AEM Accepted Manuscript Posted Online 14 February 2020
Appl. Environ. Microbiol. doi:10.1128/AEM.02808-19

Identification of genes required for glucan exopolysaccharide production in *Lactobacillus johnsonii* suggests a novel mechanism of biosynthesis

Melinda J. Mayer[a]#, Alfonsina D'Amato[b†], Ian J. Colquhoun[b], Gwénaëlle Le Gall[c], Arjan Narbad[a]

[a]Gut Microbes and Health, Quadram Institute Bioscience, Norwich, UK
[b]Analytical Sciences Unit, Quadram Institute Bioscience, Norwich, UK
[c]Department of Medicine, Faculty of Medicine and Health Sciences, Bob Champion Research and Educational Building, University of East Anglia, Norwich, UK

 # corresponding author, Melinda J. Mayer, email: melinda.mayer@quadram.ac.uk

[†]**current address**, Pharmaceutical Department, University of Milan, Via Mangiagalli 25, Milan, Italy

Running title, EPS synthesis in *L. johnsonii*

Abstract

Lactobacillus johnsonii FI9785 makes two capsular exopolysaccharides –a heteropolysaccharide (EPS2) encoded by the *eps* operon, and a branched glucan homopolysaccharide (EPS1). The homopolysaccharide is synthesised in the absence of sucrose and there are no typical glucansucrase genes in the genome. Quantitative proteomics was used to compare the wild type to a mutant where EPS production was reduced, to attempt to identify proteins associated with EPS1 biosynthesis. A putative bactoprenol glycosyltransferase, 242, was less abundant in the Δ*eps_cluster* mutant than in the wild type. NMR analysis of isolated EPS showed that deletion of the *242* gene prevented the accumulation of EPS1, without affecting EPS2 synthesis, while plasmid complementation restored EPS1 production. The deletion of *242* also produced a slow growth phenotype, which could be rescued by complementation. 242 shows amino acid homology to bactoprenol glycosyltransferase GtrB, involved in O-antigen glycosylation, while *in silico* analysis of neighbouring gene *241* suggested it encodes a putative flippase with homology to the GtrA superfamily. Deletion of *241* also prevented production of EPS1, and again caused a slow growth phenotype, while plasmid complementation reinstated EPS1 synthesis. Both genes are highly conserved in *L. johnsonii* strains isolated from different environments. These results suggest there may be a novel mechanism for homopolysaccharide synthesis in the Gram-positive *L. johnsonii*.

Importance

Exopolysaccharides are key components of the surfaces of their bacterial producers, contributing to protection, microbial and host interactions and even virulence. They also have significant applications in industry, and understanding biosynthetic mechanisms may allow improved production of novel and valuable polymers. Four categories of bacterial

exopolysaccharide biosynthesis have been described in detail, but novel enzymes and glycosylation mechanisms are still being described. Our findings that a putative bactoprenol glycosyltransferase and flippase are essential to homopolysaccharide biosynthesis in *Lactobacillus johnsonii* FI9785 indicate that there may be an alternative mechanism of glucan biosynthesis to the glucansynthase pathway. Disturbance of this synthesis leads to a slow growth phenotype. Further elucidation of this biosynthesis may give insight into exopolysaccharide production and its impact on the bacterial cell.

Keywords exopolysaccharide, alpha glucan, *Lactobacillus johnsonii*, proteomics, glycosyltransferase, Nuclear Magnetic Resonance

Introduction

Production of exopolysaccharides (EPS) has a large impact on the nature of the bacterial surface and hence on interactions with the environment, hosts and host defence systems, and other microbes (1, 2). EPS can protect bacteria against environmental conditions, both outside and inside the host (1, 3, 4), and in the case of pathogens such as *Streptococcus pneumoniae* they can have an important association with immune evasion and virulence (5). EPS can have immunomodulatory and protective properties in the host (6-9) and can affect the composition and function of the gut microbiota (10, 11). EPS can also play a crucial role in biofilm formation, adhesion to host cells and colonisation (3, 12-15). In addition to their biological importance, bacterial EPS have a range of technological applications in food, pharmaceutical and other industries and may also have potential health benefits, due to their activities in immune stimulation, anti-tumour activity and lowering of blood cholesterol, or as prebiotics (1, 2, 16, 17).

68 *Lactobacillus johnsonii* FI9785 is a poultry isolate which has shown promise as a competitive
69 exclusion agent against *Clostridium perfringens* (18) and *Campylobacter jejuni* (19). This
70 strain makes 2 capsular exopolysaccharides – EPS2, a heteropolysaccharide containing
71 glucose and galactose encoded by a 14 gene *eps* operon of the Wzx/Wzy type, and EPS1, a
72 branched dextran homopolysaccharide with an α-(1→6) backbone and α-(1→2) branches
73 which are present on every unit of the backbone and consist of a single glucose (Glc) residue
74 (20, 21). This is an unusual structure which has not been described in other bacteria, although
75 a small percentage of α-(1→2) branches were seen in dextran produced by *Leuconostoc*
76 *citreum* E497 (22). Glucansucrases have been shown to synthesise homopolysaccharides in
77 lactic acid bacteria, using sucrose as a substrate (17). However, *L. johnsonii* FI9785 makes
78 EPS1 in the absence of sucrose and there is no glucansucrase gene present in the genome,
79 suggesting a different mode of biosynthesis (20). In previous work, the 14 gene *eps* operon
80 (loci *FI9785_1170* to *FI9785_1183* inclusive, now renamed FI9785_RS05260 to
81 FI9785_RS05325) was removed by deletion mutagenesis to create the mutant Δ*eps_cluster*
82 (20), and a second mutant strain where just the transcriptional regulator *epsA* (*FI9785_1183*)
83 was deleted was also constructed (23). Although these mutations were expected to just affect
84 the synthesis of EPS2 and not EPS1, these strains did not show an EPS layer by transmission
85 electron microscopy (TEM), and NMR analysis of EPS extractions failed to identify either
86 EPS1 or EPS2 (20, 23, 24). In this work we compared the proteome of the wild type *L.*
87 *johnsonii* FI9785 EPS producer with the Δ*eps_cluster* mutant to attempt to identify proteins
88 involved in homopolysaccharide biosynthesis.

89

90 **Results**

91 *Comparative quantitative proteomic analyses identified proteins affected by deletion of the*
92 eps *cluster*

In order to identify proteins involved in EPS biosynthesis, the proteome of the wild type was compared to that of a mutant with a reduced EPS capsule, to highlight proteins which were missing or down-regulated in the mutant. Proteomic analysis of the soluble fractions of *L. johnsonii* FI9785 and Δ*eps_cluster* identified several proteins which were differently expressed between the two strains. The protein samples were trypsin digested and labelled by iTRAQ (isobaric tag for relative and absolute quantitation) reagents, mixed and analysed by nLC MS/MS, or directly analysed without labelling for the label free experiment. Andromeda analyses resulted in the identification of 699 soluble proteins (Supplementary Dataset S1), 49 of which were differentially expressed in the Δ*eps_cluster* strain versus the wild type (WT, Table 1). Volcano plots in Fig. 1 show the proteins which changed in abundance, obtained respectively in iTRAQ (A) and label free experiments (B). The two different quantitative approaches allowed the quantitation of identical proteins with a similar ratio in the mutant versus control, eg D0R1R2, supporting the accuracy of the analyses, but also identified different proteins, allowing an in-depth characterization of proteins altered in the Δ*eps_cluster* strain. 20 proteins were found at a higher level in Δ*eps_cluster*, 4 identified by iTRAQ and 17 by the label free approach, with only one found by both methods; the remaining 29 proteins were at higher levels in the WT, 17 found by iTRAQ and 14 by the label free method, with 2 proteins identified by both methods (Table 1). In Fig. 2, enriched Gene Ontology (GO) terms of proteins found at different levels in *L. johnsonii* FI9785 and Δ*eps_cluster* strains are described. Soluble proteins, mainly present in the cytoplasm, are involved in ATP binding, GO:0005524, translation, GO:0006412, nucleotide binding GO:0000166 and transferase activity, GO:0016740 in the mutant strain. Almost half of the proteins with altered abundance were associated with ribosomal structure, translation and protein biosynthesis, but some were more and some less abundant in the Δ*eps_cluster* mutant, with no discernible pattern. No other biological processes seemed to be strongly

118 impacted in the Δ*eps_cluster* mutant. Although EPS is known to protect the cells from stress
119 there were no notable changes in stress response, except a higher level of thiol peroxidase,
120 involved in cell redox homeostasis (Table 1).

121 One protein found at a lower level in the Δ*eps_cluster* mutant - D0R1R2, encoded by
122 *FI9785_242* (now renamed FI9785_RS00855) - was identified by RAST analysis as a
123 bactoprenol glycosyltransferase, involved in cell wall biosynthesis. This was one of the three
124 proteins identified by both iTRAQ and the label free protocol. Blastp analysis indicated
125 homology to the glycosyltransferase 2 superfamily, particularly to domains cd04187 (DPM1-
126 like bac, 7.24 e-81), PRK10714 superfamily (undecaprenylphosphate 4-deoxy-4-
127 formamidoL-arabinose transferase, 1.28e-33), pfam00535 (glycosyl transferase family 2,
128 6.63e-28) and COG0463 (glycosyltransferase involved in cell wall biosynthesis, 2.2e-26).
129 This protein was selected for gene deletion to investigate a possible role in EPS1
130 biosynthesis.

131

132 *Deletion of* 242 *prevents biosynthesis of homopolysaccharide EPS1*

133 The coding sequence for 242 was deleted from the *L. johnsonii* FI9785 genome to create
134 strain Δ*242*. Comparison of ^{1}H-NMR profiles of EPS extracted from WT and Δ*242* showed
135 that EPS1 production was undetectable in samples extracted both from cell pellets and from
136 supernatants (Fig. 3, Supplemental Fig. S1A), indicating that 242 is essential for EPS1
137 production. NMR analysis of EPS extracted from a derivative of Δ*242* containing a plasmid
138 expressing the *242* gene under the regulation of a strong constitutive promoter (Δ*242*-p*242*)
139 showed that complementation restored EPS1 expression, with an increased ratio of EPS1 to
140 EPS2 compared to the WT (Fig. 3, Supplemental Fig. S1A).

141 Previous NMR analysis of EPS extracted from Δ*eps_cluster* and Δ*epsA* and then purified by
142 TCA precipitation failed to detect EPS1 or EPS2 (20, 23). However, our analysis here of

crude EPS preparations prior to TCA purification, using an increased temperature and higher number of scans, revealed the presence of EPS1 in both strains (Supplemental Fig. S1B). This indicates that the genes in the *eps* cluster encoding EPS2 are not required for EPS1 production.

241-242 *show homology to* GtrA-GtrB *and have homologues in Gram-positive bacteria*

Blastp analysis showed that amino acid homologues of 242 are widely distributed among *Lactobacillus* spp, with a high conservation of amino acid sequence (71-100% in the first seventy matches). Alignment of 242 with GtrB proteins from *Shigella* phage SfII and *Escherichia coli*, a putative bactoprenol glycosyltransferase CsbB from *Bacillus subtilis*, and a polyisoprenyl-phosphate glycosyltransferase from *Synechocystis* sp. whose crystal structure has been solved (25), show areas of homology across the whole sequence, including the motifs DXD and DXSXD which have previously been identified as being conserved in glycosyltransferases (25-27) (Fig. 4A). Mutation of selected amino acids in the *Synechosystis* sp. GtrB was previously shown to affect enzymatic activity (25) – all but one of these amino acids are conserved in 242 (Fig. 4A).

Blastp analysis of the translated product of the gene upstream of *242*, *FI9785_241* (now renamed FI9785_RS00850), shows homology to domains pfam04138 (GtrA-like protein, 3.04e-18) and COG2246 (putative flippase GtrA, 2.78e-07). When aligned to the GtrA sequence pairing the SfII and *E. coli* GtrBs, 241 shows some conservation of sequence but less than that seen with the GtrB counterparts (Fig. 4B). GtrAB pairs have been identified in a range of Gram-negative bacteria and their bacteriophages, and are commonly found with a glycosyltransferase GtrX, with the three protein complex engineering the glycosylation of O-antigens with a single sugar moiety (28). However, we could not identify any further

167 glycosyltransferases in the *L. johnsonii* FI9785 genome in the immediate vicinity of *241* and
168 *242*.

169 The *241-242* pair and surrounding genes show strong nucleotide conservation in other strains
170 of *L. johnsonii* isolated from different sources. A surrounding 11.1 kb section encompassing
171 15 open reading frames (ORFs) from *L. johnsonii* FI9785 was compared with equivalent
172 regions from annotated genomes of strains isolated from the human gut (NCC533), pig
173 intestine (DPC6026), rat faeces (N6.2), turkey (UMNLJ22) and mouse faeces (Byun-jo-01),
174 selecting the area between homologues of 2,3-diphosphoglycerate-dependent
175 phosphoglycerate mutase and an aldose 1-epimerase family protein (Fig. 5). The conservation
176 of ORFs surrounding the *gtrAB* pair varies among strains, with some ORFs being present but
177 interrupted by stop codons. The section encoding the 30S ribosomal protein, *241* and *242* is
178 present in all genomes. Translated sequences of ORFs which are present in more than one
179 genome show high amino acid similarity between strains; the 242 sequence (WP_012845545)
180 shows 99-100% identity with the equivalent sequences in the other genomes
181 (WP_012845545, WP_011161379 and WP_014567007). Alignment of the surrounding
182 nucleotide region showed high conservation of the region covering the *241-242* pair, and
183 analysis of these two genes in the 6 genomes showed between 97.1 and 99.8% nucleic acid
184 identity with the FI9785 sequence (Fig. 5B). The central region of strong nucleotide
185 conservation stretches from upstream of the 30S ribosomal gene to the non-coding sequence
186 after *242*.

187

188 241 *is required for EPS1 biosynthesis*

189 To confirm the involvement of the putative flippase 241 in EPS1 production, a deletion
190 mutant (Δ*241*) and its derivatives containing a *241* expression plasmid (Δ*241*-p*241*) or an
191 empty plasmid control (Δ*241*-pQI0001) were constructed and their EPS analysed by NMR.

As with Δ*242*, gene deletion prevented EPS1 production while complementation restored biosynthesis (Fig. 6, Supplemental Fig. S1C). A mutant where both *241* and *242* were deleted also showed production of EPS2 only (data not shown).

EPS1 production affects growth

Both the Δ*242* and Δ*241* strains showed a slower growth phenotype than the wild type, both in liquid and on solid medium (Fig. 7). This phenotype was similar when the strain contained an empty vector control, but normal growth was restored in liquid by overexpression of the *242* or *241* genes, although plate growth remained slightly retarded in the *242* complemented mutant. Mutant colonies did reach the size of typical 1 d WT colonies after further incubation, within 2 d. The slow growth phenotype was maintained during growth in anaerobic conditions and at a lower temperature (30°C). The presence or absence of EPS1 did not seem to affect aggregation, while as noted previously non-production of EPS2 in Δ*epsE* caused a strong aggregation phenotype (21), suggesting that EPS2 is a primary contributor to low aggregation of the WT (Fig. 7C). Deletion of *242* also did not have a strong effect on colony phenotype, with colonies retaining a rough and crinkled appearance, although overexpression of *242* resulted in a smoother colony upon longer incubation. TEM showed that the Δ*242* and Δ*241* mutants retained a visible EPS layer; this was more frequently irregular than in WT samples (Fig. 7E). Cells overexpressing *242* or *241* also exhibited a thick EPS layer, and in the case of Δ*242*p*242* this layer was consistently paler, suggesting a different response to the osmium staining.

Discussion

Effect of EPS2 *loss on the* L. johnsonii *FI9785 proteomic profile*

Apart from variations in proteins associated with ribosome structure, translation and protein synthesis, very few biological processes seemed strongly affected in the soluble protein content by the loss of EPS2 synthesis in the Δ*eps_cluster* mutant. Comparative analysis of proteins from *Lactobacillus plantarum* grown at two temperature conditions, which gave a 10-fold difference in EPS production, also found few changing proteins (29). It is interesting that loss of EPS2 production correlated with lower abundance of 242 in the Δ*eps_cluster* mutant compared to the WT. We have now determined that this mutant is able to produce EPS1, but its biosynthesis is affected, either by the absence of the *eps* cluster genes or EPS2 itself, or in response to changed cell conditions responding to reduction of a protective layer. The regulation of EPS synthesis has been linked to external signal and quorum sensing in a range of bacteria, including *L. plantarum* (30). Blast analysis of a putative transcriptional regulator, D0R501, which was also less abundant in the Δ*eps_cluster* mutant, showed a relationship to the YebC/PmpR family; regulators of this family are involved in a range of other processes, including quorum sensing (31). Further investigation of the regulation of EPS1 and EPS2 genes, proteins and polymers and how they relate to each other will be an interesting area for future study.

Involvement of putative flippase and bactoprenol glycosyltransferase in homopolysaccharide biosynthesis in L. johnsonii

The evidence from EPS NMR profiles from deletion and complementation strains indicate that putative bactoprenol glycosyltransferase 242 and neighbouring putative flippase 241 are key components in the production of the branched glucan EPS1. In lactic acid bacteria, α-glucans such as dextran are commonly synthesised by glucansucrases, which cleave sucrose then add glucose to a growing chain (17). Three other mechanisms of EPS and O-antigen polysaccharide (O-PS) biosynthesis have been described in bacteria – the Wzx/Wzy-

241 dependent pathway, the ATP-binding cassette (ABC) transporter-dependent pathway and the
242 synthase-dependent pathway (32). The first two mechanisms begin with the addition of a
243 phosphorylated monosaccharide from a UDP-sugar to a lipid carrier, commonly thought to be
244 undecaprenyl phosphate (5, 33, 34) while the synthase pathway utilises cytosolic nucleotide-
245 activated sugars (35, 36). Guan and co-workers described a three gene operon - *gtrABX* –
246 involved in O-antigen glycosylation in a bacteriophage infecting *Shigella flexneri* and
247 demonstrated that bactoprenol glucose transferase GtrB transferred ^{14}C-glucose to decaprenyl
248 phosphate *in vitro* (28). They proposed a model where GtrB catalyses the transfer of glucose
249 from UDP-glucose to bactoprenol, GtrA flips the complex across the cytoplasmic membrane
250 and specific glycosyltransferase GtrX transfers the glucose to a specific residue on the O-
251 antigen repeating unit (28). More recently GtrB homologues have been shown to be involved
252 in glycosylation of lipoteichoic and wall teichoic acids and a similar 3-component mechanism
253 has been proposed (37-39).

254 Our hypothesis is that 242 acts as a GtrB homologue, adding a glucose molecule to a lipid
255 carrier, while the product of neighbouring gene *241* functions as a flippase. However, the full
256 process of chain and branch formation, and the possible involvement of glycosyltransferases
257 elsewhere in the genome, remains to be determined. *241-242* may be involved in the
258 decoration of a linear chain synthesised by other enzymes, or may be an integral part of a
259 biosynthetic cluster. The ability of bacterial glycosyltransferases to act on different substrates
260 and even in different pathways has been noted (40). The genes encoding the three-component
261 system involved in *S. aureus* lipoteichoic acid glycosylation are not all located together on
262 the chromosome (38), so it would not be unprecedented for a distant gene/s to be involved in
263 a three- or four- component EPS biosynthetic pathway. The genome of *L. johnsonii* FI9785
264 contains several other glycosyltransferase genes which may be involved in synthesis of a
265 linear chain, acting in concert with 241-242 to produce the final external EPS1. It is hoped

266 that further examination of these genes will lead to a clearer model for the synthesis of this
267 unusual EPS.

268

269 *Effect of* 242 *or* 241 *deletion on* L. johnsonii

270 Mutations affecting *L. johnsonii* FI9785 EPS synthesis have been shown to affect
271 aggregation, biofilm formation, adhesion to human HT29 cells and chicken gut explants and
272 resistance to stress, suggesting that EPS has a protective capacity (20, 21, 23, 41). We found
273 that gene deletion of *242* or *241* slowed bacterial growth. The slow growth phenotype is still
274 seen at lowered temperatures or in the absence of oxygen, suggesting that it is not caused by
275 increased sensitivity of cells to these conditions due to a reduction of the EPS layer. Further,
276 removal of EPS2 did not seem to have the same effect, as the *eps_cluster* mutant showed a
277 similar growth rate to the wild type when grown for proteomic analysis. It has been noted that
278 mutations which might prevent the release of undecaprenyl phosphate by blocking the full
279 EPS biosynthetic process affect cell viability, either by reducing the amount of undecaprenyl
280 available for other processes or by membrane destabilisation in the presence of lipid
281 intermediates (5, 42). However, it is not obvious as to why deletion of a protein proposed to
282 glycosylate the lipid carrier might have a similar effect, unless there are other components of
283 EPS1 biosynthesis that might also interact with the carrier.

284 In conclusion, we found that a putative glycosyltransferase, 242, was less abundant in the
285 Δ*eps_cluster* mutant, and that deletion of its gene prevented the accumulation of EPS1 while
286 plasmid complementation restored production. *In silico* analysis indicated that *242* and its
287 preceding gene *241* show similarity to two members of a three-component system, *gtrABX*,
288 shown to mediate O-antigen glycosylation in Gram-negative bacteria and more recently to be
289 involved in teichoic acid glycosylation in Gram-positive species. Further deletion and
290 complementation studies showed that *241* was also essential for EPS1 production. High

conservation of nucleotide sequence with other *L. johnsonii* strains and presence of analogous genes in other Lactobacilli suggest that this might be part of a novel mechanism for homopolysaccharide EPS biosynthesis in Gram-positive bacteria. EPS/O-PS biosynthetic pathways have been studied in detail, but many questions remain unanswered, and new enzymes are still being discovered (43). Given the potential technological applications of EPS, there is significant interest in engineering novel forms (32), and their important roles in protection and biofilm formation makes EPS biosynthesis a valid target for novel strategies to control pathogens. Further discovery of alternative mechanisms may give future opportunities to both understand and exploit bacterial EPS synthesis.

Experimental procedures

Bacterial strains and growth conditions

L. johnsonii strains were grown as described previously (41) in homemade De Man Rogosa Sharpe medium (MRS) using 2% glucose as a carbon source, at 37°C. *Lactococcus lactis* MG1614 (44) was grown in GM17 (Oxoid) at 30°C. Plasmids were selected and maintained using chloramphenicol (pFI2560, pQI0001) at 7.5 µg ml^{-1} or 5 µg ml^{-1} and erythromycin (pG^{+}host9) at 10 µg ml^{-1} and 5 µg ml^{-1} for *L. johnsonii* and *L. lactis* respectively. *L. johnsonii* strains and plasmids produced and/or used in this study are listed in Table 2.

Isolation of proteins

Soluble protein extracts were prepared from *L. johnsonii* FI9785 and Δ*eps_cluster* strains inoculated from overnight cultures at 2% into prewarmed media and grown to an optical density (OD_{600}) of 2.0 (6-7 h). Cells from 15 ml aliquots were harvested by centrifugation at 3000 x g for 15 min at 4°C, washed with 5 ml phosphate buffered saline (PBS) containing 1X

315 cOmplete protease inhibitor (Roche), re-centrifuged, washed with 1 ml PBS/protease
316 inhibitor and recentrifuged at 13,000 x g for 2 min at 4°C before removal of the supernatant
317 and freezing on dry ice. Three biological replicates and one technical replicate were prepared
318 for each strain. Pellets were resuspended in 500 µl extraction buffer (50 mM HEPES pH 7.7,
319 0.3% SDS, 1x protease inhibitor, 5U ml^{-1} RNase-free DNase (Promega), 10 mM $MgSO_4$, 1
320 mM $CaCl_2$) then sonicated using a Soniprep 150 (Sanyo) for 7 cycles of 15 s with 30 s
321 incubation on ice between cycles. After centrifugation at 13,000 x g for 25 min at 4°C to
322 pellet debris, the supernatant was precipitated overnight with 5 volumes of cold acetone at -
323 20°C. Proteins were collected by centrifugation at 14,000 x g for 10 min at 4°C and stored at
324 -20°C. Total soluble protein was resuspended in 250 µl 0.5 M trimethylammonium
325 bicarbonate buffer (Sigma), 0.05% SDS, 1X protease inhibitor and stored in LoBind tubes
326 (Eppendorf). Concentrations were measured using Bradford reagent (Bioline).
327
328 *Quantitative mass spectrometry*
329 Bacterial protein samples, three biological replicates of mutants and controls and one
330 technical replicate, were digested by trypsin and the tryptic peptides were labelled by the
331 iTRAQ 8plex kit (AB Sciex Pte. Ltd., USA) following the manufacturer's instructions. The
332 samples of each experiments were pooled and fractionated by a high pH reversed phase
333 peptide fractionation kit (Pierce, Thermo Fisher Scientific). Each single fraction was analysed
334 by a nLC MS/MS Orbitrap Fusion trihybrid mass spectrometer coupled with a nano flow
335 UHPLC system (Thermo Fisher Scientific). The peptides were separated, after being trapped
336 on a C18 pre-column, using a gradient of 3-40% acetonitrile in 0.1% formic acid, over 50
337 min at a flow rate of 300 nl min^{-1}, at 40°C. The peptides were fragmented in the linear ion
338 trap by a data-dependent acquisition method, selecting the 40 most intense ions. For label free
339 experiments, each tryptic peptide sample was analysed in triplicate as described above. All

analyses were performed in triplicate. The raw data were analysed by MaxQuant (version 1.6.2.3; RRID:SCR_014485), using Andromeda software and consulting the Uniprot_ *Lactobacillus johnsonii* (strain FI9785) (1726 sequences) protein database; the tolerance on parents was 10 ppm and on fragments was 0.02 ppm. The variable modifications allowed were oxidation on methionine and carboxyamidomethylation on cysteine as fixed modification. The false discovery rate was below 1%, using a decoy and reverse database, and the identified proteins contained at least 2 peptides with at least 6 amino acids sequenced. iTRAQ and label free quantitative analyses were also performed by MaxQuant software and evaluated by Perseus statistical software (RRID:SCR_015753) by a two-sided t test, setting a p value less than 0.05 and FDR less than 0.01. Gene Ontology analyses were performed by the QuickGO algorithm (EMBL-EBI; RRID:SCR_004608).

Plasmid construction and gene deletion

Genes were deleted from the *L. johnsonii* FI9785 chromosome as described previously (21) using the thermosensitive vector pG+host9 (45) containing a knockout cassette of the partial upstream and downstream genes, amplified and joined by splice overlap extension PCR using primers designed to generate restriction sites for cloning and to create spliced products (see supplemental text and Table 3). Initial cloning was performed using electrocompetent *Lactococcus lactis* MG1614 (46) with growth at 28°C. After sequence confirmation, plasmids were transformed into electrocompetent *L. johnsonii* FI9785 (47) and gene replacement was performed as described previously (45) using 30°C as the permissive temperature and 42°C as the non-permissive temperature. For recovery of Δ*242* and Δ*241* it was necessary to recover deletions at 30°C, and excised plasmids were cured by successive subculturing. For complementation, the *242* gene was cloned into the *L. johnsonii* expression plasmid pFI2560 (21) and the *241* gene was cloned into pFI2560-derivative pQI0001 (see supplemental data);

ligation products and control vector were transformed into electrocompetent *L. johnsonii* FI9785 as before.

Bioinformatic analysis

Translated gene sequence homologies and domain searches were performed using Blastp (RRID:SCR_001010) (48). The *L. johnsonii* genome FN298497 (49) was reanalysed using RAST (RRID:SCR_014606) (50). Amino acid alignments were performed using the clustalW algorithm (RRID:SCR_002909) in Vector NTI (Invitrogen; RRID:SCR_014265) and visualised using Genedoc. Nucleotide alignments were performed using Geneious (Biomatters Ltd., New Zealand; RRID:SCR_010519): short sequences were aligned with Geneious alignment and larger genome segments were aligned using Mauve (RRID:SCR_012852) (51).

Isolation and NMR spectroscopy of EPS

Crude EPS was isolated from 2 d 500 ml cultures grown in MRS at 37°C as described previously (21), except that the initial extraction of capsular EPS from the washed bacterial pellet was performed by sonication in 50 ml 1 M NaCl for 7 cycles of 45 s with 30 s incubation on ice between cycles, followed by centrifugation at 6000 x g and 4°C for 30 min to remove bacterial debris before the rounds of ethanol precipitation, the initial ethanol precipitation was for 3 d instead of overnight, and crude EPS was not further purified by TCA precipitation. EPS samples were analysed by NMR as before (20) but with heating to 338°K (Δ*242* series) and an increased number of scans (1024). Samples in the Δ*241* series were measured at 300°K.

389 *Growth, aggregation and phenotype studies*

390 Overnight (15 h) cultures of WT, Δ*epsE*, Δ*242*-p*242* and Δ*241*-p*241* and 20 h cultures of
391 Δ*242*, Δ*242*-pFI2560, Δ*241* and Δ*241*-pQI0001 were used for growth and aggregation
392 studies. For liquid growth, 20 ml broths were inoculated at 2% and the OD_{600} of 10-fold
393 diluted samples was measured at each hour during aerobic growth at 37°C. Colony size on
394 plates was monitored aerobically at 30°C and 37°C and anaerobically at 37°C. All liquid
395 growth of plasmid-containing strains was supplemented with chloramphenicol, while plate
396 growth was non-selective. For aggregation, triplicate 1 ml samples from vortexed overnight
397 cultures were transferred to cuvettes and the OD_{600} was measured hourly during incubation at
398 room temperature. Growth and aggregation assays were each performed three times and
399 representative curves are shown. TEM images were taken from overnight cultures as
400 described previously (20).

401

402 **Acknowledgements** This work was supported by funding from the Biotechnology and
403 Biological Sciences Research Council Institute Strategic Programme grants Gut Health and
404 Food Safety BB/J004529/1 and Gut Microbes and Health BB/R012490/1, project
405 BBS/E/F/000PR10356. TEM images were prepared by Kathryn Cross and Catherine Booth
406 (QIB). We are grateful to Emmanuelle Maguin for kind provision of the pG^{+}host9 vector and
407 would like to thank Prof. Nathalie Juge for critical reading of the manuscript.

408

409 **Author contributions** MJM, AdA and AN contributed to the conception and design of the
410 study; MJM, AdA, IC and GLG performed the acquisition, analysis and interpretation of the
411 data; all authors contributed to the writing of the manuscript.

412

413 **References**

414

415 1. Caggianiello G, Kleerebezem M, Spano G. 2016. Exopolysaccharides produced by
416 lactic acid bacteria: from health-promoting benefits to stress tolerance mechanisms.
417 Appl Microbiol Biotechnol 100:3877-86.
418 2. Zeidan AA, Poulsen VK, Janzen T, Buldo P, Derkx PMF, Oregaard G, Neves AR.
419 2017. Polysaccharide production by lactic acid bacteria: from genes to industrial
420 applications. FEMS Microbiol Rev 41:S168-S200.
421 3. Lebeer S, Claes IJJ, Verhoeven TLA, Vanderleyden J, De Keersmaecker SCJ. 2011.
422 Exopolysaccharides of *Lactobacillus rhamnosus* GG form a protective shield against
423 innate immune factors in the intestine. Microb Biotechnol 4:368-374.
424 4. Donot F, Fontana A, Baccou JC, Schorr-Galindo S. 2012. Microbial
425 exopolysaccharides: Main examples of synthesis, excretion, genetics and extraction.
426 Carbohyd Polym 87:951-962.
427 5. Yother J. 2011. Capsules of *Streptococcus pneumoniae* and other bacteria: paradigms
428 for polysaccharide biosynthesis and regulation. Annu Rev Microbiol 65:563-81.
429 6. Fanning S, Hall LJ, Cronin M, Zomer A, MacSharry J, Goulding D, Motherway MO,
430 Shanahan F, Nally K, Dougan G, van Sinderen D. 2012. Bifidobacterial surface-
431 exopolysaccharide facilitates commensal-host interaction through immune modulation
432 and pathogen protection. Proc Natl Acad Sci U S A 109:2108-13.
433 7. Gorska S, Sandstrom C, Wojas-Turek J, Rossowska J, Pajtasz-Piasecka E,
434 Brzozowska E, Gamian A. 2016. Structural and immunomodulatory differences
435 among lactobacilli exopolysaccharides isolated from intestines of mice with
436 experimentally induced inflammatory bowel disease. Sci Rep 6:37613.

437 8. Jones SE, Paynich ML, Kearns DB, Knight KL. 2014. Protection from intestinal
438 inflammation by bacterial exopolysaccharides. J Immunol 192:4813-20.
439 9. Zivkovic M, Hidalgo-Cantabrana C, Kojic M, Gueimonde M, Golic N, Ruas-Madiedo
440 P. 2015. Capability of exopolysaccharide-producing *Lactobacillus paraplantarum*
441 BGCG11 and its non-producing isogenic strain NB1, to counteract the effect of
442 enteropathogens upon the epithelial cell line HT29-MTX. Food Res Int 74:199-207.
443 10. Lindstrom C, Xu J, Oste R, Holst O, Molin G. 2013. Oral administration of live
444 exopolysaccharide-producing *Pediococcus parvulus*, but not purified
445 exopolysaccharide, suppressed *Enterobacteriaceae* without affecting bacterial
446 diversity in ceca of mice. Appl Environ Microbiol 79:5030-7.
447 11. Salazar N, Ruas-Madiedo P, Kolida S, Collins M, Rastall R, Gibson G, de Los Reyes-
448 Gavilan CG. 2009. Exopolysaccharides produced by *Bifidobacterium longum* IPLA
449 E44 and *Bifidobacterium animalis* subsp. *lactis* IPLA R1 modify the composition and
450 metabolic activity of human faecal microbiota in pH-controlled batch cultures. Int J
451 Food Microbiol 135:260-7.
452 12. Guttenplan SB, Blair KM, Kearns DB. 2010. The EpsE flagellar clutch is bifunctional
453 and synergizes with EPS biosynthesis to promote *Bacillus subtilis* biofilm formation.
454 PLoS Genet 6:e1001243.
455 13. Kim HS, Park SJ, Lee KH. 2009. Role of NtrC-regulated exopolysaccharides in the
456 biofilm formation and pathogenic interaction of *Vibrio vulnificus*. Mol Microbiol
457 74:436-53.
458 14. Koo H, Xiao J, Klein MI, Jeon JG. 2010. Exopolysaccharides produced by
459 *Streptococcus mutans* glucosyltransferases modulate the establishment of
460 microcolonies within multispecies biofilms. J Bacteriol 192:3024-3032.

15. Walter J, Schwab C, Loach DM, Ganzle MG, Tannock GW. 2008. Glucosyltransferase A (GtfA) and inulosucrase (Inu) of *Lactobacillus reuteri* TMW1.106 contribute to cell aggregation, *in vitro* biofilm formation, and colonization of the mouse gastrointestinal tract. Microbiology-SGM 154:72-80.

16. Freitas F, Alves VD, Reis MA. 2011. Advances in bacterial exopolysaccharides: from production to biotechnological applications. Trends Biotechnol 29:388-98.

17. Leemhuis H, Pijning T, Dobruchowska JM, van Leeuwen SS, Kralj S, Dijkstra BW, Dijkhuizen L. 2013. Glucansucrases: three-dimensional structures, reactions, mechanism, alpha-glucan analysis and their implications in biotechnology and food applications. J Biotechnol 163:250-72.

18. La Ragione RM, Narbad A, Gasson MJ, Woodward MJ. 2004. *In vivo* characterization of *Lactobacillus johnsonii* FI9785 for use as a defined competitive exclusion agent against bacterial pathogens in poultry. Lett Appl Microbiol 38:197-205.

19. Manes-Lazaro R, Van Diemen PM, Pin C, Mayer MJ, Stevens MP, Narbad A. 2017. Administration of *Lactobacillus johnsonii* FI9785 to chickens affects colonisation by *Campylobacter jejuni* and the intestinal microbiota. Br Poult Sci 58:373-381.

20. Dertli E, Colqhoun IJ, Gunning AP, Bongaerts RJ, Le Gall G, Bonev BB, Mayer MJ, Narbad A. 2013. Structure and biosynthesis of two exopolysaccharides produced by *Lactobacillus johnsonii* FI9785. J Biol Chem 288:31938-51.

21. Horn N, Wegmann U, Dertli E, Mulholland F, Collins SRA, Waldron KW, Bongaerts RJ, Mayer MJ, Narbad A. 2013. Spontaneous mutations reveals influence of exopolysaccharide on *Lactobacillus johnsonii* surface characteristics. PlosOne 8:e59957.

485 22. Maina NH, Tenkanen M, Maaheimo H, Juvonen R, Virkki L. 2008. NMR
486 spectroscopic analysis of exopolysaccharides produced by *Leuconostoc citreum* and
487 *Weissella confusa*. Carbohydr Res 343:1446-55.
488 23. Dertli E, Mayer MJ, Colquhoun IJ, Narbad A. 2016. *EpsA* is an essential gene in
489 exopolysaccharide production in *Lactobacillus johnsonii* FI9785. Microb Biotechnol
490 9:496-501.
491 24. Erickson AK, Jesudhasan PR, Mayer MJ, Narbad A, Winter SE, Pfeiffer JK. 2018.
492 Bacteria facilitate enteric virus co-infection of mammalian cells and promote genetic
493 recombination. Cell Host Microbe 23:77-88 e5.
494 25. Ardiccioni C, Clarke OB, Tomasek D, Issa HA, von Alpen DC, Pond HL, Banerjee S,
495 Rajashankar KR, Liu Q, Guan Z, Li C, Kloss B, Bruni R, Kloppmann E, Rost B,
496 Manzini MC, Shapiro L, Mancia F. 2016. Structure of the polyisoprenyl-phosphate
497 glycosyltransferase GtrB and insights into the mechanism of catalysis. Nat Commun
498 7:10175.
499 26. Inoue H, Suzuki D, Asai K. 2013. A putative bactoprenol glycosyltransferase, CsbB,
500 in *Bacillus subtilis* activates SigM in the absence of co-transcribed YfhO. Biochem
501 Biophys Res Commun 436:6-11.
502 27. Mavris M, Manning PA, Morona R. 1997. Mechanism of bacteriophage SfII-mediated
503 serotype conversion in *Shigella flexneri*. Mol Microbiol 26:939-50.
504 28. Guan S, Bastin DA, Verma NK. 1999. Functional analysis of the O antigen
505 glucosylation gene cluster of *Shigella flexneri* bacteriophage SfX. Microbiology 145
506 (5):1263-73.
507 29. PingitoreEsteban V, Pessione A, Fontana C, Mazzoli R, Pessione E. 2016.
508 Comparative proteomic analyses for elucidating metabolic changes during EPS

production under different fermentation temperatures by *Lactobacillus plantarum* Q823. Int J Food Microbiol 238:96-102.

30. Sturme MH, Nakayama J, Molenaar D, Murakami Y, Kunugi R, Fujii T, Vaughan EE, Kleerebezem M, de Vos WM. 2005. An agr-like two-component regulatory system in *Lactobacillus plantarum* is involved in production of a novel cyclic peptide and regulation of adherence. J Bacteriol 187:5224-35.

31. Liang H, Li L, Dong Z, Surette MG, Duan K. 2008. The YebC family protein PA0964 negatively regulates the *Pseudomonas aeruginosa* quinolone signal system and pyocyanin production. J Bacteriol 190:6217-27.

32. Schmid J. 2018. Recent insights in microbial exopolysaccharide biosynthesis and engineering strategies. Curr Opin Biotechnol 53:130-136.

33. Greenfield LK, Whitfield C. 2012. Synthesis of lipopolysaccharide O-antigens by ABC transporter-dependent pathways. Carbohydr Res 356:12-24.

34. Islam ST, Lam JS. 2014. Synthesis of bacterial polysaccharides via the Wzx/Wzy-dependent pathway. Can J Microbiol 60:697-716.

35. Hubbard C, McNamara JT, Azumaya C, Patel MS, Zimmer J. 2012. The hyaluronan synthase catalyzes the synthesis and membrane translocation of hyaluronan. J Mol Biol 418:21-31.

36. Whitney JC, Howell PL. 2013. Synthase-dependent exopolysaccharide secretion in Gram-negative bacteria. Trends Microbiol 21:63-72.

37. Eugster MR, Haug MC, Huwiler SG, Loessner MJ. 2011. The cell wall binding domain of *Listeria* bacteriophage endolysin PlyP35 recognizes terminal GlcNAc residues in cell wall teichoic acid. Mol Microbiol 81:1419-32.

38. Kho K, Meredith TC. 2018. Salt-induced stress stimulates a lipoteichoic acid-specific three component glycosylation system in *Staphylococcus aureus*. J Bacteriol 200:e00017-18.

39. Rismondo J, Percy MG, Grundling A. 2018. Discovery of genes required for lipoteichoic acid glycosylation predicts two distinct mechanisms for wall teichoic acid glycosylation. J Biol Chem 293:3293-3306.

40. Tytgat HL, Lebeer S. 2014. The sweet tooth of bacteria: common themes in bacterial glycoconjugates. Microbiol Mol Biol Rev 78:372-417.

41. Dertli E, Mayer MJ, Narbad A. 2015. Impact of the exopolysaccharide layer on biofilms, adhesion and resistance to stress in *Lactobacillus johnsonii* FI9785. BMC Microbiol 15:8.

42. Ou L, Ang L, Chujun Z, Jingyu H, Yongli M, Shenjing Y, Junhua H, Xu G, Yulong Y, Rui Y, Jinpan H, Bin D, Xiufang H. 2018. Identification and characterization of six glycosyltransferases involved in the biosynthesis of a new bacterial exopolysaccharide in *Paenibacillus elgii*. Appl Microbiol Biotechnol 102:1357-1366.

43. Williams DM, Ovchinnikova OG, Koizumi A, Mainprize IL, Kimber MS, Lowary TL, Whitfield C. 2017. Single polysaccharide assembly protein that integrates polymerization, termination, and chain-length quality control. Proc Natl Acad Sci U S A 114:E1215-E1223.

44. Gasson MJ. 1983. Plasmid complements of *Streptococcus lactis* NCDO 712 and other lactic streptococci after protoplast-induced curing. J Bacteriol 154:1-9.

45. Maguin E, Prevost H, Ehrlich SD, Gruss A. 1996. Efficient insertional mutagenesis in lactococci and other Gram-positive bacteria. J Bacteriol 178:931-5.

555 46. Holo H, Nes IF. 1989. High-frequency transformation, by electroporation, of
556 *Lactococcus lactis* subsp. *cremoris* grown with glycine in osmotically stabilized
557 media. Appl Environ Microbiol 55:3119-23.
558 47. Horn N, Wegmann U, Narbad A, Gasson MJ. 2005. Characterisation of a novel
559 plasmid p9785S from *Lactobacillus johnsonii* FI9785. Plasmid 54:176-83.
560 48. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ.
561 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search
562 programs. Nucleic Acids Res 25:3389-3402.
563 49. Wegmann U, Overweg K, Horn N, Goesmann A, Narbad A, Gasson MJ, Shearman C.
564 2009. Complete genome sequence of *Lactobacillus johnsonii* FI9785, a competitive
565 exclusion agent against pathogens in poultry. J Bacteriol 191:7142-3.
566 50. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes
567 S, Glass EM, Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL, Overbeek RA,
568 McNeil LK, Paarmann D, Paczian T, Parrello B, Pusch GD, Reich C, Stevens R,
569 Vassieva O, Vonstein V, Wilke A, Zagnitko O. 2008. The RAST Server: rapid
570 annotations using subsystems technology. BMC Genomics 9:75.
571 51. Darling ACE, Mau B, Blattner FR, Perna NT. 2004. Mauve: Multiple alignment of
572 conserved genomic sequence with rearrangements. Genome Res 14:1394-1403.
573 52. George DT, Stephenson DP, Tran E, Morona R, Verma NK. 2013. Complete genome
574 sequence of SfII, a serotype-converting bacteriophage of the highly prevalent *Shigella*
575 *flexneri* serotype 2a. Genome Announc 1:e00626-13.
576 53. Blattner FR, Plunkett G, 3rd, Bloch CA, Perna NT, Burland V, Riley M, Collado-
577 Vides J, Glasner JD, Rode CK, Mayhew GF, Gregor J, Davis NW, Kirkpatrick HA,
578 Goeden MA, Rose DJ, Mau B, Shao Y. 1997. The complete genome sequence of
579 *Escherichia coli* K-12. Science 277:1453-62.

580 54. Kunst F, Ogasawara N, Moszer I, Albertini AM, Alloni G, Azevedo V, Bertero MG,
581 Bessieres P, Bolotin A, Borchert S, Borriss R, Boursier L, Brans A, Braun M,
582 Brignell SC, Bron S, Brouillet S, Bruschi CV, Caldwell B, Capuano V, Carter NM,
583 Choi SK, Cordani JJ, Connerton IF, Cummings NJ, Daniel RA, Denziot F, Devine
584 KM, Dusterhoft A, Ehrlich SD, Emmerson PT, Entian KD, Errington J, Fabret C,
585 Ferrari E, Foulger D, Fritz C, Fujita M, Fujita Y, Fuma S, Galizzi A, Galleron N,
586 Ghim SY, Glaser P, Goffeau A, Golightly EJ, Grandi G, Guiseppi G, Guy BJ, Haga
587 K, et al. 1997. The complete genome sequence of the gram-positive bacterium
588 *Bacillus subtilis*. Nature 390:249-56.
589 55. Leonard MT, Valladares RB, Ardissone A, Gonzalez CF, Lorca GL, Triplett EW.
590 2014. Complete genome sequences of *Lactobacillus johnsonii* strain N6.2 and
591 *Lactobacillus reuteri* strain TD1. Genome Announc 2:e00397-14.
592 56. Guinane CM, Kent RM, Norberg S, Hill C, Fitzgerald GF, Stanton C, Ross RP. 2011.
593 Host specific diversity in *Lactobacillus johnsonii* as evidenced by a major
594 chromosomal inversion and phage resistance mechanisms. PLoS One 6:e18740.
595 57. Pridmore RD, Berger B, Desiere F, Vilanova D, Barretto C, Pittet AC, Zwahlen MC,
596 Rouvet M, Altermann E, Barrangou R, Mollet B, Mercenier A, Klaenhammer T,
597 Arigoni F, Schell MA. 2004. The genome sequence of the probiotic intestinal
598 bacterium *Lactobacillus johnsonii* NCC 533. Proc Natl Acad Sci U S A 101:2512-7.

599

600 **Figures and tables**

601 **Figure 1 Volcano plots of differentially expressed proteins.** Results compare *L. johnsonii*
602 Δ*eps_cluster* (DC) versus FI9785 (WT), obtained by two-sided t test, p value less than 0.05
603 in **A**, iTRAQ and **B**, label free experiments. Red, abundance higher in DC than WT; green,
604 abundance lower in DC than WT.

605 **Figure 2 Gene Ontology analyses of differentially expressed proteins.** On the x axis the
606 gene ontology enriched terms are shown, and on the y axis the percentage of enrichment. Up,
607 processes enriched in the Δ*eps_cluster* mutant compared to the WT; Down, processes
608 enriched in the WT compared to the mutant.

609 **Figure 3 NMR analysis of pellet-associated EPS.** 600 MHz ^{1}H NMR spectra of EPS from
610 WT and modified *L. johnsonii* (pellet samples, D_2O, 338°K). Anomeric signals of EPS1 and
611 EPS2 are labelled '1' and '2' respectively, imp = impurities. The presence of EPS1 is
612 indicated by two H1 signals of equal intensity at 5.17 ppm ((1,2,6)-α-Glc) and 5.11 ppm (t-α-
613 Glc). There are multiple H1 signals associated with EPS2 as indicated at the chemical shifts
614 listed previously (20, 23).

615 **Figure 4 Amino acid alignments with GtrA and GtrB proteins. A**, Translation of *242*
616 coding sequence (WP_012845545) aligned with GtrB proteins from *Shigella* phage SfII
617 (YP_008318506 (52)), *E. coli* K12 (P77293 (53)), *B. subtilis* CsbB (Q45539 (54)) and
618 *Synechocystis* sp. (Q55487, 5EKP (25)). Conserved motifs DXD and DXSXD are underlined,
619 residues affecting activity in 5EKP are marked #. **B**, Translation of *241* coding sequence
620 (WP_004896037) aligned with GtrA family proteins from *Shigella* phage SfII
621 (YP_008318507 (52)) and *E. coli* K12 (P77682 (53)). Black, 100%, dark grey 80%, light
622 grey 60% homology.

623 **Figure 5 Conservation of genes with *L. johnsonii* strains from different environments.**
624 **A**, ORFs are shown from genomes of *L. johnsonii* strains FI9785 (FN298497 (49),
625 nucleotides 184194-194938, loci FI9785_RS00820 to FI9785_RS00875); UMNLJ22
626 (NZ_CP021704, Johnson T. J. and Youmans B., unpublished, nucleotides 699996-711750,
627 loci A3P32_RS03290 to A3P32_RS03350); N6.2 (NC_022909 (55), nucleotides 210473-
628 221016, loci T285_RS00860 to T825_RS00915); DPC6026 (NC_017477 (56), nucleotides
629 202698-210932, loci LJP_RS00920 to LJP_RS00960); NCC533 (NC_005362 (57),

630 nucleotides 196136-202659, loci LJ_RS00845 to LJ_RS00880); Byun-jo-01 (NZ_CP029614,
631 Kim D., unpublished, nucleotides shown in complement 1111505 to 1117990, loci
632 C0060_RS05265 to C0060_RS05300), with the GtrA-GtrB pairs aligned. **B**, nucleotide
633 alignment of sequences in **A** with Mauve to indicate areas of high sequence conservation.
634 HicB, Hic B family antitoxin; phage tail, putative phage tail-related protein; HP, hypothetical
635 protein; 30S, 30S ribosomal protein S14; MFS transporter, major facilitator family
636 transporter; sug-trans, sugar transporter.

637 **Figure 6 NMR analysis of pellet-associated EPS showing effect of *241* deletion and**
638 **complementation.** 600 MHz ^{1}H NMR spectra of EPS from WT and engineered *L. johnsonii*
639 (pellet samples, D_2O, 300°K). Anomeric signals of EPS1 and EPS2 are labelled '1' and '2'
640 respectively, imp = impurities.

641 **Figure 7 Phenotypic characterisation of *241* and *242* deletion. A, B,** growth of *L.*
642 *johnsonii* strains in liquid at 37°C showing increase in optical density; **C,** aggregation of
643 overnight cultures; **D,** differences in colony size in strains given the same incubation time at
644 37°C, **E,** TEM analysis of cells from overnight cultures (bar = 200 nm); WT, wild type.

645

646

647 **Table 1** Quantified *Lactobacillus johnsonii* proteins, using MaxQuant software in iTRAQ and label free experiments

Protein Accession number	Protein name	Gene name	Razor + unique peptides	Mol. weight [kDa]	Score	log2 (D/WT)	ratio D/WT	ratio WT/D	iTRAQ	label free	GO Biological process
D0R1P2	Uncharacterized protein	FI9785_401	4	22	234	3.71	13.12	0.08	x		-
D0R498	Thiol peroxidase	tpx	3	18	29	2.41	5.30	0.19		x	cell redox homeostasis; oxidation/reduction process; cellular oxidant detoxification
D0R5C5	Ribosomal silencing factor RsfS	rsfS	3	14	111	2.22	4.67	0.21	x		Mature ribosome assembly; negative regulation of ribosome biogenesis; negative regulation of translation; regulation of translation
D0R3E4	50S ribosomal protein L28	rpmB	3	7	26	2.17	4.49	0.22	x		translation
D0R3T4	Aspartate--tRNA ligase	aspS	26	71	155	1.38	2.60	0.39		x	translation; tRNA aminoacylation for protein translation; aspartyl-tRNA aminoacylation
D0R3V1	Glycine--tRNA ligase beta subunit	glyS	16	78	150	1.23	2.35	0.43		x	translation; arginyl tRNA aminoacylation; glycyl tRNA aminoacylation
D0R277	Uncharacterized protein	FI9785_219	7	17	87	1.16	2.24	0.45		x	-
D0R5K0	Aspartyl/glutamyl-tRNA(Asn/Gln) amidotransferase subunit C	gatC	3	12	42	1.12	2.17	0.46		x	translation; regulation of translational fidelity
D0R6B7	Deoxynucleoside kinase	dgk1	10	25	60	1.05	2.07	0.48		x	nucleobase-containing compound metabolic process; deoxyribonucleoside monophosphate biosynthetic process; nucleotide biosynthetic process; phosphorylation
D0R362	Ribokinase	rbsK	16	33	146	1.04	2.05	0.49		x	carbohydrate metabolic process; D-ribose metabolic process ; phosphorylation ; D-ribose catabolic process; carbohydrate phosphorylation
D0R1J3	30S ribosomal protein S12	rpsL	9	15	134	0.93	1.90	0.53	x	x	translation
D0R2C4	Recombination protein RecR	RecR	2	22	44	0.89	1.85	0.54		x	DNA repair; DNA recombination; cellular response to DNA damage stimulus
D0R4S3	Isoleucine--tRNA ligase	ileS	16	107	100	0.81	1.75	0.57		x	translation; tRNA aminoacylation for protein translation; isoleucyl tRNA aminoacylation;aminoacyl-tRNA metabolism involved in translational fidelity
D0R1L4	30S ribosomal protein S5	rpsE	15	19	278	0.77	1.71	0.59		x	translation
D0R5T0	Uncharacterized protein	FI9785_1588	10	21	53	0.76	1.70	0.59		x	-
D0R434	Asparagine--tRNA ligase	asnS	18	50	72	0.74	1.67	0.60		x	translation; tRNA aminoacylation for protein translation; asparagyl tRNA aminoacylation
D0R3G2	30S ribosomal protein S16	rpsP	5	11	227	0.73	1.66	0.60		x	translation
D0R5D2	50S ribosomal protein L35	rpmL	7	8	80	0.69	1.61	0.62		x	translation
D0R4W9	ATP synthase subunit b	atpF	7	18	42	0.65	1.57	0.64		x	ATP biosynthetic process; ion transport; ATP synthesis coupled proton transport; ATP hydrolysis coupled cation transmembrane transport

D0R5D1	50S ribosomal protein L20	rplT	8	13	101	0.63	1.55	0.65		x	ribosomal large subunit assembly; translation
D0R608	Chromosome partitioning protein ParB	parB	6	33	51	-3.26	0.10	9.60	x		
D0R4H4	Pseudouridine synthase	FI9785_1123	5	27	115	-3.02	0.12	8.09	x		psuedouridine synthesis; RNA modification
D0R5U7	Elongation factor P	efp	8	21	105	-2.99	0.13	7.96	x		translation; translational elongation; peptide biosynthetic process
D0R1R2	Putative glycosyl transferase	FI9785_242	8	35	46	-2.50	0.18	5.64	x	x	-
D0R254	Extracellular solute-binding protein PhnD	phnD	4	34	86	-2.26	0.21	4.78	x		transmembrane transport
D0R5M6	Aggregation promoting factor	apf2	3	33	190	-2.04	0.24	4.12		x	-
D0R1L2	50S ribosomal protein L6	rplF	12	19	205	-2.01	0.25	4.04	x		translation
D0R588	Peptide chain release factor 3	prfC	10	59	63	-1.99	0.25	3.97	x		translation; translational termination; regulation of translational termination
D0R5Z4	Tagatose-6-phosphate kinase	fruB	4	33	153	-1.96	0.26	3.88	x		carbohydrate metabolic process; lactose metabolic process; phosphorylation; carbohydrate phosphorylation
D0R4W2	MreB-like protein	mbl	8	35	318	-1.83	0.28	3.56	x		cell morphogenesis
D0R4I9	Ribonuclease Z	rnZ	4	35	17	-1.82	0.28	3.52	x		tRNA processing; tRNA 3'-trailer cleavage, endonucleolytic; tRNA 3'-trailer cleavage; RNA phosphodiester bond hydrolysis, endonucleolytic
D0R1L5	50S ribosomal protein L30	rpmD	2	6	37	-1.81	0.29	3.50	x	x	translation
D0R1U3	Tryptophan--tRNA ligase	trpS	8	39	269	-1.80	0.29	3.47	x		translation; tRNA aminoacylation for protein translation; tryptophanyl tRNA aminoacylation
D0R5K3	ATP-dependent DNA helicase	pcrA	18	84	294	-1.60	0.33	3.04	x		DNA unwinding involved in DNA replication
D0R2Q2	30S ribosomal protein S6	rpsF	9	12	255	-1.60	0.33	3.03	x		translation
D0R268	Putative secreted protein	FI9785_210	21	102	323	-1.55	0.34	2.93		x	-
D0R3I4	Proline--tRNA ligase	proS	16	63	297	-1.38	0.39	2.60		x	translation; tRNA aminoacylation for protein translation;prolyl tRNA aminoacylation; aminoacyl-tRNA metabolism involved in translational fidelity
D0R4U6	Valine--tRNA ligase	valS	20	101	323	-1.37	0.39	2.59	x		translation; tRNA aminoacylation for protein translation;valyl tRNA aminoacylation; aminoacyl-tRNA metabolism involved in translational fidelity
D0R662	Phosphoenolpyruvate-dependent sugar phosphotransferase system EIIAB, probably mannose specific	manL	13	36	323	-1.16	0.45	2.24		x	Phosphoenolpyruvate-dependent sugar phosphotransferase system; carbohydrate transmembrane transport
D0R1P7	Muramidase	FI9785_225	8	64	323	-1.03	0.49	2.04		x	metabolic process; peptidoglycan catabolic process; cell wall macromolecule catabolic process
D0R3J0	Translation initiation factor IF-2	infB	13	99	122	-1.02	0.49	2.03		x	translation; translational initiation

D0R383	Methionine aminopeptidase	pepM	10	30	323	-0.96	0.52	1.94	x		proteolysis; protein initiator methionine removal
D0R501	Probable transcriptional regulatory protein FI9785_1304	FI9785_1304	5	27	99	-0.88	0.54	1.84		x	regulation of transcription, DNA-templated
D0R4I0	Pyruvate kinase	pyk	35	64	323	-0.88	0.54	1.84		x	glycolytic process; phosphorylation
D0R395	Aminopeptidase	pepN	13	96	140	-0.87	0.55	1.83		x	proteolysis
D0R2B3	50S ribosomal protein L10	rplJ	9	21	191	-0.85	0.56	1.80		x	translation; ribosome biogenesis
D0R3F4	Oligopeptide-binding protein OppA	OppA	7	65	50	-0.81	0.57	1.76	x		transmembrane transport
D0R1L8	Adenylate kinase	adK	13	24	323	-0.74	0.60	1.67		x	nucleobase-containing compound metabolic process; nucleotide biosynthetic process; phosphorylation; AMP salvage; nucleoside monophosphate phosphorylation
D0R5L2	NH(3)-dependent NAD(+) synthetase	nadE	6	31	149	-0.71	0.61	1.63		x	NAD biosynthetic process

WT, *L. johnsonii* FI9785; D, Δ*eps_cluster*; GO, gene ontology; -, no process identified

650 **Table 2** List of *L. johnsonii* strains created and used in this study

651

Strain	Genotype	Description	Plasmid	reference
FI9785	wild type	poultry isolate		(18)
FI10754	Δ*eps_cluster*	*eps* gene cluster deleted		(20)
FI11504	Δ*242*	FI9785 with *242* gene deleted		this study
FI11646	Δ*242*-p*242*	FI11504 complemented with the *242* gene in expression plasmid pFI2560	pFI2843	this study
FI11647	Δ*242*-pFI2560	FI11504 with pFI2560 empty vector control	pFI2560	this study
FI11669	Δ*241*	FI9785 with *241* gene deleted		this study
FI11670	Δ*241*-p*241*	FI11669 complemented with the *241* gene in expression plasmid pQI0001	pQI0002	this study
FI11671	Δ*241*-pQI0001	FI11669 with pQI0001 empty vector control	pQI0001[a]	this study
FI10785	Δ*epsA*	*epsA* transcriptional regulator from *eps* gene cluster deleted		(23)
FI10844	Δ*epsE*	*epsE* priming glycosyltransferase from *eps* gene cluster deleted		(21)

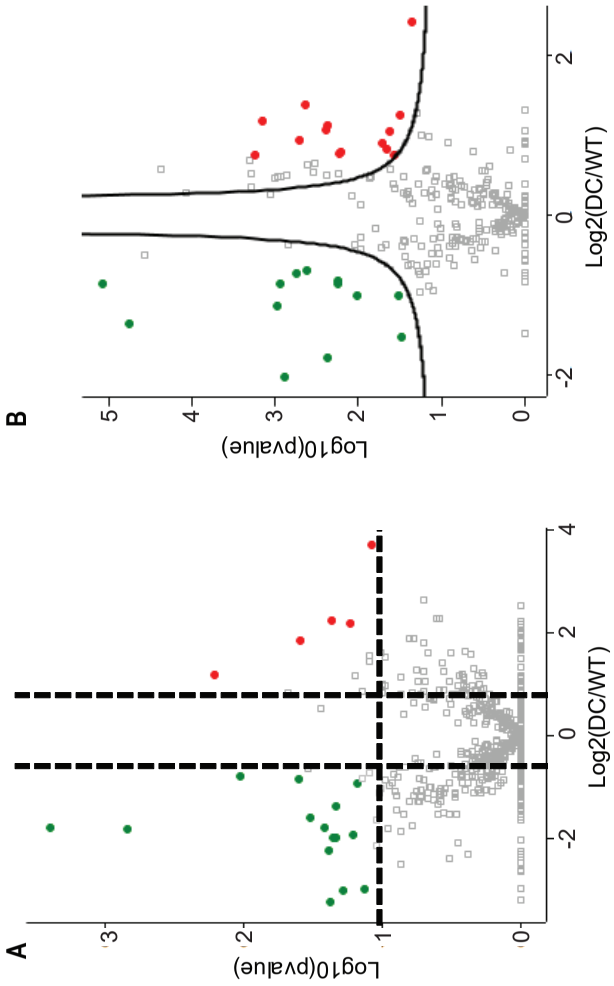
652 [a]plasmid pFI2560 with cloning site NcoI altered to NdeI-BamHI
653

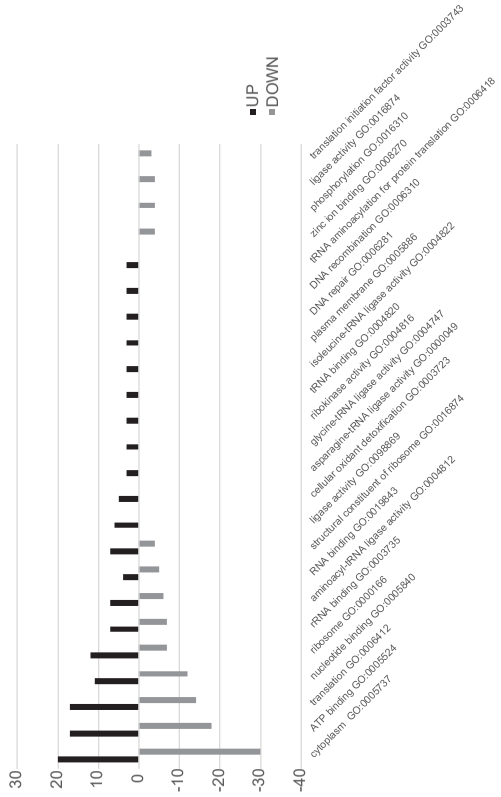
654 **Table 3** Oligonucleotide primers used for creation of deletion constructs and plasmids and
655 assessment of sequence integrity, integration and excision.

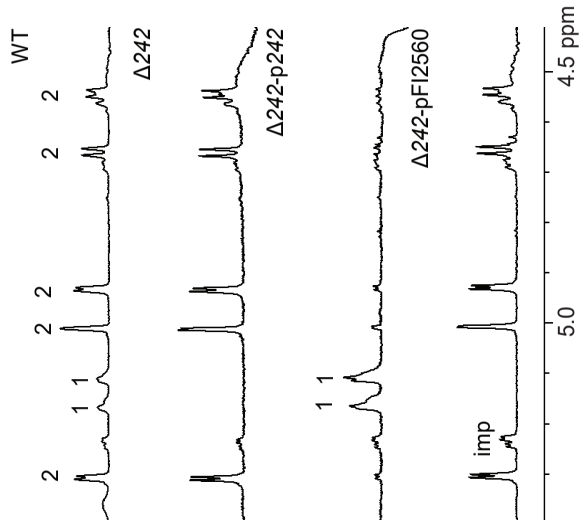
Primer	Sequence 5'-3'
241Eco_F	GAT**G**AATTCACGCTGCTTAG
241splice243_R	**CGGCTTTTTGTCATAT**ACTTTAACAGTCTTTCTTAT
243Spe_R	CT**AC**TAGTCATGATTGATTTTGGT
243splice241_F	**AGAAAGACTGTTAAAGT**ATATGACAAAAAGCCGA
241_IF	GCTTCTACGTCACCAGCTTCT
243_IR	TCCACAGTTTCGAACTGGTG
240_F	ATGTCTAAAGTGTGACTATATGTT
240splice242_R	**TACTTTAACAGTCTTTCTTA**GGCTTATTTTCCCTTCT
242splice240_F	**AGAAGGGAAAATAAGCC**TAAGAAAGACTGTTAAAGTA
242Spe_R	CATTTGAC**T**AGTCATCATTCGGTAGTC
240_IF	GAATGTCTAAAGTGTGACTATATGTT
242_IR	ACGGTTGTATTCAGGCATATTC

pGhost1	AGTCACGACGTTGTAAAACGACG
pGhostR	TACTACTGACAGCTTCCAAGG
pForVec	ACAGCAATGTTACAAGTTGAAAT
p181	GCGAAGATAACAGTGACTCTA
242_COD2F	AAAAAATTATCAATTATAGTTCCTTG
242_C_R	GAAGCTCCACGTGAACTTC
241_NdeF	TAA**CATA**TGGGTATTTTTAAAAGAATAC
241_BamR	TTT**GG**AT**C**CTTTAACAGTCTTTCTTATTAC

Mismatching base pairs to insert restriction sites or for splice overlap extension are in bold







A

```
                  *        20         *        40         *        60         *        80
FI9785_242 : -----MKKLSIIVPCYNEEESVPLFYPAVNKVMDT-IPDLEPEYWIINDGSKDNTLKEIKELRKKLPEHVHFVSFSRNFGKESALY :  80
SfII_GtrB  : ------MKISLVVPVFNEEEAIPVFYKTVREFQEL--KPYEVELVFINDGSKDATESIINALAVSD-PLVVPLSFTRNFGKEPALF :  77
Ecoli_GtrB : ------MKISLVVPVFNEEEAIPIFYKTVREFEEL--KSYEVELVFINDGSKDATESIINALAVSD-PLVVPLSFTRNFGKEPALF :  77
Bsub_CsbB  : ---MKQGLISIIIPSYNEGYNVKLIHESLKKEFKN--IHYDYEIFIINDGSVDDTLQQIKDLAATC-SRVKYISFSRNFGKEAAIL :  80
Syne_GtrB  : ----MTIELSIVIPMYNEEDNLEHLYARLLEVLTP--LKITYEIICVNDGSKDKTLKQLIDCYQSN-RQIKIVNLSRNFGKEIALS :  79

                *       100        *       120        *       140        *       160        *
FI9785_242 : AGIQATGDYVVVMDVDLQDPPKFLPQMYDLIKTGEYDCIGTRRVDRTGEAKFKSFLSDMFYKVVNKISDTEIVPGARDYRMMDRQ : 166
SfII_GtrB  : AGIDHASGDAVIPIDVDLQDPIEVIPHLIEKWQAG-ADMVLAKRSDRSTDGRLKRKTAEWFYKLHNKISTPKIEENVGDFRLMSRE : 162
Ecoli_GtrB : AGIDHATGDAIIPIDVDLQDPIEVIPHLIEKWQAG-ADMVLAKRSDRSTDGRLKRKTAEWFYKLHKKISNPKIEENVGDFRLMSRD : 162
Bsub_CsbB  : AGFEHVQGEAVIVMDADLQHPTYLLKEFIKGYEEG-YQQVIAQRNRK-GDSFVRSLLSSMYYKFINKAVEVDLRDGVGDFRLLSRQ : 164
Syne_GtrB  : AGIDYAQGNAVIPIDADLQDPPELIHELVDKWREG-YDIVYATRRSRQGETWVKQFTAKMFYKVIGRMTEIKIPPNTGDFRLMDRK : 164

                180        *       200        *       220        *       240        *       2
FI9785_242 : WVDAVLNMEEYNRESKGIFSWVGFKTKYLDYHNVERVACESDWSTWKLFKYAMDGIADFSQAPLNLAVWLGTGSFILSLIGLIAVI : 252
SfII_GtrB  : WVENIKLLERNLEMKGILSWVGGQTDVVEYVRARRVACISKFNGWKLWNLALEGITGFSTFPLRVWTYLGLFVASLSFLYGAWMI : 248
Ecoli_GtrB : WVENIKLMERNLEMKGILSWVGGKTDIVEYVRARRIACDTKFNGWKLWNLALEGITGFSTFPLRIWTYLGLVVASVAFIYGAWMI : 248
Bsub_CsbB  : AVNALLKLSEGNRESKGLFCWIGFDQKIVEYENVERKNGTSKWSFSSLFYYGMDGVVGFNHKPLRLCFYTGIFILLLSIIYIIATF : 250
Syne_GtrB  : VVNAIKQLERTREMKGLFAWVGYRQTFVLYDREFRFGQSKYNYWKLWNFALDGIFGFSLLPLKVWTYIGSIISLLSLAVASFLI : 250

                60         *       280        *       300        *       320        *       340
FI9785_242 : IRRVLYPGSSIFGWASMVCIILLGGLQLICIGILGKYIGRIYIQVKNRPIYIIEEKK------------------------ : 310
SfII_GtrB  : IDILVF-GNPVRGYFSLLVSILFLGGVQLICIGVLGEYIGRIYIEVKNRPKYIIKSHRGNP------------------- : 309
Ecoli_GtrB : LDIIIF-GNAVRGYPSLLVSILFLGGIQMICIGVLGEYIGRTYIETKKRPKYIIKRVKK--------------------- : 306
Bsub_CsbB  : VKILTN-GISVPGYFTIISAVLFLGGVQLLSLGIIGEYIGRIYYETKKRPHYLIKEANIPNKDLPETNELKSMRRLTKMH-- : 329
Syne_GtrB  : LKTITL-GVDVPGYASLMVAILFLGGVQLISLGVIGEYLGRIYEEVKARPIYLVSDLWGLEYLPLEKLN------------ : 318

                *       100        *       120        *       140
FI9785_241 : QTTTLFTTVCLFIFVTQLGYNQLIVKAVSTFIATLIQFVINKLLTFRKIKTTKPKVDVRK : 145
SfII_GtrA  : ILS------VIVGWAADKGSLPFIVTLITSSIISIVCGFVYSKIVFRDAK---------- : 120
Ecoli_GtrA : TLS------ATVGWAADRGALPFMITLVTFSIISIVCGFVYSKIVFRDAK---------- : 120
```