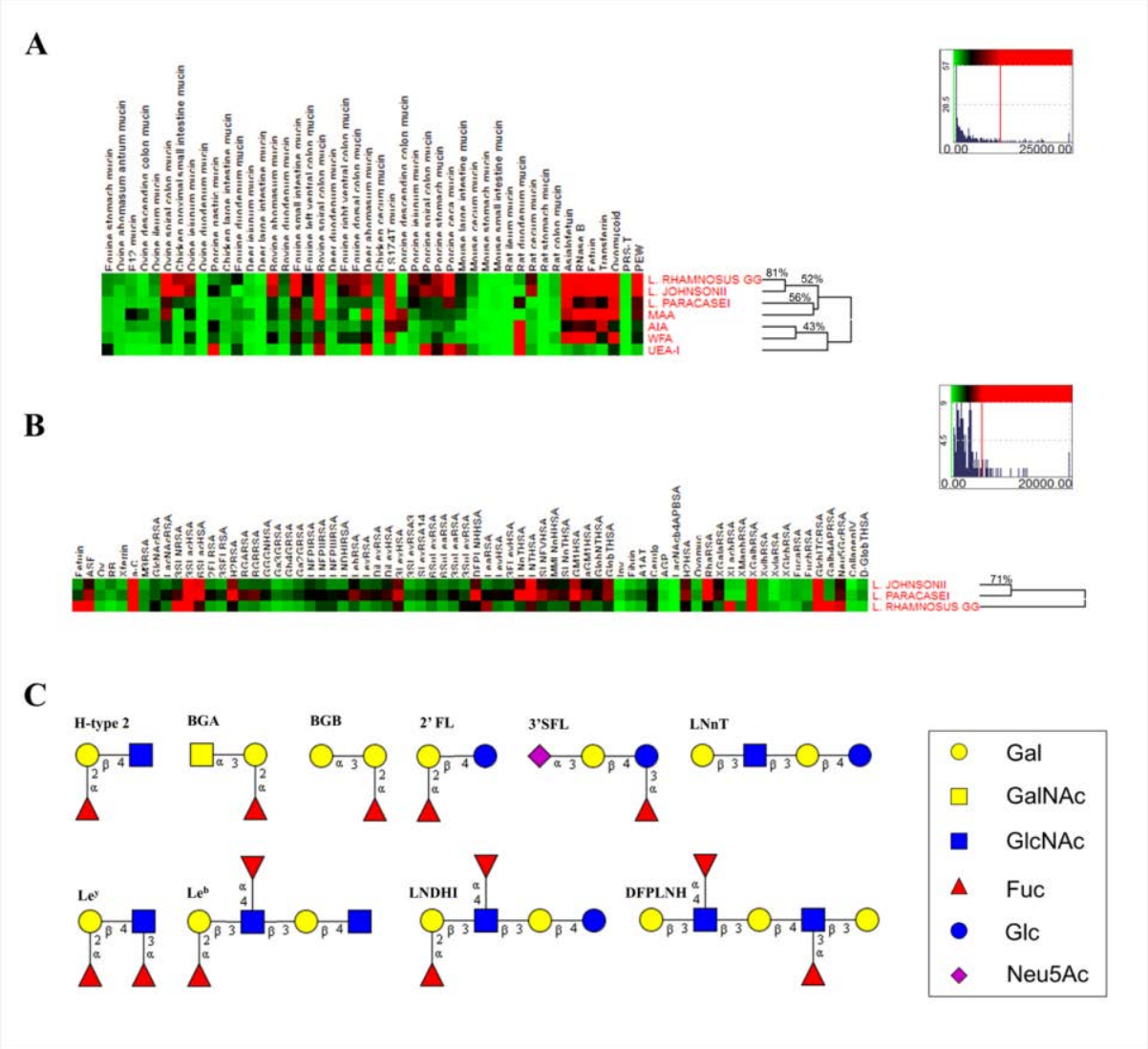

Fig. 5. Hierarchically clustered heat maps depicting binding intensities of *L. paracasei*, *L. rhamnosus* GG and *L. johnsonii* whole cells to (A) GIT mucin microarrays, and (B) carbohydrate microarrays. Binding intensities were hierarchically clustered using complete linkage and Euclidean distances of scale normalised binding intensities and similarity percentages are shown. GIT mucin microarray data additionally includes binding intensities of the lectins MAA, AIA, WFA and UEA-I. (C) Carbohydrate structures bound differentially by *L. paracasei* compared to *L. rhamnosus* GG and *L. johnsonii* whole cells.

Table 1. Plant lectins used, their abbreviations and main carbohydrate binding specificity.

Lectin	Abbreviation	Carbohydrate specificity
Artocarpus integrifolia agglutinin	AIA	β-linked Gal (sialylation independent)
Maackia amurensis agglutinin	MAA	Neu-α-(2,3)-Gal(NAc)-R
Wisteria floribunda agglutinin	WFA	GalNAc/sulfated GalNAc
Ulex europaeus agglutinin isolectin-1	UEA-I	Fuc-α-(1,2)-GlcNAc-R

Table 2. Putative adhesin proteins in *L. paracasei* subsp. *paracasei* ATCC 25302, their accession number in NCBI database, number of amino acids (AA), predicted isoelectric point (pI), and homologous sequences in the NCBI database.

NCBI accession no.	AA	pI	Protein name	Conserved domain	Potential adhesin homologue
EEI67294	396	4.85	Elongation factor Tu	EF_Tu [cd01884]; EFTU_II [cd03697]; EFTU_III [cd03707]; PRK00049 [PRK00049]	EF-Tu (87%, AAS08831, *L. johnsonii* NCC 533)
EEI68802	544	4.89	GROEL chaperonin	GroEL [cd03344]; groEL[PRK12849]	GROEL (84%, AAS08453, *L. johnsonii* NCC 533)
EEI67357	574	9.15	Fibronectin binding protein A	DUF814 super family[cl05307]; FbpA[pfam05833]	FBPA (47%, AAV42987, *L. acidophilus* NCFM)
EEI68137	270	9.60	Collagen binding protein	PBPb [cd00134]	CnBP/MapA (57%, CAA68052, CnBP *L. reuteri*)
EEI68395	2240	5.07	LPXTG-motif cell wall anchor domain protein	Collagen_Bind super family [cl27951]: 10 repeats of Cna_B [pfam05738]; COG4932 [COG4932]	Collagen adhesion protein (EEN79876, *L. rhamnosus* LMS2-1)
EEI68842	611	6.49	Outer membrane protein	Collagen_Bind super family [cl27951]: Cna_B [pfam05738]; Cna_A [pfam05737]	Collagen binding protein (ACT98655, *L. plantarum*)
EEI68452	655	8.88	LPXTG-motif cell wall anchor domain protein	Collagen_Bind super family [cl27951]: 2 repeats of Cna_B [pfam05738]	Collagen adhesion protein (EEV50121, *Enterococcus faecium*)
EEI68453	683	5.09	LPXTG-motif cell wall anchor domain protein	Collagen_Bind super family [cl27951]: 2 repeats of Cna_B [pfam05738]; GramPos_pilinBB [pfam16569]	Putative pilus subunit protein PilA (WP_002330570, *E. faecium*)
EEI68918	1069	4.61	KxYKxGKxW signal domain protein	KxYKxGKxW super family [cl14016]; LRR_5 [pfam13306]: 4 repeats of 6 Leu	Adhesion exoprotein (CAR88818, *L. rhamnosus* GG)
EEI68919	1173	4.48	LPXTG-motif cell wall anchor domain protein	LPXTG_anchor [TIGR01167]	Adhesin [BAI42720, *L. rhamnosus GG*]

NCBI accession no.	AA	pI	Protein name	Conserved domain	Potential adhesin homologue
EEI68682	931	5.24	LPXTG-motif cell wall anchor domain protein	LPXTG_anchor [TIGR01167]; Collagen_Bind super family [cl27951]: 2 repeats of Cna_B [pfam05738]	Pilus specific protein, ancillary protein involved in adhesion, SpaF (CAR88267, *L. rhamnosus* GG)
EEI68684	519	5.79	LPXTG-motif cell wall anchor domain protein	Collagen_Bind super family [cl27951]: Cna_B [pfam05738]; Fimbrial subunit type 2 [TIGR04226]; GramPos_pilinD1 super family [cl24952]	Pilus specific protein, SpaD, (CAR88265, *L. rhamnosus* GG)
EEI69168	334	5.35	LPXTG-motif cell wall anchor domain protein	LPXTG_anchor [TIGR01167]; GramPos_pilinD1 super family [cl24952]	Pilus specific protein, major backbone protein, SpaA [CAR86337, *L. rhamnosus GG*]

Table 3. Primer sets for adhesin candidate and house keeping genes from *L. paracasei.* 'F' refers to the sense primer and 'R' to the anti-sense primer for RT-qPCR amplification.

NCBI identification	Name	Oligonucleotide primers
EEI67294	Elongation factor Tu	LEFTU_F: GCGCATCAATATGTGCGTAG
		LEFTU_R: TGCTGCTCCAGAAGAAAAGG
EEI68802	GROEL chaperonin	LGROEL_F: CAAGGACAACACCACGATTG
		LGROEL_R: TTTCACGGTCGAAGTCACTG
EEI67357	Fibronectin binding protein A	LFBPA_F: AATGACACGTAGGGGATTGC
		LFBPA_R: TGCGTTACGACTCAACCAAG
EEI68137	Collagen binding protein	LCnBP_F: GCTCTTTAAACCGGTTGTGG
		LCnBP_R: CACTGTTTCGCTCAATCTGG
EEI69168	LPXTG-motif cell wall anchor domain protein	Lpara5075_F: TCTCGGGTTTAATGGCACTC
		Lpara5075_R: GATCAGTCGCACGATCAATG
EEI68918	KxYKxGKxW signal domain adhesion exoprotein	Lpara28086_F: ATTCCGACGCTGACAATACC
		Lpara28086_R: CCTGAAAGCCCAATGTTGTC
EEI68919	Adhesion exoprotein	Lpara31701_F: TGAAGCGTCAGGTCTCAATG
		Lpara31701_R: CCGTCAACAACAATGTCCTG
EEI68842	Outer membrane protein	Lpara8786_F: TCAACCACTGCAAAGTCCAC
		Lpara8786_R: ATCACTGGCGAATTCCTGAC
EEI68395	Collagen adhesion protein	Lpara8692_F: AGTAGCGTTGAATGGGATGG
		Lpara8692_R: ATACCGGAACCCCAATAAGG
EEI68452	LPXTG-motif cell wall anchor domain protein	Lpara48083_F: TTGAGACGACTGCACCAAAG
		Lpara48083_R: CGTGAGTGCATTGGTATTGG
EEI68453	LPXTG-motif cell wall anchor domain protein	Lpara50135_F: CAAGAAGTAAAGGCGCCAAC
		Lpara50135_R: TCATCGCTGGCAACATAGAC
EEI68682	LPXTG-motif cell wall anchor domain protein	Lpara2797_F: ACTGGATTACCGCGAACAAG
		Lpara2797_R: CAATCGGATCAACCCAGTTC
EEI68684	LPXTG-motif cell wall anchor domain protei	Lpara5677_F: ACCTGTCGATTCAAGGGATG
		Lpara5677_R: ATCTCAAAGTCCGTGGTTGG
AF182724	16S ribosomal RNA	LAF182724_F: ATACATAGCTGGCCGGCGGC
		LAF182724_R: CCCACTTCGCTCGCCGCTAC

Table 4. MS identification of 1 M LiCl extracted *L. paracasei* subsp. *paracasei* proteins from gel bands (samples Fpr57, Lpp35, Lpp102) and chunks (samples Lpp48 and LppW). The protein identification against the bespoke in-house database is reported, except for Lpp102 which reports the identification from the NCBI database.

Sample	Sequence coverage (%)	Number of peptides (number of unique peptides)	Protein identification
Lpp35	20	5 (5)	*L. paracasei* elongation factor Tu
Lpp35	10	4 (4)	*L. paracasei* chaperonin GroEL
Lpp48	65	58 (21)	*L. paracasei* elongation factor Tu
Lpp48	13	5 (5)	*L. paracasei* chaperonin GroEL
Fpr57	57	17 (16)	*L. paracasei* elongation factor Tu
Fpr57	33	16 (16)	*L. paracasei* chaperonin GroEL
LppW	68	45 (21)	*L. paracasei* chaperonin GroEL
LppW	64	53 (33)	*L. paracasei* elongation factor Tu
Lpp102	24	37 (36)	*L. paracasei* PII-type proteinase

Supplementary information

Identification of putative adhesins and carbohydrate ligands of *Lactobacillus paracasei* using a combinatorial *in silico* and microarray profiling approach

Benoit Houeix,[a,b] Silvia Synowsky,[c] Michael T. Cairns,[a,b] Marian Kane,[a,b] Michelle Kilcoyne,[b,d] Lokesh Joshi[a,b]

[a]Glycoscience Group, National Centre for Biomedical Engineering Science, Biomedical Sciences, National University of Ireland Galway, Galway, Ireland

[b]Advanced Glycoscience Research Cluster, National Centre for Biomedical Engineering Science, Biomedical Sciences, National University of Ireland Galway, Galway, Ireland

[c] Biomedical Sciences Research Complex, University of St Andrews, St Andrews, KY16 9ST, UK

[d]Carbohydrate Signalling Group, Discipline of Microbiology, School of Natural Sciences, National University of Ireland Galway, Galway, Ireland

Table of contents

Fig. S1. SurfG+ packages flow scheme for surface-exposed protein prediction. 3

Fig. S2. Growth curves of *Lactobacillus paracasei* subsp. *paracasei* in various media. 4

Fig. S3. Differential binding between *L. johnsonii* (Ljns), *L. rhamnosus* GG (LGG) and *L. paracasei* (Lpp) to carbohydrate microarrays. 5

Table S1. SurfG+ analysis of predicted surface exposed proteins (SEPs). 6

Table S2. *L. paracasei* RNA samples and quality control. 13

Table S3. Expression of reference gene 16S rRNA from *L. paracasei* between growth conditions. 15

Table S4. List of features printed on GIT mucin microarray. 16

Table S5. List of features printed on the carbohydrate microarray A. 18

Table S6. List of features printed on the carbohydrate microarray B. 21

Table S7. Characterised adhesins and conserved domains from human GIT commensal *Lactobacillus* species and strains. 23

Table S8. Expression of candidate adhesins from *L. paracasei.* 25

References 26

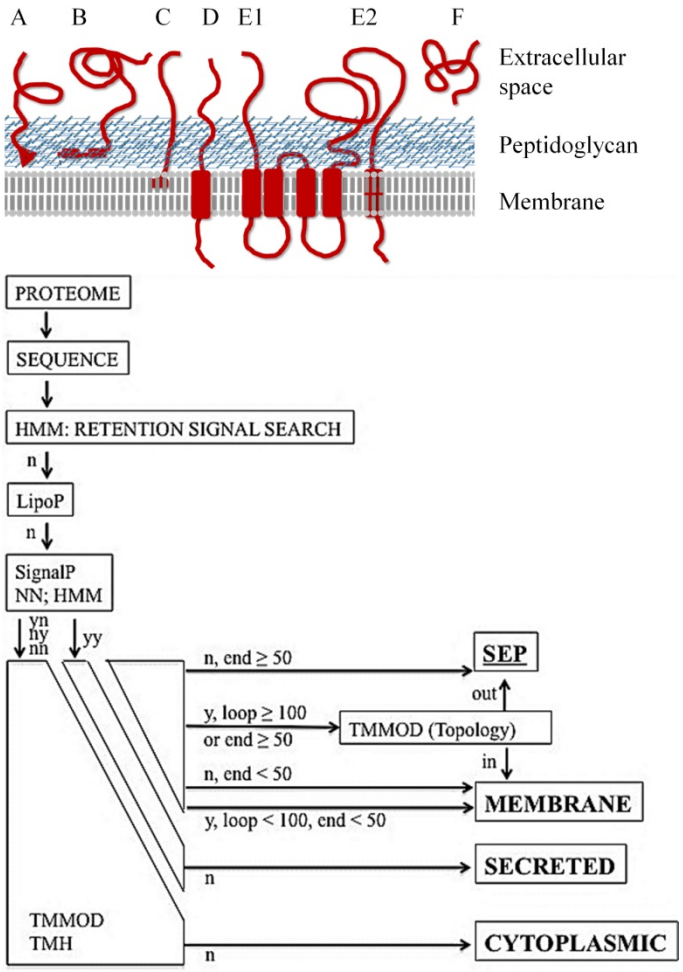

Fig. S1. SurfG+ packages flow scheme for surface-exposed protein prediction. Protein sequences were analysed through a series of logical decisions represented by the arrows and classified into four categories (Surface-exposed protein (SEP), Membrane, Secreted or Cytoplasmic). 'y' refers to a positive search for the feature and 'n' to a negative search. For SignalP, which predicts the presence and location of signal peptide cleavage sites, the first character represents the result of the neural network (NN) method and the second character the result of the Hidden Markov Model (HMM) method. Mature proteins were analysed using transmembrane proteins topology prediction (TMMOD). Commas represent the Boolean expression "AND". The 'end' and 'loop' cut off values are expressed as numbers of amino acids; 'in' are proteins predicted at the cytoplasmic side of the cell membrane; 'out' are predicted outside the cytoplasmic membrane. Types of SEPs predicted by SurfG+ are shown in red. A, surface exposed proteins (SEP) covalently bound to the cell wall; B, SEP non covalently bound to the cell wall; C, membrane-bound lipoprotein; D, membrane anchored protein through N or C-terminal trans-membrane helix (TMH); E, membrane anchored through several TMHs; E1: surface exposed N or C-terminal end; E2: surface exposed loop; F: secreted protein. Adapted from Barinov, *et al.*[1]

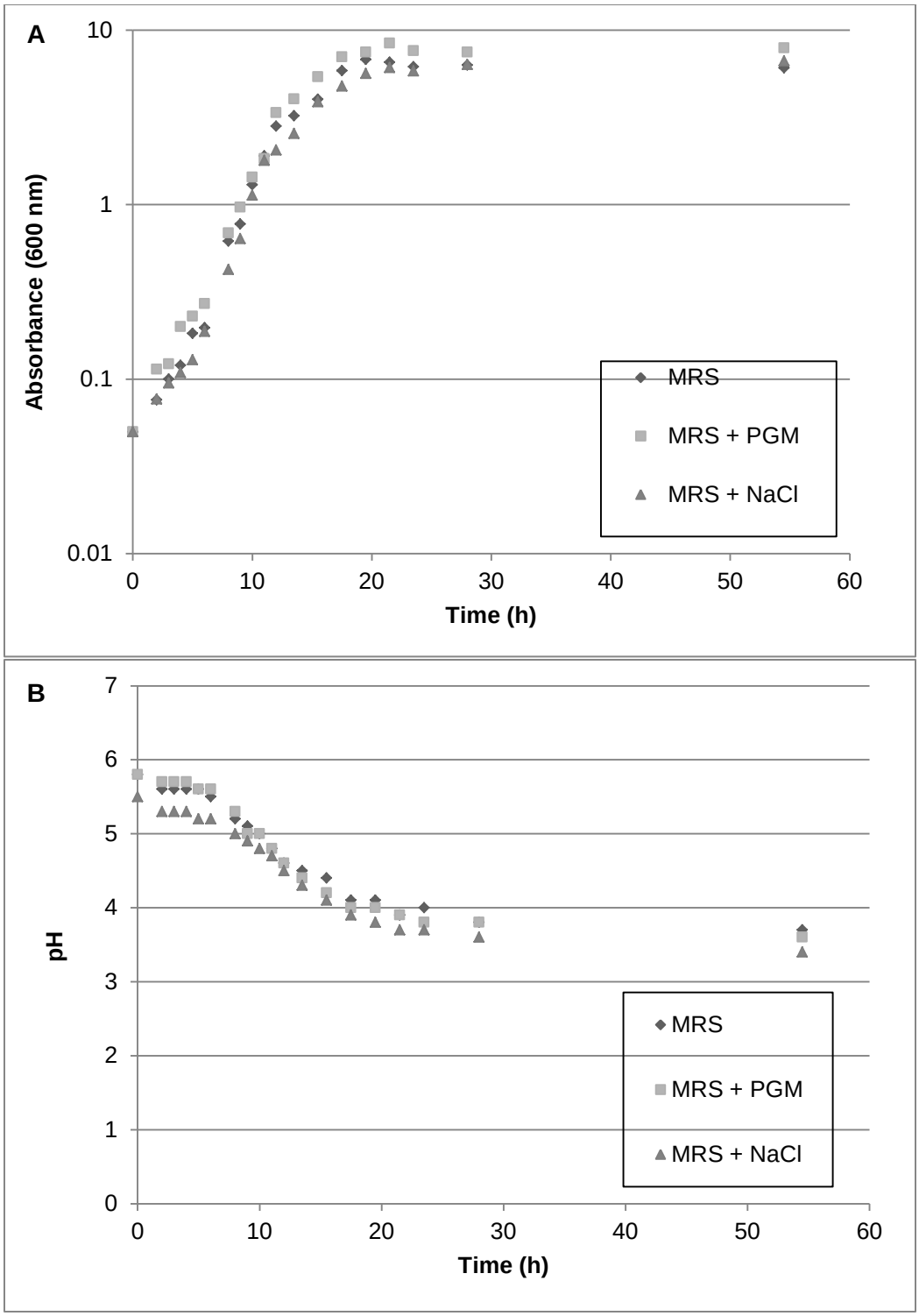

Fig. S2. Growth curves of *Lactobacillus paracasei* subsp. *paracasei* in various media. (**A**) Bacterial growth measured by absorbance at 600 nm over 54.5 h in MRS media (MRS), MRS supplemented with 0.1% porcine gastric mucin (PGM) (MRS + PGM), MRS supplemented with 0.3 M NaCl (MRS + NaCl). (**B**) Culture supernatant pH measured at corresponding time-points.

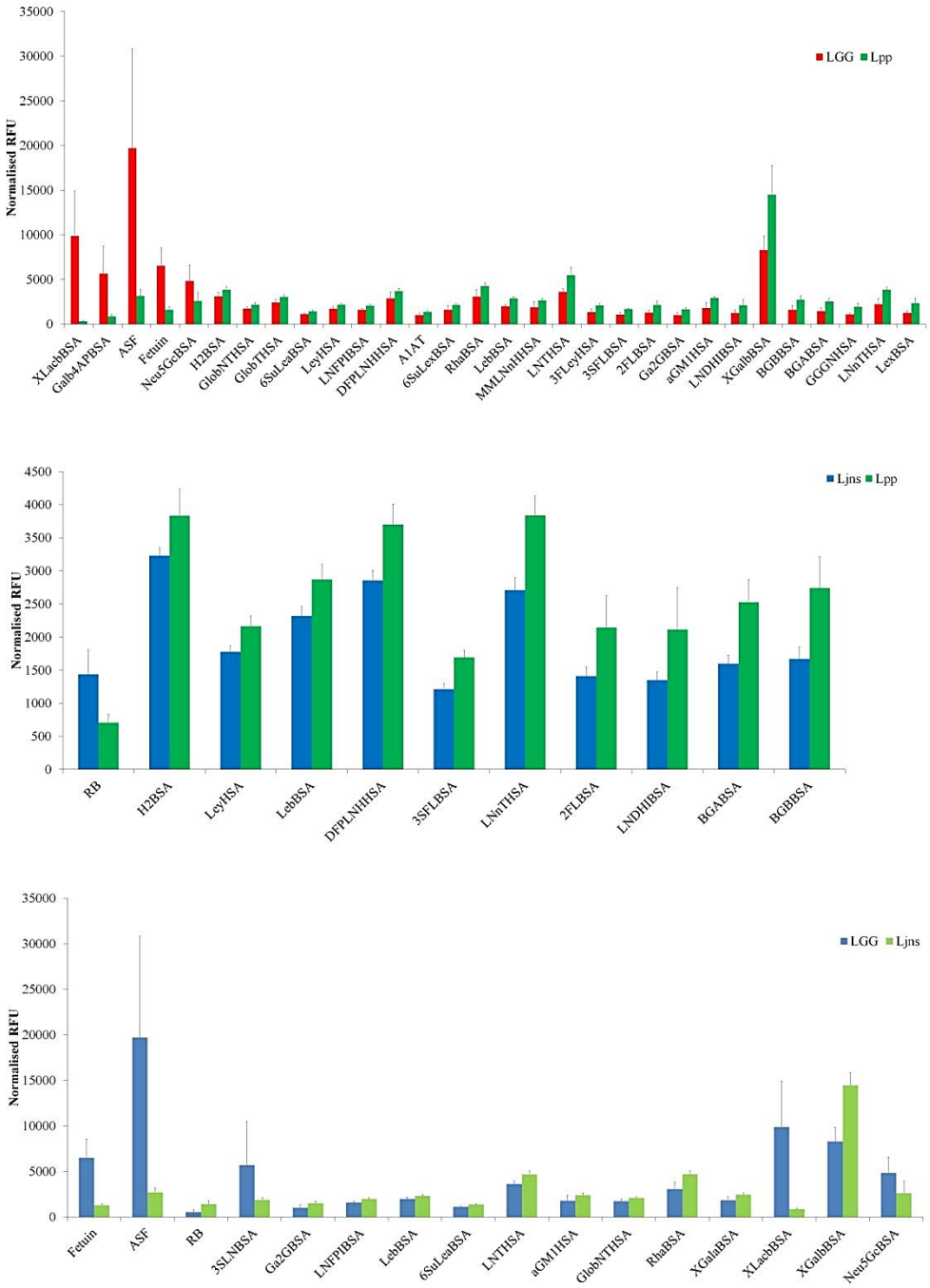

Figure S3. Differential binding between *L. johnsonii* (Ljns), *L. rhamnosus* GG (LGG) and *L. paracasei* (Lpp) to carbohydrate microarrays. Neoglycoconjugates showing significant differences in binding when each pair of lactobacilli species were compared are depicted in bar charts. All signals are averages from three microarrays following scale normalisation. Error bars represent the standard deviation of the mean. Top bar chart: pairwise comparison between LGG and Lpp. Middle bar chart: pairwise comparison between Ljns and Lpp. Bottom bar chart: pairwise comparison between LGG) and Ljns.

Table S1. SurfG+ analysis of predicted surface exposed proteins (SEPs), with NCBI accession entry, a categorisation and number of amino acids (AA) in the predicted SEP. Categorisation is for the predicted topology of the SEP in SurfG+ with L indicating surface exposed loop, Ct indicating transmembrane helix with surface exposed C-terminal, Nt indicating transmembrane helix with surface exposed N-terminal, and r indicating protein with retention membrane signal.

NCBI accession entry	Description	Category	AA
gi|227532720|ref|ZP_03962769.1|	possible sensor histidine kinase	Nt	303
gi|227532730|ref|ZP_03962779.1|	membrane associated transglutaminase family enzyme	L	720
gi|227532757|ref|ZP_03962806.1|	PTS family mannose/fructose/sorbose porter component IID	L	303
gi|227532779|ref|ZP_03962828.1|	histidine kinase	Nt	636
gi|227532780|ref|ZP_03962829.1|	yycH protein	Ct	478
gi|227532781|ref|ZP_03962830.1|	conserved hypothetical protein	Ct	265
gi|227532784|ref|ZP_03962833.1|	S1C family peptidase	Ct	442
gi|227532789|ref|ZP_03962838.1|	conserved hypothetical protein	Ct	420
gi|227532810|ref|ZP_03962859.1|	endonuclease/exonuclease/phosphatase	Ct	355
gi|227532820|ref|ZP_03962869.1|	possible glycosyl hydrolase	r	620
gi|227532855|ref|ZP_03962904.1|	protein-N(pi)-phosphohistidine--sugar phosphotransferase	Nt	372
gi|227532868|ref|ZP_03962917.1|	alpha-glucosidase	rNt	573
gi|227532873|ref|ZP_03962922.1|	signal transduction histidine kinase	Ct	443
gi|227532878|ref|ZP_03962927.1|	hypothetical membrane protein	Ct	654
gi|227532897|ref|ZP_03962946.1|	conserved hypothetical protein	Ct	250
gi|227532947|ref|ZP_03962996.1|	membrane protein	Nt	225
gi|227532967|ref|ZP_03963016.1|	beta-propeller domain of methanol dehydrogenase	Nt	286
gi|227532991|ref|ZP_03963040.1|	signal transduction histidine kinase regulating citrate/malate metabolism	Ct	303
gi|227533018|ref|ZP_03963067.1|	integral membrane protein	Ct	318
gi|227533038|ref|ZP_03963087.1|	ErfK family cell surface protein	Ct	234
gi|227533041|ref|ZP_03963090.1|	ErfK family protein	Ct	483
gi|227533047|ref|ZP_03963096.1|	septum formation initiator	Ct	143
gi|227533051|ref|ZP_03963100.1|	M41 family endopeptidase FtsH	Nt	715
gi|227533055|ref|ZP_03963104.1|	cell surface protein	Nt	203
gi|227533056|ref|ZP_03963105.1|	cell surface protein	Nt	352
gi|227533103|ref|ZP_03963152.1|	beta-N-acetylglucosaminidase precursor	Nt	569
gi|227533123|ref|ZP_03963172.1|	oligopeptide ABC superfamily ATP binding cassette transporter substrate binding protein	E	546
gi|227533152|ref|ZP_03963201.1|	succinate dehydrogenase	E	508
gi|227533164|ref|ZP_03963213.1|	PTS family mannose/fructose/sorbose porter component IID	Nt	281

NCBI accession entry	Description	Category	AA
gi|227533175|ref|ZP_03963224.1|	protein-N(pi)-phosphohistidine-sugar phosphotransferase	Ct	654
gi|227533187|ref|ZP_03963236.1|	autolysin	r	255
gi|227533201|ref|ZP_03963250.1|	ABC superfamily ATP binding cassette transporter	Ct	603
gi|227533204|ref|ZP_03963253.1|	hypothetical protein HMPREF0530_0490	Nt	128
gi|227533205|ref|ZP_03963254.1|	phage infection protein	L	910
gi|227533249|ref|ZP_03963298.1|	possible yvcC protein	Nt	303
gi|227533250|ref|ZP_03963299.1|	Bee2 protein	rNt	238
gi|227533251|ref|ZP_03963300.1|	LPXTG-motif cell wall anchor domain protein	rNt	334
gi|227533252|ref|ZP_03963301.1|	sortase	Ct	348
gi|227533261|ref|ZP_03963310.1|	cell envelope-associated proteinase	rNt	2232
gi|227533262|ref|ZP_03963311.1|	cell envelope-associated proteinase	L	1783
gi|227533282|ref|ZP_03963331.1|	conserved hypothetical protein	Nt	235
gi|227533283|ref|ZP_03963332.1|	conserved hypothetical protein	Ct	212
gi|227533303|ref|ZP_03963352.1|	possible TrsG protein	Ct	825
gi|227533304|ref|ZP_03963353.1|	hypothetical protein HMPREF0530_0590	Ct	174
gi|227533306|ref|ZP_03963355.1|	conserved hypothetical protein	Ct	393
gi|227533338|ref|ZP_03963387.1|	conserved hypothetical protein	Ct	122
gi|227533341|ref|ZP_03963390.1|	phage infection protein	L	981
gi|227533344|ref|ZP_03963393.1|	phage infection protein	L	869
gi|227533354|ref|ZP_03963403.1|	cell surface protein	Nt	359
gi|227533357|ref|ZP_03963406.1|	conserved hypothetical protein	Nt	119
gi|227533397|ref|ZP_03963446.1|	ABC superfamily ATP binding cassette transporter	Nt	924
gi|227533403|ref|ZP_03963452.1|	protein-N(pi)-phosphohistidine-sugar phosphotransferase	Ct	663
gi|227533407|ref|ZP_03963456.1|	DHH family phosphoesterase	Ct	668
gi|227533429|ref|ZP_03963478.1|	conserved hypothetical protein	Nt	212
gi|227533430|ref|ZP_03963479.1|	conserved hypothetical protein	L	410
gi|227533442|ref|ZP_03963491.1|	indole-3-glycerol phosphate synthase (IGPS)	Nt	209
gi|227533462|ref|ZP_03963511.1|	RND superfamily resistance-nodulation-cell division:proton (H+) antiporter	Nt	1103
gi|227533475|ref|ZP_03963524.1|	conserved hypothetical protein	Ct	114
gi|227533485|ref|ZP_03963534.1|	ABC superfamily ATP binding cassette transporter permease protein	Nt	857
gi|227533494|ref|ZP_03963543.1|	cytochrome bd-I ubiquinol oxidase subunit I	L	498

NCBI accession entry	Description	Category	AA
gi|227533522|ref|ZP_03963571.1|	adhesion exoprotein	Ct	1069
gi|227533523|ref|ZP_03963572.1|	Adhesion exoprotein	rNt	1173
gi|227533548|ref|ZP_03963597.1|	ABC superfamily ATP binding cassette transporter membrane permease protein	Ct	282
gi|227533558|ref|ZP_03963607.1|	enolase 2 (2-phosphoglycerate dehydratase 2)	Nt	86
gi|227533566|ref|ZP_03963615.1|	possible cell surface protein	L	729
gi|227533571|ref|ZP_03963620.1|	possible outer membrane protein	Nt	611
gi|227533618|ref|ZP_03963667.1|	lactococcin A ABC superfamily ATP binding cassette transporter permease protein	Ct	465
gi|227533621|ref|ZP_03963670.1|	integral membrane protein	Nt	451
gi|227533631|ref|ZP_03963680.1|	possible outer membrane protein	rNt	931
gi|227533632|ref|ZP_03963681.1|	conserved hypothetical protein	Nt	440
gi|227533633|ref|ZP_03963682.1|	pilus subunit protein	Nt	519
gi|227533636|ref|ZP_03963685.1|	conserved hypothetical protein	Ct	407
gi|227533640|ref|ZP_03963689.1|	permease of the major facilitator superfamily protein	Nt	410
gi|227533655|ref|ZP_03963704.1|	conserved hypothetical protein	L	398
gi|227533656|ref|ZP_03963705.1|	conserved hypothetical protein	L	366
gi|227533666|ref|ZP_03963715.1|	conserved hypothetical protein	rL	432
gi|227533670|ref|ZP_03963719.1|	transcriptional regulator	Nt	276
gi|227533690|ref|ZP_03963739.1|	possible secreted protein	Ct	161
gi|227533692|ref|ZP_03963741.1|	pheromone cAD1 precursor lipoprotein	E	304
gi|227533693|ref|ZP_03963742.1|	ApbE protein	Ct	379
gi|227533702|ref|ZP_03963751.1|	conserved hypothetical protein	E	133
gi|227533717|ref|ZP_03963766.1|	peptidylprolyl isomerase	E	299
gi|227533718|ref|ZP_03963767.1|	PII family proteinase	rNt	1902
gi|227533724|ref|ZP_03963773.1|	conserved hypothetical protein	Ct	122
gi|227533747|ref|ZP_03963796.1|	conserved hypothetical protein	Nt	342
gi|227533754|ref|ZP_03963803.1|	APC family amino acid-polyamine-organocation transporter	Nt	611
gi|227533760|ref|ZP_03963809.1|	conserved hypothetical protein	Ct	164
gi|227533762|ref|ZP_03963811.1|	hemolysin	Ct	466
gi|227533791|ref|ZP_03963840.1|	ysaD protein	Ct	124
gi|227533792|ref|ZP_03963841.1|	conserved hypothetical protein	Ct	402
gi|227533794|ref|ZP_03963843.1|	conserved hypothetical protein	E	384
gi|227533814|ref|ZP_03963863.1|	ABC superfamily ATP binding cassette transporter permease protein	Nt	423
gi|227533815|ref|ZP_03963864.1|	ABC superfamily ATP binding cassette transporter ATP binding protein	Ct	353

NCBI accession entry	Description	Category	AA
gi|227533834|ref|ZP_03963883.1|	flotillin	Ct	505
gi|227533849|ref|ZP_03963898.1|	sortase	Ct	235
gi|227533859|ref|ZP_03963908.1|	membrane protein	Nt	379
gi|227533861|ref|ZP_03963910.1|	conserved hypothetical protein	Ct	186
gi|227533934|ref|ZP_03963983.1|	capsular polysaccharide biosynthesis protein	L	310
gi|227533951|ref|ZP_03964000.1|	collagen adhesion protein	Nt	2240
gi|227534006|ref|ZP_03964055.1|	conserved hypothetical protein	E	261
gi|227534008|ref|ZP_03964057.1|	Cna B domain protein	rNt	655
gi|227534009|ref|ZP_03964058.1|	cell wall surface anchor family protein	Nt	683
gi|227534010|ref|ZP_03964059.1|	possible sortase	L	278
gi|227534011|ref|ZP_03964060.1|	conserved hypothetical protein	Nt	287
gi|227534014|ref|ZP_03964063.1|	conserved hypothetical protein	Ct	415
gi|227534020|ref|ZP_03964069.1|	TRAG protein precursor	Ct	849
gi|227534021|ref|ZP_03964070.1|	conserved hypothetical protein	L	767
gi|227534022|ref|ZP_03964071.1|	hypothetical protein HMPREF0530_1308	Ct	111
gi|227534025|ref|ZP_03964074.1|	possible extracellular protein precursor	Ct	325
gi|227534026|ref|ZP_03964075.1|	hypothetical protein HMPREF0530_1312	E	208
gi|227534027|ref|ZP_03964076.1|	conserved hypothetical protein	Ct	270
gi|227534041|ref|ZP_03964090.1|	UDP-N-acetylglucosamine:LPS N-acetylglucosamine transferase	Nt	132
gi|227534057|ref|ZP_03964106.1|	sugar transferase	Ct	234
gi|227534095|ref|ZP_03964144.1|	undecaprenyl-phosphate galactosephosphotransferase	Ct	466
gi|227534096|ref|ZP_03964145.1|	conserved hypothetical protein	E	288
gi|227534100|ref|ZP_03964149.1|	conserved hypothetical protein	L	495
gi|227534102|ref|ZP_03964151.1|	conserved hypothetical protein	L	579
gi|227534103|ref|ZP_03964152.1|	conserved hypothetical protein	L	557
gi|227534108|ref|ZP_03964157.1|	MPA1 family cytoplasmic membrane-periplasmic auxiliary transporter	Nt	230
gi|227534111|ref|ZP_03964160.1|	possible membrane-oligosaccharide glycerophosphotransferase	Ct	774
gi|227534117|ref|ZP_03964166.1|	peptide ABC superfamily ATP binding cassette transporter permease	L	602
gi|227534134|ref|ZP_03964183.1|	conserved hypothetical protein	Ct	192
gi|227534188|ref|ZP_03964237.1|	endolysin	rCt	382
gi|227534192|ref|ZP_03964241.1|	conserved hypothetical protein	rCt	537
gi|227534224|ref|ZP_03964273.1|	conserved hypothetical protein	Nt	192
gi|227534250|ref|ZP_03964299.1|	cell surface protein	Nt	344
gi|227534273|ref|ZP_03964322.1|	hypothetical protein HMPREF0530_1559	Nt	159

NCBI accession entry	Description	Category	AA
gi|227534278|ref|ZP_03964327.1|	collagen binding protein	E	270
gi|227534285|ref|ZP_03964334.1|	cell surface protein	Nt	339
gi|227534286|ref|ZP_03964335.1|	conserved hypothetical protein	rNt	125
gi|227534290|ref|ZP_03964339.1|	conserved hypothetical protein	rCt	129
gi|227534316|ref|ZP_03964365.1|	nucleoside-diphosphate-sugar epimerase	Ct	254
gi|227534334|ref|ZP_03964383.1|	DNA-entry nuclease	Ct	272
gi|227534386|ref|ZP_03964435.1|	ATP-binding cassette transporter	E	318
gi|227534387|ref|ZP_03964436.1|	ATP-binding cassette transporter	E	322
gi|227534394|ref|ZP_03964443.1|	glycosyltransferase	Ct	716
gi|227534397|ref|ZP_03964446.1|	possible histidine kinase	L	431
gi|227534415|ref|ZP_03964464.1|	IS5 family transposase	Nt	472
gi|227534427|ref|ZP_03964476.1|	conserved hypothetical protein	E	157
gi|227534441|ref|ZP_03964490.1|	ABC superfamily ATP binding cassette transporter ATPase and permease protein	Ct	530
gi|227534456|ref|ZP_03964505.1|	conserved hypothetical protein	L	325
gi|227534459|ref|ZP_03964508.1|	conserved hypothetical protein	Ct	153
gi|227534505|ref|ZP_03964554.1|	penicillin-binding protein 1B	Ct	927
gi|227534512|ref|ZP_03964561.1|	multitransmembrane protein	Nt	335
gi|227534518|ref|ZP_03964567.1|	membrane protein	Nt	207
gi|227534547|ref|ZP_03964596.1|	conserved hypothetical protein	Ct	210
gi|227534548|ref|ZP_03964597.1|	beta-propeller domains of methanol dehydrogenase	Ct	447
gi|227534606|ref|ZP_03964655.1|	type I signal peptidase family protein	Ct	199
gi|227534613|ref|ZP_03964662.1|	peptidoglycan-binding protein	rNt	716
gi|227534614|ref|ZP_03964663.1|	conserved hypothetical protein	Ct	186
gi|227534674|ref|ZP_03964723.1|	endolysin	r	382
gi|227534713|ref|ZP_03964762.1|	acetyl-CoA C-acetyltransferase	Nt	393
gi|227534727|ref|ZP_03964776.1|	conserved hypothetical protein	Nt	116
gi|227534761|ref|ZP_03964810.1|	peptide ABC superfamily ATP binding cassette transporter permease	L	691
gi|227534766|ref|ZP_03964815.1|	glycerophosphoryl diester phosphodiesterase	Ct	604
gi|227534774|ref|ZP_03964823.1|	penicillin-binding protein	Ct	698
gi|227534778|ref|ZP_03964827.1|	peptidylprolyl isomerase	E	300
gi|227534814|ref|ZP_03964863.1|	GTP-binding protein	Nt	374
gi|227534825|ref|ZP_03964874.1|	sensor histidine kinase	L	528
gi|227534826|ref|ZP_03964875.1|	Oxa1 family cytochrome oxidase biogenesis protein	Nt	332
gi|227534834|ref|ZP_03964883.1|	aminodeoxychorismate lyase	Ct	400
gi|227534841|ref|ZP_03964890.1|	membrane protein	L	868
gi|227534881|ref|ZP_03964930.1|	possible non-specific serine/threonine protein kinase	Ct	666
gi|227534886|ref|ZP_03964935.1|	alkaline shock protein	Ct	132

NCBI accession entry	Description	Category	AA
gi|227534895|ref|ZP_03964944.1|	oligopeptide ABC trasporter substrate binding protein	E	597
gi|227534913|ref|ZP_03964962.1|	possible xenobiotic-transporting ATPase	Ct	615
gi|227534927|ref|ZP_03964976.1|	M50 family peptidase	L	413
gi|227534961|ref|ZP_03965010.1|	conserved hypothetical protein	Ct	141
gi|227535016|ref|ZP_03965065.1|	conserved hypothetical protein	L	347
gi|227535032|ref|ZP_03965081.1|	peptidoglycan glycosyltransferase	Ct	770
gi|227535049|ref|ZP_03965098.1|	cardiolipin synthetase	Ct	489
gi|227535098|ref|ZP_03965147.1|	peptide ABC superfamily ATP binding cassette transporter permease	L	663
gi|227535121|ref|ZP_03965170.1|	Carboxy-terminal processing proteinase	r	390
gi|227535125|ref|ZP_03965174.1|	conserved hypothetical protein	Ct	212
gi|227535140|ref|ZP_03965189.1|	LysM domain protein	rCt	206
gi|227535156|ref|ZP_03965205.1|	conserved hypothetical protein	E	303
gi|227535165|ref|ZP_03965214.1|	acyltransferase	Ct	664
gi|227535190|ref|ZP_03965239.1|	possible competence protein EC	L	734
gi|227535249|ref|ZP_03965298.1|	cardiolipin synthetase	Ct	477
gi|227535267|ref|ZP_03965316.1|	Septation ring formation regulator ezrA	Ct	577
gi|227535290|ref|ZP_03965339.1|	conserved hypothetical protein	Ct	451
gi|227535293|ref|ZP_03965342.1|	signal transduction histidine kinase	Ct	370
gi|227535300|ref|ZP_03965349.1|	two-component sensor kinase	L	542
gi|227535373|ref|ZP_03965422.1|	possible xenobiotic-transporting ATPase	Ct	628
gi|227535390|ref|ZP_03965439.1|	conserved hypothetical protein	Ct	291
gi|227535391|ref|ZP_03965440.1|	DNA-entry nuclease	E	283
gi|227535403|ref|ZP_03965452.1|	phosphatidylglycerol--membrane-oligosaccharide glycerophosphotransferase	Ct	725
gi|227535404|ref|ZP_03965453.1|	conserved hypothetical protein	Nt	228
gi|227535415|ref|ZP_03965464.1|	membrane protein	Ct	323
gi|227535416|ref|ZP_03965465.1|	conserved hypothetical protein	E	212
gi|227535418|ref|ZP_03965467.1|	conserved hypothetical protein	L	606
gi|227535420|ref|ZP_03965469.1|	membrane protein	L	869
gi|227535447|ref|ZP_03965496.1|	conserved hypothetical protein	Ct	118
gi|227535448|ref|ZP_03965497.1|	extracellular glycosyl hydrolase	Ct	261
gi|227535465|ref|ZP_03965514.1|	Sex pheromone staph-cAM373	E	383
gi|227535469|ref|ZP_03965518.1|	acyl-CoA thioesterase family protein	E	263
gi|227535470|ref|ZP_03965519.1|	membrane protein	Ct	398
gi|227535479|ref|ZP_03965528.1|	possible histidine kinase	Ct	491
gi|227535494|ref|ZP_03965543.1|	competence protein ComGC	Ct	153
gi|227535497|ref|ZP_03965546.1|	competence protein ComGC	Ct	153
gi|227535501|ref|ZP_03965550.1|	integral membrane protein	Nt	163
gi|227535574|ref|ZP_03965623.1|	conserved hypothetical protein	Nt	236
gi|227535594|ref|ZP_03965643.1|	signal transduction histidine kinase	Nt	555

NCBI accession entry	Description	Category	AA
gi|227535597|ref|ZP_03965646.1|	cell divison ABC superfamily ATP binding cassette transporter FtsX	L	295
gi|227535610|ref|ZP_03965659.1|	possible transcription regulation protein	Ct	309
gi|227535614|ref|ZP_03965663.1|	FtsK/SpoIIIE family DNA translocase	Ct	799
gi|227535626|ref|ZP_03965675.1|	thiamine biosynthesis lipoprotein	E	352
gi|227535650|ref|ZP_03965699.1|	conserved hypothetical protein	Ct	241
gi|227535658|ref|ZP_03965707.1|	lysin	r	432
gi|227535659|ref|ZP_03965708.1|	holin	Ct	157
gi|227535717|ref|ZP_03965766.1|	conserved hypothetical protein	rNt	220
gi|227535729|ref|ZP_03965778.1|	phosphatidylglycerol-membrane-oligosaccharide glycerophosphotransferase	Ct	703
gi|227535735|ref|ZP_03965784.1|	conserved hypothetical protein	Ct	329

Table S2. *L. paracasei* RNA sample numbers and descriptions, yields of RNA preparations, ratio of absorbance at 260 nm/ 280 nm and RNA integrity number (R.I.N.).

Sample	Description	Yield (µg)	260/280	R.I.N.
1	Calibrator controls (lag phase, regular broth)	8.79	2.2	10
2	Calibrator controls (lag phase, regular broth)	7.10	2.09	10
3	Calibrator controls (lag phase, regular broth)	12.47	2.15	10
4	Regular media (log phase)	11.40	2.15	10
5	Regular media (log phase)	11.23	2.12	10
6	Regular media (log phase)	14.29	2.14	10
7	Broth supplemented with PGM 0.1% (log phase)	22.30	2.14	10
8	Broth supplemented with PGM 0.1% (log phase)	16.27	2.14	10
9	Broth supplemented with PGM 0.1% (log phase)	19.21	2.14	9.9
10	Media supplemented with 0.3 M NaCl (log phase)	12.88	2.11	10
11	Media supplemented with 0.3 M NaCl (log phase)	13.94	2.09	10
12	Media supplemented with 0.3 M NaCl (log phase)	8.44	2.19	7.4
13	Broth supplemented with both PGM and NaCl (log phase)	4.96	2.04	10
14	Broth supplemented with both PGM and NaCl (log phase)	10.16	2.12	10
15	Broth supplemented with both PGM and NaCl (log phase)	10.81	2.13	10

Table S3. Expression of reference gene 16S rRNA from *L. paracasei* between growth conditions.

	L EFTU	L GROEL	L FBPA	L CnBP	L 28086	L 57075	L 50135	L 31701	L 8786	L 8692	L5677	L 2797	L 48083
CONTROL	10.17	9.77	9.71	9.66	10.79	9.77	9.76	9.65	9.30	9.34	9.73	9.61	9.66
CONTROL	9.70	9.26	10.90	11.21	11.32	10.91	10.96	11.49	10.86	10.79	12.31	11.09	10.89
CONTROL	11.15	10.86	10.79	10.97	10.74	10.77	10.44	10.82	10.68	11.00	10.67	10.87	10.86
REGULAR	11.62	11.54	12.71	12.73	12.82	12.23	14.03	12.33	12.62	12.59	12.70	13.00	12.25
REGULAR	10.39	10.14	10.43	10.30	10.96	10.45	11.02	12.90	10.19	10.42	10.56	10.65	10.58
REGULAR	9.65	9.84	10.21	9.66	10.67	10.38	10.24	10.28	10.08	9.92	10.11	10.24	11.60
M	11.57	11.34	10.90	10.88	10.54	10.90	11.01	10.80	10.65	10.69	10.59	10.62	10.55
M	11.39	11.18	10.83	10.91	10.99	10.88	10.58	10.60	10.36	10.22	10.40	10.67	10.65
M	10.15	9.68	11.03	10.90	11.60	10.74	11.02	11.06	10.98	11.09	10.98	11.01	11.08
N	9.91	9.46	11.55	11.21	11.72	11.27	11.06	11.28	11.43	11.07	11.38	11.54	11.14
N	11.53	11.44	9.63	8.89	9.94	9.27	9.77	9.75	9.24	9.14	9.41	9.39	9.29
N	11.66	11.23	11.09	10.54	11.11	10.98	11.18	11.30	11.02	11.14	12.12	10.81	10.89
M+N	11.24	11.15	8.44	8.14	8.56	8.59	8.52	8.44	8.37	8.38	8.59	8.55	8.55
M+N	11.18	11.08	12.28	12.47	12.71	12.35	12.25	12.31	12.49	12.09	12.39	12.85	12.25
M+N	13.19	11.04	11.16	10.76	11.17	11.27	10.94	11.19	10.99	11.08	10.98	11.03	10.93

Table S3. Expression of candidate adhesins from *L. paracasei* subsp. *paracasei* indicating fold change with standard deviation (sd) in expression of candidate genes normalised with 16S rRNA reference gene and with lag phase used as the calibrator, between MRS media, MRS + PGM, MRS + NaCl or MRS + PGM/NaCl. Significant data are coloured and underlined in red.

Canditate adhesion	**Fold change**							
NCBI entry	**MRS**	**sd**	**MRS+PGM**	**sd**	**MRS+NaCl**	**sd**	**MRS+PGM / NaCl**	**sd**
EEI67294.1	1.03	0.31	1.17	0.03	1.22	0.05	1.59	0.33
EEI68802.1	1.71	0.79	1.25	0.29	0.53	0.38	1.85	1.46
EEI67357.1	1.25	0.84	0.91	0.31	1.49	1.12	1.16	0.23
EEI68137.1	1.63	0.48	1.63	0.3	1.28	0.7	1.19	0.23
EEI69168.1	1.43	0.26	1.42	0.27	1.10	0.2	2.58	0.89
EEI68918.1	2.08	2.23	1.22	0.34	2.35	1.94	1.58	1.53
EEI68919.1	1.50	0.55	1.04	0.25	1.26	0.42	1.42	0.26
EEI68842.1	1.26	0.5	1.20	0.37	0.93	0.18	1.75	0.65
EEI68395.1	1.40	0.73	0.69	0.41	1.04	1.29	1.50	1.29
EEI68452.1	1.16	0.21	1.61	0.64	1.19	0.52	2.51	1.9
EEI68453.1	1.06	0.13	1.11	0.16	1.14	0.39	1.71	0.71
EEI68682.1	1.02	0.34	0.86	0.07	1.11	0.49	2.54	1.02
EEI68684.1	1.72	1.4	0.94	0.2	1.09	0.88	2.61	1.71
Mean	1.40	0.67	1.16	0.28	1.21	0.66	1.85	0.94

Table S4. Print list for the GIT mucin microarray with sample identification (ID), sample description, print concentration (mg/mL) and details of print buffer of PBS supplemented with various concentrations of Tween 20 (% T). The same colour indicates the same species.

Sample ID	Sample description	Concentration (mg/mL)	Print buffer
6	Equine stomach mucin	0.25	PBS 0.01% T
10	Ovine abomasum antrum mucin	0.1	PBS 0.01% T
11	E12 cell line mucin	0.5	PBS 0.025% T
12	Ovine descending colon mucin	0.15	PBS 0.01% T
13	Ovine ileum mucin	0.15	PBS 0.01% T
18	Ovine spiral colon mucin	0.5	PBS
71	Chicken proximal small intestine mucin	0.25	PBS
35	Ovine jejunum mucin	0.5	PBS 0.01% T
36	Ovine duodenum mucin	0.15	PBS 0.01% T
37	Porcine gastric mucin	0.33	PBS 0.01% T
41	Chicken large intestine mucin	0.2	PBS 0.01% T
52	Equine duodenum mucin	0.3	PBS 0.01% T
55	Deer jejunum mucin	0.25	PBS 0.025% T
56	Deer large intestine mucin	0.75	PBS 0.025% T
57	Bovine abomasum mucin	0.25	PBS 0.01% T
58	Bovine duodenum mucin	0.5	PBS 0.01% T
59	Equine small intestine mucin	0.25	PBS
60	Equine left ventral colon mucin	0.25	PBS 0.01% T
61	Bovine spiral colon mucin	0.25	PBS 0.01% T
62	Deer duodenum mucin	0.5	PBS 0.025% T
65	Equine right ventral colon mucin	0.15	PBS 0.01% T
66	Equine dorsal colon mucin	0.25	PBS 0.01% T
67	Deer abomasum mucin	0.25	PBS 0.01% T
70	Chicken cecum mucin	0.5	PBS 0.01% T
204	LS174T cell line mucin	0.5	PBS 0.01% T
85	Porcine descending colon mucin	0.6	PBS 0.025% T
86	Porcine jejunum mucin	0.25	PBS
87	Porcine spiral colon mucin	0.6	PBS 0.025% T
102	Porcine stomach mucin	0.5	PBS 0.025% T
103	Porcine ceca mucin	0.5	PBS 0.025% T
105	Mouse large intestine	0.4	PBS 0.025% T
106	Mouse cecum mucin	0.3	PBS 0.025% T
107	Mouse stomach mucin	0.5	PBS 0.025% T
121	Mouse small intestine mucin	0.25	PBS 0.025% T
139	Rat ileum mucin	0.5	PBS 0.025% T
140	Rat duodenum mucin	0.5	PBS 0.025% T
141	Rat cecum mucin	0.5	PBS 0.025% T
146	Rat stomach mucin	0.5	PBS 0.025% T
147	Rat colon mucin	0.75	PBS 0.01% T

Sample ID	Sample description	Concentration (mg/mL)	Print buffer
ASF	Asialofetuin glycoprotein	0.5	PBS
RB	RNase B glycoprotein	0.5	PBS
Fetuin	Fetuin glycoprotein	0.5	PBS
Xferrin	Transferrin glycoprotein	0.5	PBS
Ovomuc	Ovomucoid glycoprotein	0.5	PBS
PEW	Pigeon egg white glycoprotein	0.5	PBS
PBS	PBS		PBS
PBST	PBS 0.025% Tween 20		PBS 0.025% T

Table S5. List of features printed on carbohydrate microarray A.

Row	Probe	Abbreviation	Structure
1	Fetuin	Fetuin	Fetuin
2	Asialofetuin	ASF	ASF
3	PBS	PBS	PBS
4	Ovalbumin	Ov	Ovalbumin
5	RNAse B	RB	RNAseB
6	Transferrin	Xferrin	Transferrin
7	4AP-HSA	4APHSA	4-aminopyridine-HSA
8	α-Crystallin from bovine lens	a-C	α-crystalline (*Bos taurus*)
9	Mana1,3(Mana1,6)Man-BSA	M3BSA	Man-α-(1→3)-[Man-α-(1→6)]-Man-BSA
10	GlcNAc-BSA	GlcNAcBSA	GlcNAc-Sp14-NH_2(Lys)-BSA
11	LacNAc-BSA	LacNAcBSA	Gal-β-(1→4)-GlcNAc-Sp3-BSA
12	3'SialylLacNAc-BSA	3SLNBSA	Neu5Ac-α-(2→3)-Gal-β-(1→4)-GlcNAc-APD-HSA
13	3´-Sialyllactose-APD-HAS	3SLacHSA	Neu5Ac-α-(2→3)-Gal-β-(1→4)-Glc-APD-HSA
14	6´-Sialyllactose-APD-HSA	6SLacHSA	Neu5Ac-α-(2→6)-Gal-β-(1→4)-Glc-APD-HSA
15	2'Fucosyllactose-BSA	2FLBSA	Fuc-α-(1→2)-Gal-β-(1→4)-Glc-Sp3-BSA
16	3'Sialyl-3-fucosyllactose-BSA	3SFLBSA	Neu5Ac-α-(2→3)-Gal-β-(1→4)[Fuc-α-(1→3)]-Glc-Sp3-BSA
17	H Type II-APE-BSA	H2BSA	Fuc-α-(1→2)-Gal-β-(1→4)-GlcNAc-β1-APE-BSA
18	Blood Group A-BSA	BGABSA	GalNAc-α-(1→3)-[Fuc-α-(1→2)]-Gal
19	Blood Group B-BSA	BGBBSA	Gal-α-(1→3)-[Fuc-α-(1→2)]-Gal
20	Gala1,3Galb1,4GlcNAc-HSA	GGGNHSA	Gal-α-(1→3)-Gal-β-(1→4)-GlcNAc-HSA
21	Gala1,3Gal-BSA	Ga3GBSA	Gal-α-(1→3)-Gal-Sp3-BSA
22	Galb1,4GalBSA	Gb4GBSA	Gal-β-(1→4)-Gal-Sp3-BSA
23	Gala1,2GalBSA	Ga2GBSA	Gal-α-(1→2)-Gal-Sp3-BSA
24	4AP-BSA	4APBSA	4-aminopyridine-BSA
25	Lacto-*N*-fucopentaose I-BSA	LNFPIBSA	Fuc-α-(1→2)-Gal-β-(1→3)-GlcNAc-β-(1→3)-Gal-β-(1→4)-Glc-BSA
26	Lacto-*N*-fucopentaose II-BSA	LNFPIIBSA	Fuc-α-(1→3)Gal-β-(1→3)-GlcNAcβ-(1→3)-Gal-β-(1→4)-Glc-BSA
27	Lacto-*N*-fucopentaose III-BSA	LNFPIIIBSA	Gal-β-(1→4)-[Fuc-α-(1→3)]-GlcNAc-β-(1→3)-Gal-β-(1→4)-Glc-BSA
28	Lacto-*N*-difucohexaose I-BSA	LNDHIBSA	Fuc-α-(1→2)-Gal-β-(1→3)-[Fuc-α-(1→4)]-GlcNAc-β-(1→3)-Gal-β-(1→4)-Glc-Sp3-BSA
29	LNDI-BSA/ Lewis b-BSA	LebBSA	Fuc-α-(1→2)-Gal-β-(1→3)-[Fuc-α-

Row	Probe	Abbreviation	Structure
			(1→4)]-GlcNAc-β-(1→3)-Gal-β-(1→4)-Glc-APD-BSA
30	Lewis x-BSA	LexBSA	Gal-β-(1→4)-[Fuc-α-(1→3)]-GlcNAc-Sp3-BSA
31	Di-Lex-APE-BSA, 0.5 mg	DiLexBSA	Gal-β-(1→4)-[Fuc-α(1→3)]-GlcNAc-β-(1→3)-Gal-β-(1→4)[Fuc-α-(1→3)]-GlcNAc-β1-O-APE-BSA
32	Di-Lewisx-APE-HSA	DiLexHSA	Gal-β-(1→4)-[Fuc-α-(1→3)]-GlcNAc-β-(1→3)-Gal-β(1→4)[Fuc-α-(1→3)]GlcNAc-β1-O-APE-HSA
33	Tri-Lex-APE-HSA	3LexHSA	Gal-β(1→4)-[Fuc-α-(1→3)]GlcNAc-β(1→3)Gal-β(1→4)-[Fuc-α-(1→3)]-GlcNAc-β(1→3)Gal-β(1→4)-[Fuc-α-(1→3)]-GlcNAc-β1-O-APE-HSA
34	3'Sialyl Lewis x-BSA	3SLexBSA3	Neu5Ac-α(2-3)-Gal-β(1→4)-[Fuc-α(1→3)]-GlcNAc-Sp3-BSA
35	3'Sialyl Lewis x-BSA	SLexBSA14	Neu5Ac-α-(2→3)-Gal-β-(1→4)-[Fuc-α(1→3)]-GlcNAc-Sp3-BSA
36	6-Sulfo Lewis x-BSA	6SuLexBSA	(SO4)-6Gal-β-(1→4)-[Fuc-α-(1→3)]-GlcNAc-Sp3-BSA
37	6-Sulfo Lewis a-BSA	6SuLeaBSA	(SO4)-6Gal-β-(1→3)-[Fuc-α-(1→4)]-GlcNAc-Sp3-BSA
38	3-Sulfo Lewis a-BSA	3SuLeaBSA	(SO4)-3Gal-β-(1→3)-[Fuc-α-(1→4)]-GlcNAc-Sp3-BSA
39	3-Sulfo Lewis x-BSA	3SuLexBSA	(SO4)-3Gal-β-(1→4)-[Fuc-α-(1→3)]-GlcNAc-Sp3-BSA
40	Difucosyl-para-lacto-N-hexaose-APD-HSA (Lea/Lex)	DFPLNHHSA	Gal-β-(1→3)-[Fuc-α(1→4)]-GlcNAc-β-(1→3)-Gal-β-(1→4)-[Fuc-α(1→3)]-GlcNAc-β-(1→3)-Gal-β(1→4)-Glc-APD-HSA
41	Lewis a-BSA	LeaBSA	Gal-β-(1→3)-[Fuc-α-(1→4)]-GlcNAc-Sp3-BSA
42	Lewis y-tetrasaccharide-APE-HSA	LeyHSA	Fuc-α-(1→2)-Gal-β-(1→4)-[Fuc-α-(1→3)]-GlcNAc-β1-O-APE-HSA
43	Tri-fucosyl-Ley-heptasaccharide-APE-HSA	3FLeyHSA	Fuc-α-(1→2)-Gal-β-(1→4)-[Fuc-α-(1→3)]-GlcNAc-β-(1→3)-Gal-β-(1→4)-[Fuc-α-(1→3)]-GlcNAc-β1-O-APE-HSA
44	Lacto-N-neotetraose-APD-HSA	LNnTHSA	Gal-β-(1→4)-GlcNAc-β-(1→3)-Gal-β-(1→4)-Glc-APD-HSA
45	Lacto-N-tetraose-APD-HSA	LNTHSA	Gal-β-(1→3)-GlcNAc-β-(1→3)-Gal-β-(1→4)-Glc-APD-HSA
46	Sialyl-LNF V-APD-HSA	SLNFVHSA	Fuc-α-(1→2)-Gal-β-(1→3)-[NeuAc-α-(2→6)]-GlcNAc-β-(1→3)-Gal-β-(1→4)-Glc-APD-HSA

Row	Probe	Abbreviation	Structure
47	Monofucosyl, monosialyllacto-N-neohexaose-APD-HSA	MMLNnHHSA	Neu5Ac-α-(2→3)-Gal-β-(1→4)-GlcNAc-β-(1→3)-[Gal-β-(1→4)-[Fuc-α-(1→3)]-GlcNAc-β-(1→6)]-Gal-β-(1→4)Glc-APD-HSA
48	Sialyl-LNnT-penta-APD-HSA	SLNnTHSA	Neu5Ac-α-(2→3)-Gal-β-(1→4)-GlcNAc-β-(1→3)-Gal-β-1→4)-Glc-APD-HSA
49	GM1-pentasaccharide-APD-HSA	GM1HSA	Gal-β-(1→3)-GalNAc-β-(1→4)-[Neu5Ac-α-(2→3)-]Gal-β-(1→4)-Glc-APD-HSA
50	Asialo-GM1-tetrasaccharide-APD-HSA	aGM1HSA	Gal-β-(1→3)-GalNAc-β-(1→4)-Gal-β-(1→4)-Glc-APD-HSA
51	Globo-N-tetraose-APD-HSA	GlobNTHSA	GalNAc-β-(1→3)-Gal-α-(1→4)-Gal-β-(1→4)-Glc-APD-HSA
52	Globotriose-APD-HSA	GlobTHSA	Gal-α-(1→4)-Gal-β-(1→4)-Glc-β1-APE-HSA

Table S6. List of features printed on carbohydrate microarray B.

Row	Probe	Abbreviation	Structure
1	Fetuin	Fetuin	Fetuin
2	Invertase	Inv	Invertase from yeast
3	PBS	PBS	PBS
4	Ovalbumin	Ovalbumin	Ovalbumin from hen egg
5	Fibrinogen	Fibrin	Fibrinogen
6	alpha-1-antitrypsin	A1AT	alpha-1-antitrypsin
7	4AP-HSA-	4APHSA	4AP-HSA-
8	α-Crystallin from bovine lens	a-C	a-Crystallin
9	Mana1,3(Mana1,6)Man-BSA	M3BSA	Man-α-(1→3)-[Man-α-(1→6)-]Man-BSA
10	GlcNAc-BSA	GlcNAcBSA	GlcNAc-BSA
11	Ceruloplasmin	Cerulo	Ceruloplasmin
12	alpha-1-acid glycoprotein	AGP	α-1-acid glycoprotein
13	3´-Sialyllactose-APD-HSA	3SLacHSA	Neu5Ac-α-(2→3)-Gal-β-(1→4)-Glc-APD-HSA
14	6´-Sialyllactose-APD-HSA	6SLacHSA	Neu5Ac-α-(2→6)-Gal-β-(1→4)-Glc-APD-HSA
15	LacNAc-a-4AP-BSA	LacNAcaBSA	Gal-β-(1→4)-GlcNAc-α-4AP-BSA
16	LacNAc-b-4AP-BSA	LacNAcb4APBSA	Gal-β-(1→4)-GlcNAc-β-4AP-BSA
17	H Type II-APE-BSA	H2BSA	Fuc-α-(1→2)-Gal-β(1→4)-GlcNAc-β1-APE-BSA
18	H-Type 2-APE-HSA	H2HSA	Fuc-α-(1→2)-Gal-β(1→4)-GlcNAc-β1-APE-HSA
19	Ovomucoid	Ovomuc	Ovomucoid
20	Gala1,3Galb1,4GlcNAc-HSA	GGGNHSA	Gal-α-(1→3)-Gal-β-(1→4)-GlcNAc-HSA
21	Gala1,3Gal-BSA	Ga3GBSA	Gal-α-(1→3)-Gal-Sp3-BSA
22	L-Rhamnose-Sp14-BSA	RhaBSA	L-Rha-Sp14
23	Gal-a-ITC-BSA	XGalaBSA	Gal-α-ITC
24	4AP-BSA	4APBSA	4AP-BSA
25	Lacto-*N*-fucopentaose I-BSA	LNFPIBSA	Fuc-α-(1→2)-Gal-β-(1→3)-GlcNAc-β-(1→3)-Gal-β-(1→4)-Glc-BSA
26	Man-a-ITC-BSA	XManaBSA	Man-α-ITC-BSA
27	Lac-b-4AP-BSA	XLacbBSA	Lac-β-4AP-BSA
28	Man-b-4AP-BSA	XManbBSA	Man-β-4AP-BSA
29	LNDI-BSA/ Lewis b-BSA	LebBSA	Fuc-α-(1→2)-Gal-β-(1→3)-[Fuc-α(1→4)]-GlcNAc-β-(1→3)-Gal-β-(1→4)-Glc-APD-BSA

Row	Probe	Abbreviation	Structure
30	Lewis x-BSA	LexBSA	Gal-β-(1→4)-[Fuc-α(1→3)]-GlcNAc-Sp3-BSA
31	Gal-b-ITC-BSA	XGalbBSA	Gal-β-ITC-BSA
32	Xyl-b-4AP-BSA	XylbBSA	Xyl-β-4AP-BSA
33	Xyl-a-4AP-BSA	XylaBSA	Xyl-α-4AP-BSA
34	3'Sialyl Lewis x-BSA	3SLexBSA3	Neu5Ac-α-(2→3)-Gal-β-(1→4)-[Fuc-α-(1→3)]-GlcNAc-Sp3-BSA
35	Glc-b-4AP-BSA	XGlcbBSA	Glc-β-4AP-BSA
36	Fuc-a-4AP-BSA	FucaBSA	Fuc-α-4AP-BSA
37	6-Sulfo Lewis a-BSA	6SuLeaBSA	(SO4)-6-Gal-β-(1→3)-[Fuc-α(1→4)]-GlcNAc-Sp3-BSA
38	Fuc-b-4AP-BSA	FucbBSA	Fuc-β-4AP-BSA
39	Glc-b-ITC-BSA	GlcbITCBSA	Glc-β-ITC-BSA
40	Gal-b-4AP-BSA	Galb4APBSA	Gal-β-4AP-BSA
41	Neu5Gc-4AP-BSA	Neu5GcBSA	Neu5Gc-4AP-BSA
42	Lewis y-tetrasaccharide-APE-HSA	LeyHSA	Fuc-α-(1→2)-Gal-β-(1→4)-[Fuc-α(1→3)]-GlcNAc-β1-O-APE-HSA
43	PBS	PBS	PBS
44	PBS	PBS	PBS
45	Lacto-N-tetraose-APD-HSA	LNTHSA	Gal-β-(1→3)-GlcNAc-β-(1→3)-Gal-β-(1→4)-Glc-APD-HSA
46	PBS	PBS	PBS
30	PBS	PBS	PBS
31	PBS	PBS	PBS
32	Collagen type IV	CollagenIV	Collagen type IV
33	Globotriose-HSA	D-GlobTHSA	Gal-α-(1→4)-Gal-β-(1→4)-Glc-Sp3-BSA
34	Globo-N-tetraose-APD-HSA	GlobNTHSA	GalNAc-β-(1→3)-Gal-α-(1→4)-Gal-β-(1→4)-Glc-APD-HSA
35	Globotriose-APD-HSA	GlobTHSA	Gal-α(1→4)-Gal-β-(1→4)-Glc-β1-APE-HSA

Table S7. Characterised adhesins and conserved domains from human GIT commensal *Lactobacillus* species and strains bacteria. IECs: Intestinal epithelial cells.

Name	Amino acids	Binding ligand	NCBI entry	Conserved Domain	Species and strain	Reference
Mub	4326	Human IECs, mucus	AAV43217	YSIRK_signal super family[cl04650]; MucBP super family[cl05785] x2	*Lactobacillus acidophilus* NCFM	2
Mub	3269	Mucus	AAF25576	MucBP super family[cl05785] x14	*Lactobacillus reuteri* 1063	3
SlpA	1017	Human IECs	AAV43202	none	*Lactobacillus acidophilus* NCFM	2
SlpA	465	Human IECs, collagen, laminin and fibronectin	CAA78618	none	*Lactobacillus brevis* ATCC 8287	4
Slp	437	Human IECs	AAZ99044	PBP2_NikA_DppA_OppA_like super family[cl01709]; SLAP[pfam03217]	*Lactobacillus helveticus* R0052	5
FbpA	563	Human IECs, fibronectin	AAV42987	DUF814 super family[cl05307]	*Lactobacillus acidophilus* NCFM	2
CbsA	440	Collagen I and IV, laminin, bacterial lipoteichoic acids (LTA)	AAB58734	SLAP[pfam03217]	*Lactobacillus crispatus*	6
EF-Tu	396	Human IECs, mucus	AAS08831	EF_Tu[cd01884]; EFTU_III[cd03707]; EFTU_II[cd03697]	*Lactobacillus johnsonii* NCC533	7
GroEL	543	Human IECs, mucus	AAS08453	GroEL[cd03344]	*Lactobacillus johnsonii* NCC533	7
Msa	1010	Mucus *via* mannose binding	CCC78612	lectin_L-type[cd01951]; MucBP super family[cl05785] x3	*Lactobacillus plantarum* WCSF	8
CnBP	263	Human IECs, mucus	CAA68052	PBPb[cd00134]	*Lactobacillus reuteri*	3

Name	Amino acids	Binding ligand	NCBI entry	Conserved Domain	Species and strain	Reference
MapA	263	Human Caco-2 cells, mucus	CAC05301	PBPb[cd00134]	*Lactobacillus reuteri* 104R	9
LspA	1209	Human IECs	YP_535207	none	*Lactobacillus salivarius* UCC118	10
SpaC	895	Human mucus	CAR86339	vWFA[cd00198]; Peptidase_M14NE-CP-C_like super family[cl15700] x3	*Lactobacillus rhamnosus GG*	11
SpaF	983	Human mucus	CAR88267	Peptidase_M14NE-CP-C_like super family[cl15700] x2	*Lactobacillus rhamnosus* GG	11
Lar0958	1229	Human mucus	BAG25474.1	YSIRK_signal [TIGR01168]	*Lactobacillus reuteri* JCM 1112	12

Table S8. Expression of candidate adhesins from *L. paracasei*. Fold change in expression of candidate genes normalised with 16S rRNA reference gene and with lag phase used as the calibrator, between regular MRS media, MRS+PGM, MRS+NaCl or MRS+PGM/NaCl. Significant data (P<0.05) are coloured and underlined in red.

	Fold change							
NCBI entry	**MRS**	**sd**	**MRS +PGM**	**sd**	**MRS +NaCl**	**sd**	**MRS +PGM/NaCl**	**sd**
EEI67294.1	1.03	0.31	1.17	0.03	1.22	0.05	1.59	0.33
EEI68802.1	1.71	0.79	1.25	0.29	0.53	0.38	1.85	1.46
EEI67357.1	1.25	0.84	0.91	0.31	1.49	1.12	1.16	0.23
EEI68137.1	1.63	0.48	1.63	0.3	1.28	0.7	1.19	0.23
EEI69168.1	1.43	0.26	1.42	0.27	1.10	0.2	2.58	0.89
EEI68918.1	2.08	2.23	1.22	0.34	2.35	1.94	1.58	1.53
EEI68919.1	1.50	0.55	1.04	0.25	1.26	0.42	1.42	0.26
EEI68842.1	1.26	0.5	1.20	0.37	0.93	0.18	1.75	0.65
EEI68395.1	1.40	0.73	0.69	0.41	1.04	1.29	1.50	1.29
EEI68452.1	1.16	0.21	1.61	0.64	1.19	0.52	2.51	1.9
EEI68453.1	1.06	0.13	1.11	0.16	1.14	0.39	1.71	0.71
EEI68682.1	1.02	0.34	0.86	0.07	1.11	0.49	2.54	1.02
EEI68684.1	1.72	1.4	0.94	0.2	1.09	0.88	2.61	1.71
Mean	1.40	0.67	1.16	0.28	1.21	0.66	1.85	0.94

References

1. A. Barinov, *et al*., Prediction of surface exposed proteins in *Streptococcus pyogenes*, with a potential application to other Gram-positive bacteria. *Proteomics*, 2009, **9**, 61-73.
2. E., Altermann, Complete genome sequence of the probiotic lactic acid bacterium *Lactobacillus acidophilus* NCFM. *Proc. Natl Acad. Sci., U.S.A.*, 2005, **102**, 3906-3912.
3. S. Roos and H. Jonsson, A high-molecular-mass cell-surface protein from *Lactobacillus reuteri* 1063 adheres to mucus components. *Microbiol.*, 2002, **148**, 433-442.
4. M. Jakava-Viljanen, *et al*., Isolation of three new surface layer protein genes (slp) from *Lactobacillus brevis* ATCC 14869 and characterization of the change in their expression under aerated and anaerobic conditions. *J. Bacteriol*., 2002, **184**, 6786-6795.
5. K.C. Johnson-Henry, *et al*., Surface-layer protein extracts from *Lactobacillus helveticus* inhibit enterohaemorrhagic *Escherichia coli* O157:H7 adhesion to epithelial cells. *Cell. Microbiol.*, 2007, **9**, 356-367.
6. J. Sillanpaa, *et al*., Characterization of the collagen-binding S-layer protein CbsA of *Lactobacillus crispatus*. J. Bacteriol., 2000, **182**, 6440-6450.
7. D. Granato, *et al*., Cell surface-associated elongation factor Tu mediates the attachment of *Lactobacillus johnsonii* NCC533 (La1) to human intestinal cells and mucins." *Infect. Immun.*, 2004, **72**, 2160-2169.
8. G. Gross, *et al*., Mannose-specific interaction of *Lactobacillus plantarum* with porcine jejunal epithelium. *FEMS Immunol. Med. Microbiol*., 2008, **54**, 215-223.
9. Y. Miyoshi, *et al*., A mucus adhesion promoting protein, MapA, mediates the adhesion of *Lactobacillus reuteri* to Caco-2 human intestinal epithelial cells. *Biosci. Biotechnol. Biochem.*, 2006, **70**, 1622-1628.
10. M. J. Claesson, Multireplicon genome architecture of *Lactobacillus salivarius*. *Proc. Natl Acad. Sci., U.S.A.*, 2006, **103**, 6718-6723.
11. M. Kankainen, *et al.*, Comparative genomic analysis of *Lactobacillus rhamnosus* GG reveals pili containing a human- mucus binding protein. *Proc. Natl Acad. Sci., U.S.A.*, 2009, **106**, 17193-17198.
12. S. Etzold, *et al*., Structural and molecular insights into novel surface-exposed mucus adhesins from *Lactobacillus reuteri* human strains. *Mol. Microbiol*., 2014, **92**, 543-556.