Plasmid and chromosome-encoded adhesion-related genes of *Lactobacillus fermentum* revealed by genome sequencing

B. Lehri, A. M. Seddon and A. V. Karlyshev*

School of Life Sciences, SEC Faculty, Kingston University, Kingston upon Thames, KT1 2EE

*Corresponding author

a.karlyshev@kingston.ac.uk

Abstract

In this report we describe a *Lactobacillus fermentum* 3872 plasmid (pLF3872) not previously found in any other strain of this species. The analysis of the complete sequence of this plasmid revealed the presence of a gene encoding a large collagen binding protein (CBP), as well as the genes responsible for plasmid maintenance and conjugation. Potential roles of CBP and a chromosomally encoded fibronectin-binding protein (FbpA) in probiotic activity are discussed.

Keywords: probiotics; *Lactobacillus fermentum*; adhesion; conjugative plasmid; collagen binding protein; host cell receptors; peptidoglycan hydrolase

Introduction

Regarded as beneficial and health promoting microorganisms, probiotics have been widely used for commercial purposes (Marco *et al.*, 2006). Currently, there is increased interest in using probiotics to treat medical conditions such as allergic diseases, hypercholesterolaemia, and as additives/alternatives to antibiotic treatments (Yang *et al.*, 2013; Tomaro *et al.*, 2014; Oelschlaeger, 2010; Angelakis *et al.*, 2013). Among the most widely used probiotics are

lactic acid bacteria, particularly *Lactobacillus* spp, which are commonly found in humans as commensal microorganisms making them good candidates for probiotic research (Ljungh & Wadstrom, 2006). In particular, beneficial properties of *L. fermentum* strain CECT 5716 have been reported (Mane et al., 2009*)*. Although many strains of *Lactobacillus* spp. have GRAS (Generally Recognised as Safe) status, *L. fermentum* AGR1487 induced negative changes in gut epithelia (not observed with *L. fermentum* AGR1485) suggesting that safety and beneficial properties of probiotic bacteria could be strain- and not just genus- or species-related (Anderson *et al.*, 2013).

Probiotics provide their benefits through immune modulation, release of metabolites, and/or attachment to host cells (Oelschlaeger, 2010). These factors, particularly those involved in adhesion, are genus-, species- and even strain-specific. Expression of specific adhesins allows probiotics to colonise and stay within the host, while exerting anti-adhesive effects on other bacteria (Ljungh & Wadstrom, 2006; M. Andrea Azcárate-Peril *et al.*, 2011; Ouwehand *et al.*, 2002). Some of the genes involved in probiotic action, may be carried by plasmids (Ainsworth *et al.*, 2014).

Our previous analysis of a draft genome sequence of the *Lactobacillus fermentum* strain 3872 revealed a fragment of a Collagen Binding Protein (CBP)-encoding gene (Karlyshev et al., 2013), which in the current study was found to be located on a plasmid (pLF3872). In this report we describe a complete sequence and genetic organisation of this plasmid, as well as present an update on the whole genome assembly. Potential contribution of plasmid- and chromosome- encoded adhesins to the beneficial properties of this strain is discussed.

Materials and methods

Genome sequencing was conducted using the Ion Torrent Personal Genome Machine, 400 bp kit and 314v2 chip (Life Technologies). Contigs generated by Torrent Assembler and CLC

Genomics Workbench (GWB) were combined using CISA contig integrator (Lin & Liao, 2013) and verified by read mapping using GWB. The plasmid-related contigs were identified using NCBI Blast similarity search tool, which revealed similarity with plasmids plca36 (*L. casei* Zhang), and pWCFS103 (*L. plantarum* WCFS1). Consensus sequences generated by mapping reads onto a closely related plasmid sequence plca36 using CLC genomics workbench were merged with Torrent assembled contigs using CISA contig integrator (Lin & Liao, 2013) producing a contiguous sequence of the plasmid, named pLF3872. The latter, as well as the chromosomal genome sequence of *L. fermentum* 3872 were annotated using RAST (Overbeek *et al.,* 2014), as well as NCBI automatic gene annotation pipeline. The coding sequences were also verified using Artemis software (Rutherford *et al.,* 2000). This Whole Genome Shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession AVCT00000000. The version described in this paper is version AVCT02000000.

Results & Discussion

The genome sequence of *L. fermentum* 3872 reported in this study consists of 16 chromosomal contigs and one plasmid sequence, with the total size of 2.5Mb. Despite the presence of plasmids in some strains of *L. ferementum* (such as e.g. seven plasmids detected in *L. fermentum* MTCC 8711 (Jayashree *et al.,* 2013)) none of them are related to the 32.6 kb plasmid pLF3872 we found in *L. fermetnum* 3872. The G+C content of this plasmid (40.1%) is remarkably lower than the average G+C content of the entire genome of this strain (50.1%), suggesting its possible recent acquisition from another bacterium. This plasmid contains 33 genes (Fig.1 and Table 1), twelve of which are hypothetical. There are some conjugation-related genes, such as *traA, trsC, trsJ, trsF, trsL, trsE, trsD* and *traG,* which may be involved in type four secretion (Alvarez-Martinez & Christie, 2009, Morton *et al.,* 1993, Morton *et al.,* 1993, Laverde *et al.,* 2014). A gene encoding a peptidoglycan hydrolase

(TcpG), containing a lytic transglycosylase and amidase-5 domains, may also be involved in conjugation (Scheurwater *et al.*, 2008; Laverde *et al.*, 2014; Bantwal *et al*., 2012).

A toxin-antitoxin gene pair *eatL-zetL* present in pLF3872 is possibly required for stable maintenance of the plasmid (Zielenkiewicz *et al.*, 2009). In addition, an antitoxin encoding gene *dinJ* is also present (Hu *et al.*, 2012). Despite the presence of 'antitoxin' encoding genes in both pLF3872 and plca36 plasmids, there is very little sequence similarity between them and the respective gene products (Fig. 1S). Both proteins contain Relb superfamily domains (Fig. 2S) suggesting similarity in their functions. Functional similarity in the absence of sequence similarity can also be revealed between *parA* genes carried by these plasmids.

Plasmid pLF3872 contains a collagen-binding protein-encoding gene (*cbp*), which has not been detected in any other tested strain of *L. fermentum.* An orthologue of this gene in *Lactobacillus plantarum* 91 is known to be responsible for anti-adhesive activity of against *E. coli* 0157:H7 (Yadav *et al*., 2013).

The CBP protein consists of an N-terminal binding region 'A' and a repetitive C-terminal region 'B', forming stalks presenting the 'A' region for adhesion (Deivanayagam *et al.,* 2000). The 'B' region may provide added stability in anchoring to the host via increased protection from host proteases (Deivanayagam *et al.*, 2000). The C-terminal LPXT domain may be required for cell wall anchoring (Fig. 2a) (Davies *et al.*, 2009).

The chromosomally-located *fbpA* gene of strain 3872, which is highly conserved in various *L. fermentum* strains, belongs to a recombinatorial zone of *Streptococcus pyogenes* (Fig. 2b). The FbpA protein of *S. pyogenes* plays a role in adhesion and colonisation (Yamaguchi *et al.,* 2013), and consists of an N-terminal domain responsible for adhesion, and a conserved C-terminal DUF184 domain with no known function.

Collagen, fibronectin, and fibrinogen make up the extracellular matrix (ECM), and are ubiquitously found within the human body. Proteins binding to the ECM are known as

microbial surface components recognising adhesive matrix molecules (MSCRAMMs). In Gram-positive pathogenic bacteria, such as *S. pyogenes*, *Staphylococcus aureus* and *Arcanobacterium pyogenes*, MSCRAMM proteins play a role in the initial step of colonization (Yamaguchi *et al*., 2013; Pietrocola *et al.*, 2007; Ponnuraj *et al.*, 2003; Foster & Höök, 1998). The presence of FbpA and CBP may increase the adhesive properties of *L. fermentum* 3872 allowing it to compete against pathogenic bacteria that have an affinity towards similar target proteins. The FbpA and CBP-mediated adhesion in a close proximity to a pathogen might assist in elimination of the latter via production of anti-bacterials such as hydrogen peroxide known to be released by *L .fermentum* 3872.

Our study suggests a potential role of plasmids in the provision of beneficial properties to probiotic bacteria. The plasmid- and chromosomally-encoded adhesins of *L. fermentum* 3872 may have a synergistic effect on bacterial binding to host cell tissues, which may thus increase bacterial survival and competiveness of this probiotic microorganism against other (including potentially pathogenic) bacteria. The results of this study will assist in the development of novel antibacterials.

Acknowledgments

This work was not supported by any external funding.

Conflicts of interests

The authors declare no conflicting interests related to this publication

Genbank accession numbers

Plasmids

plca36 (CP000935.1), pWCFS103 (CR377166.1), plasmid 1 (CP002392.1)

Chromosomes

L. casei Zhang (CP001084.1), *L. fermentum* F6 (CP005958.1), *L. fermentum* 5716 (CP002033.1), *L. fermentum* 3916 (AP008937.1), *L. plantarum* WCFS1(AL935263.2), *Lactobacillus paracasei* NFBC338 contig 1 (AAO43108)

References

Ainsworth, S., Stockdale, S., Bottacini, F., Mahony, J. and van Sinderen, D. (2014) 'The *Lactococcus lactis* plasmidome: much learnt, yet still lots to discover', *FEMS Microbiology Reviews,* 38 (5), pp. 1066-1088.

Anderson, R. C., Young, W., Clerens, S., Cookson, A. L., McCann, M. J., Armstrong, K. M. and Roy, N. C. (2013) 'Human Oral Isolate *Lactobacillus fermentum* AGR1487 Reduces Intestinal Barrier Integrity by Increasing the Turnover of Microtubules in Caco-2 Cells', *Plos One,* 8 (11), pp. e78774.

Alvarez-Martinez, C. E. and Christie, P. J. (2009) 'Biological Diversity of Prokaryotic Type IV Secretion Systems', *Microbiology and Molecular Biology Reviews,* 73 (4), pp. 775-808.

Angelakis, E., Merhej, V. and Raoult, D. (2013) 'Related actions of probiotics and antibiotics on gut microbiota and weight modification', *The Lancet Infectious Diseases,* 13 (10), pp. 889-899.

Bantwal, R., Bannam, T. L., Porter, C. J., Quinsey, N. S., Lyras, D., Adams, V. and Rood, J. I. (2012) 'The peptidoglycan hydrolase TcpG is required for efficient conjugative transfer of pCW3 in *Clostridium perfringens*', *Plasmid,* 67 (2), pp. 139-147.

Broccoli, S., Phoenix, P. and Drolet, M. (2000) 'Isolation of the topB gene encoding DNA topoisomerase III as a multicopy suppressor of *topA* null mutations in *Escherichia coli*', *Molecular Microbiology,* 35 (1), pp. 58-68.

151 Carver TJ, Rutherford KM, Berriman M, Rajandream MA, Barrell BG and Parkhill J (2005)
152 'ACT: the Artemis Comparison Tool', *Bioinformatics (Oxford, England)* 21(16), pp. 3422-
153 3423.
154 Davies, J. R., Svensäter, G. and Herzberg, M. C. (2009) 'Identification of novel LPXTG-
155 linked surface proteins from *Streptococcus gordonii*', *Microbiology*, 155 (6), pp. 1977-1988.
156 Deivanayagam, C. C., Rich, R. L., Carson, M., Owens, R. T., Danthuluri, S., Bice, T., Höök,
157 M. and Narayana, S. V. (2000) 'Novel fold and assembly of the repetitive B region of the
158 *Staphylococcus aureus* collagen-binding surface protein', *Structure*, 8 (1), pp. 67-78.
159 Di´az-López, T., Lages-Gonzalo, M., Serrano-López, A., Alfonso, C., Rivas, G., Di´az-
160 Orejas, R. and Giraldo, R. (2003) 'Structural Changes in RepA, a Plasmid Replication
161 Initiator, upon Binding to Origin DNA', *Journal of Biological Chemistry*, 278 (20), pp.
162 18606-18616.
163 Foster, T. J. and Höök, M. (1998) 'Surface protein adhesins of *Staphylococcus aureus*',
164 *Trends in Microbiology*, 6 (12), pp. 484-488.
165 Jayashree, S., Pooja, S., Pushpanathan, M., Vishnu, U., Sankarasubramanian, J., Rajendhran,
166 J. and Gunasekaran, P. (2013) 'Genome Sequence of *Lactobacillus fermentum* Strain MTCC
167 8711, a Probiotic Bacterium Isolated from Yogurt', Genome Announcements, 1 (5),
168 DOI: 10.1128/genomeA.00770-13
169 Karlyshev, A. V., Raju, K. and Abramov, V. M. (2013) 'Draft Genome Sequence of
170 *Lactobacillus fermentum* Strain 3872', *Genome Announcements*, 1 (6),
171 doi:10.1128/genomeA.01006-13.
172 Laverde Gomez, J. A., Bhatty, M. and Christie, P. J. (2014) 'PrgK, a Multidomain
173 Peptidoglycan Hydrolase, Is Essential for Conjugative Transfer of the Pheromone-
174 Responsive Plasmid pCF10', *Journal of Bacteriology*, 196 (3), pp 527-539 .

Lin, S. and Liao, Y. (2013) 'CISA: Contig Integrator for Sequence Assembly of Bacterial Genomes', *Plos One,* 8 (3), art. no. e60843.

Ljungh, A. and Wadstrom, T. (2006) 'Lactic acid bacteria as probiotics', *Current Issues in Intestinal Microbiology,* 7 (2), pp. 73-89.

Azcárate-Peril, M. A., Sikes, M. and Bruno-Bárcena, J. M. (2011) 'The intestinal microbiota, gastrointestinal environment and colorectal cancer: a putative role for probiotics in prevention of colorectal cancer?' *American Journal of Physiology,* 301 (3) pp. G401-G424.

Mane, J., Loren, V., Pedrosa, E., Ojanguren, I., Xaus, J., Cabre, E., Domenech, E. and Gassull, M. A. (2009) '*Lactobacillus fermentum* CECT 5716 prevents and reverts intestinal damage on TNBS-induced colitis in mice', *Inflammatory Bowel Diseases,* 15 (8), pp. 1155-1163.

Marco, M. L., Pavan, S. and Kleerebezem, M. (2006) 'Towards understanding molecular modes of probiotic action', *Current Opinion in Biotechnology,* 17 (2), pp. 204-210.

Morton, T. M., Eaton, D. M., Johnston, J. L. and Archer, G. L. (1993) 'DNA sequence and units of transcription of the conjugative transfer gene complex (trs) of *Staphylococcus aureus* plasmid pGO1', *Journal of Bacteriology,* 175 (14), pp. 4436-4447.

Oelschlaeger, T. A. (2010) 'Mechanisms of probiotic actions – A review', *International Journal of Medical Microbiology,* 300 (1), pp. 57-62.

Ouwehand, A., Salminen, S. and Isolauri, E. (2002) 'Probiotics: an overview of beneficial effects', *Antonie Van Leeuwenhoek,* 82 (1-4), pp. 279-289.

Overbeek, R., Olson, R., Pusch, G. D., Olsen, G. J., Davis, J. J., Disz, T., Edwards, R. A., Gerdes, S., Parrello, B., Shukla, M., Vonstein, V., Wattam, A. R., Xia, F. and Stevens, R. (2014) 'The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST)', *Nucleic Acids Research,* 42:D206-D214. doi:10.1093/nar/gkt1226

199 Pietrocola, G., Valtulina, V., Rindi, S., Jost, B. H. and Speziale, P. (2007) 'Functional and
200 structural properties of CbpA, a collagen-binding protein from *Arcanobacterium pyogenes*',
201 *Microbiology*, 153 (10), pp. 3380-3389.
202 Ponnuraj, K., Bowden, M. G., Davis, S., Gurusiddappa, S., Moore, D., Choe, D., Xu, Y.,
203 Hook, M. and Narayana, S. V. L. (2003) 'A "dock, lock, and latch" Structural Model for a
204 *Staphylococcal* Adhesin Binding to Fibrinogen', *Cell*, 115 (2), pp. 217-228.
205 Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., Rajandream, M. and Barrell, B.
206 (2000) 'Artemis: sequence visualization and annotation', *Bioinformatics*, 16 (10), pp. 944-
207 945.
208 Scheurwater, E., Reid, C. W. and Clarke, A. J. (2008) 'Lytic transglycosylases: Bacterial
209 space-making autolysins', *The International Journal of Biochemistry & Cell Biology*, 40 (4),
210 pp. 586-591.
211 Tomaro-Duchesneau, C., Jones, M. L., Shah, D., Jain, P., Saha, S. and Prakash, S (2014).
212 'Cholesterol Assimilation by *Lactobacillus* Probiotic Bacteria: An In Vitro Investigation',
213 '*BioMed research international*'. 2014 (2014), Article ID 380316, 9
214 Yamaguchi, M., Terao, Y. and Kawabata, S. (2013) 'Pleiotropic virulence factor –
215 *Streptococcus pyogenes* fibronectin-binding proteins', *Cellular Microbiology*, 15 (4), pp. 503-
216 511.
217 Yang, G., Liu, Z. and Yang, P. (2013) 'Treatment of allergic rhinitis with probiotics: An
218 alternative approach', '*North American Journal Of Medical Sciences*'. 5(8), pp. 465-468
219 Zielenkiewicz, U., Kowalewska, M., Kaczor, C. and Ceglowski, P. (2009) 'In Vivo
220 Interactions between Toxin-Antitoxin Proteins Epsilon and Zeta of *Streptococcal* Plasmid
221 pSM19035 in *Saccharomyces cerevisiae*', *Journal of Bacteriology*, 191 (11), pp. 3677-3684.
222

222 **Table 1.** Putative functions of the genes carried by plasmid pLF3872. The gene products
223 were analysed using NCBI Blast server. The hits with highest similarity scores were selected
224 with the respective reference strains indicated.

Loci	Gene	Gene product	Putative function	Reference strain	Identity, % (coverage, %)
1-1335	*ykgC*	Pyridine nucleotide-disulphide oxidoreductase YkgC	Energy production (oxidoreductase activity)	*Lactobacillus plantarum* CMPG5300	100 (99)
1663-2247	*tnpR*	Putative resolvase	Recombination	*Lactobacillus salivarius* CECT 5713	99 (99)
2571-3197	*hyp4*	Conserved hypothetical protein		*Lactobacillus salivarius*	99 (99)
3409-3693	*dinJ*	DNA damage inducible protein J	Antitoxin	*Lactobacillus brevis* ATCC 367	93 (98)
4083-4535	*hyp5*	Hypothetical protein		Multiple *Lactobacillus* bacteria	99 (99)
5007-5798	*parA*	Plasmid partitioning protein ParA	Cell division, partitioning (replication)	Multiple *Lactobacillus* bacteria	100 (99)
5840-6109	*hyp6*	Hypothetical protein		*Lactobacillus hilgardii*	96 (98)
6670-7833	*repA*	Replication initiator protein RepA	Replication	*Lactobacillus crispatus* EM-LC1	100 (99)
7808-8104	*hyp7*	Hypothetical protein		*Lactobacillus crispatus* EM-LC1	100 (98)
8388-8660	*eatL*	Epsilon anti-toxin	Post-segregation killing system	Multiple *Lactobacillus* bacteria	97 (98)
8663-9493	*zetL*	Zeta toxin	Post-segregation killing system	*Lactobacillus antri*	99 (98)
9577-9855	*hyp10*	Hypothetical protein		*Lactobacillus oris* PB013-T2-3	99 (98)
9878-10120	*hyp11*	Hypothetical protein		*Lactobacillus casei* Zhang	100 (92)
10346-12421	*traA*	Nickase	Conjugation	*Lactobacillus rhamnosus*	99 (99)
12506-12817	*hyp1*	Hypothetical protein		*Lactobacillus plantarum* CMPG5300	99 (99)
12853-13467	*hyp9*	Hypothetical protein		*Lactobacillus plantarum* CMPG5300	99 (99)
13469-13804	*traB*	Transfer complex protein TraB	Conjugation	*Lactobacillus paracasei*	99 (99)
13825-14187	*trsC*	TrsC	Conjugation	*Lactobacillus paracasei*	99 (99)
14156-14815	*trsD*	TrsD	Conjugation	*Lactobacillus plantarum* CMPG5300	99 (99)
14827-16845	*trsE*	TrsE	Conjugation	*Lactobacillus paracasei*	99 (99)
16838-18256	*trsF*	TrsF	Conjugation	*Lactobacillus plantarum* CMPG5300	97 (99)

18257-19414	*tcpG*	Peptidoglycan hydrolase	Hydrolysis of peptidoglycan	*Lactobacillus casei* Zhang	99 (99)
19428-20045	*hyp12*	Hypothetical protein		*Lactobacillus oris* PB013-T2-3	99 (99)
20032-20400	*Trx*	Thioredoxin	Reduction of oxidising compounds	*Lactobacillus oris* PB013-T2-3	94 (99)
20401-20871	*trsJ*	TrsJ	Conjugation	*Lactobacillus helveticus* CIRM-BIA 101	96 (99)
21101-22651	*traG*	Conjugal transfer protein TraG	Conjugation	*Lactobacillus oris* PB013-T2-3	98 (99)
22651-23055	*hyp2*	Hypothetical protein		*Lactobacillus coryniformis* CECT 5711	99 (99)
23074-23913	*trsL*	TrsL	Conjugation	*Lactobacillus paracasei* Lpp189	100 (99)
23928-24338	*hyp3*	Hypothetical protein		Multiple *Lactobacillus* bacteria	100 (99)
24345-26480	*topB*	DNA topoisomerase III	Replication	*Lactobacillus paracasei*	98 (99)
26602-26817	*hyp8*	Hypothetical protein		*Lactobacillus oris* PB013-T2-3	99 (98)
26821-27945	*ltrC*	Low temperature requirement C protien	Phosphatidylglycerophosphatase activity	*Lactobacillus coryniformis*	97 (99)
28656-31823	*cbp*	Collagen binding protein	Adhesion	*Lactobacillus casei, Lactobacillus oris*	94 (99)

225

226

Figure legends

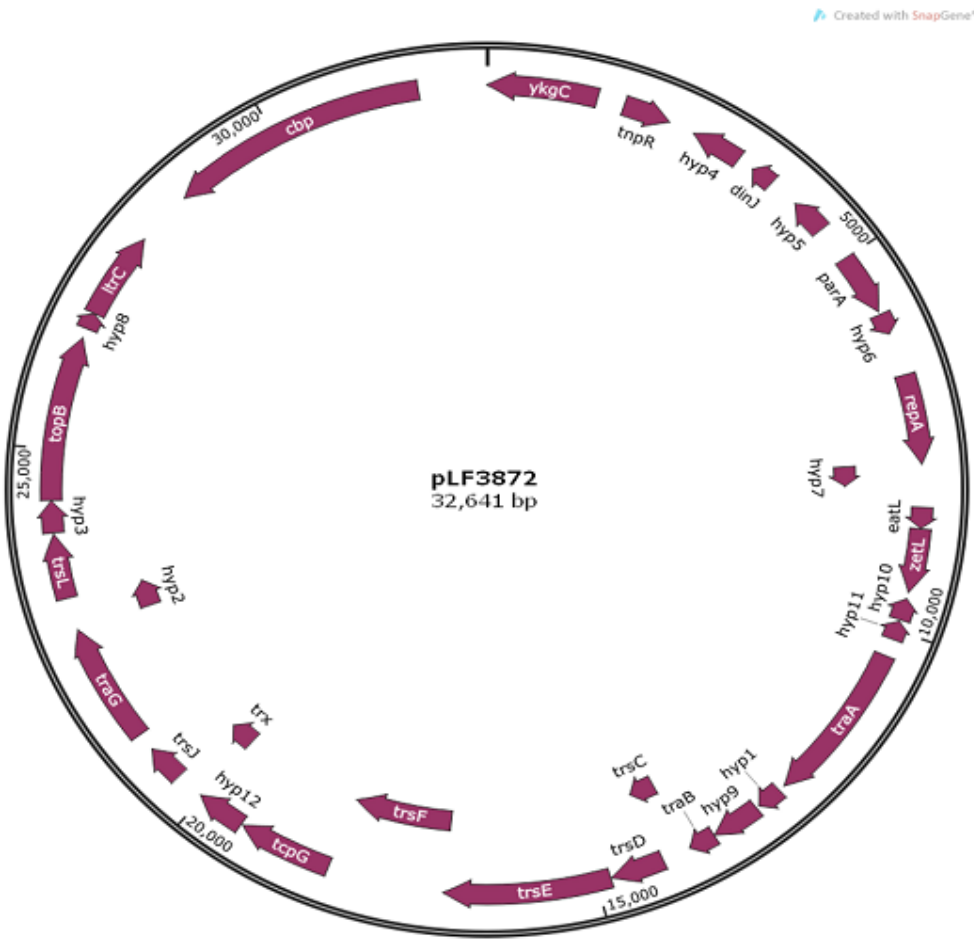
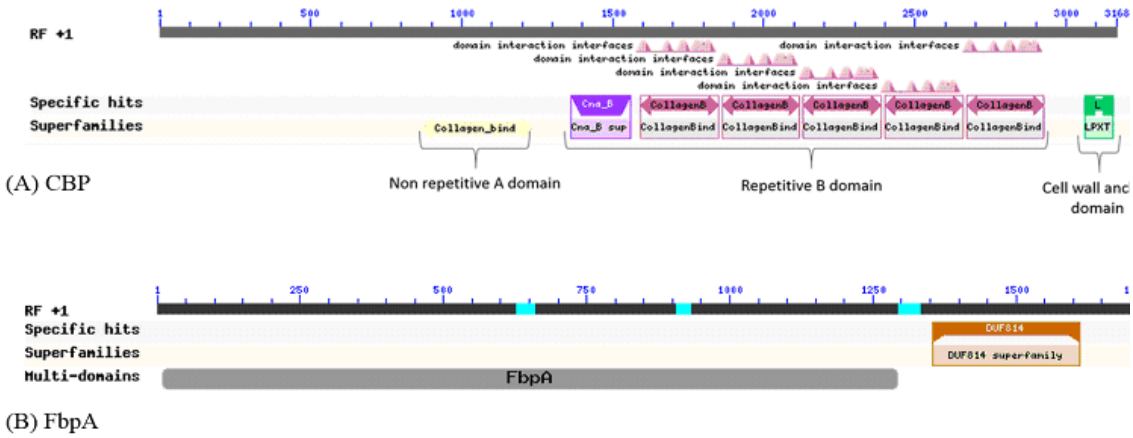


Figure 1. Plasmid pLF3872 genetic map generated by SnapGene program.



231

232

233 **Figure 2.** Domains in the fibronectin binding protein (FbpA) and collagen binding protein

234 (CBP) of *L. fermentum* 3872 detected using NCBI conserved domain (CD) program

235 (translated nucleotide sequence against a CD database).

236

237 **Figure 1S.** Comparison of plasmids plca36 (top) and pLF3872 (bottom) using WebACT

238 program (Carver *et al*., 2005). The red lines connect regions of high level of similarity.

239 *Note: the pLF3872 gene names shown are produced by SnapGene programme. The gene

240 names of plca36 have been abbreviated, with those labelled '*hyp*' referring to hypothetical

241 genes.

242 **Figure 2S.** Comparison of DinJ proteins encoded by pLF3872 (A) and plca36 (B) plasmids.

243 The RelB superfamily domain and the amino acid sequence similarities to a reference

244 sequence pfam042221 (as generated by NCBI Blast server) are shown.

245

246