Comparative Genomic Analysis of *Lactobacillus mucosae* LM1 Identifies Potential Niche-Specific Genes and Pathways for Gastrointestinal Adaptation

Valerie Diane V. Valeriano[1†], Ju Kyoung Oh[1], Bernadette B. Bagon[1], Heebal Kim[2], and Dae-Kyung Kang[1]*

[1]Department of Animal Resources Science, Dankook University, Cheonan 31116, Republic of Korea

[2]Department of Agricultural Biotechnology, Seoul National University, Seoul 08826, Republic of Korea

*Corresponding author**:** Dae-Kyung Kang

Department of Animal Resources Science, Dankook University, 119 Dandae-ro, Cheonan 31116, Republic of Korea

E-mail: dkkang@dankook.ac.kr

[†]Current address: MIMS-Department of Molecular Biology, Umeå University, SE-901 87 Umeå, Sweden

Short title: Gut Mucosa-adherent *Lactobacillus mucosae* LM1

Abstract

Lactobacillus mucosae is currently of interest as putative probiotics due to their metabolic capabilities and ability to colonize host mucosal niches. *L. mucosae* LM1 has been studied in its functions in cell adhesion and pathogen inhibition, etc. It demonstrated unique abilities to use energy from carbohydrate and non-carbohydrate sources. Due to these functions, we report the first complete genome sequence of an *L. mucosae* strain, *L. mucosae* LM1. Analysis of the pan-genome in comparison with closely-related *Lactobacillus* species identified a complete glycogen metabolism pathway, as well as folate biosynthesis, complementing previous proteomic data on the LM1 strain. It also revealed common and unique niche-adaptation genes among the various *L. mucosae* strains. The aim of this study was to derive genomic information that would reveal the probable mechanisms underlying the probiotic effect of *L. mucosae* LM1, and provide a better understanding of the nature of *L. mucosae* sp.

Keywords

Lactobacillus mucosae, Comparative genomics, Probiotics

1. Introduction

Lactobacillus spp. are members of the family *Lactobacillaceae* and are generally low-GC-content, non-motile, non-spore-forming, anaerobic bacteria. The members of this genus are subgrouped by their carbohydrate fermentation profiles, namely, obligately homofermentative, facultatively heterofermentative, and obligately heterofermentative. The genus *Lactobacillus* is a member of the lactic acid bacteria (LAB) group that plays a key role in the gastrointestinal tract [1]. It belongs to *Firmicutes*, which is a major mucosa-associated phylum of the gut microbiota, especially in the pars nonglandularis [2]. The genus *Lactobacillus* reportedly exerts beneficial effects on human and animal health [1,3,4].

Lactobacillus mucosae was first identified by Roos *et al*. [5] using the mucus-binding (MUB) gene from *Lactobacillus reuteri* as a probe. Initial research on *L. mucosae* revealed its ability to adhere to pig mucus [5]. In humans, *L. mucosae* represents one of the strongly mucosa-associated subpopulations of the intestinal microbiota, together with *Lactobacillus gasseri* and *Bifidobacterium breve* [6]. Its tight association with the mucosal layer and its presence in other mucosal niches, such as the human cervix and vagina, demonstrate the affinity of *L. mucosae* for mucin components and the host mucosal layer [7].

L. mucosae LM1 has various probiotic traits; for example, acid and bile tolerance, which facilitates survival during transit through the gastrointestinal tract [8,9], production of antibacterial substances against various Gram-positive and -negative pathogens [8], adhesion to intestinal epithelial cells [10], and inhibition of enteropathogens *in vitro* [8,10]. Its ability to aggregate, adhere to mucus, and inhibit adhesion of pathogens are prerequisites for competitiveness in, and transient colonization of, the host gut [10].

A draft genome sequence of LM1 [11] revealed the presence of numerous adhesion genes. To characterize its probiotic properties, the complete genome of LM1 was sequenced, which is a first for an *L. mucosae* strain, and compared with those of other draft genomes of *L.*

mucosae strains, and of the completed genomes of the closely associated species *L. reuteri* and *Lactobacillus fermentum*. The data will enable assessment of the biotechnological and probiotic potential of this *L. mucosae* strain and, likewise, provide a foundation for understanding *L. mucosae* species for further applications.

2. Results and Discussion

2.1. Genome Features

After resequencing, reannotation, and reassembly, the *L. mucosae* LM1 complete genome was 2.43 Mbp, with a 2.33 Mbp circular chromosome (46.13% GC ratio) (**NZ_CP011013.1**) and a 0.1 Mbp extrachromosomal plasmid, pLM1 (**NZ_CP011014.1**). A total of 2,161 open reading frame (ORF) genes were predicted from the circular chromosome, with 1,998 protein-coding genes, 24 rRNA genes, 90 tRNA genes, and 49 pseudogenes. From the plasmid pLM1, 114 ORFs were identified, comprising 111 protein-coding genes, 1 tRNA gene, and 2 pseudogenes (Table S1). A total of 1,623 proteins (70.93%) were assigned functions based on the functional COG categories of the whole genome. A previous study of the draft genome sequence of LM1 failed to identify the plasmid incorporated in the LM1 genome. Annotation of plasmid pLM1 showed that COG functions involving replication, recombination, and DNA repair were the most abundant, and phage replication components (DNA polymerase III and DNA helicase) were also identified. Further study is required to understand the functional mechanisms of this plasmid.

The COG functions of the genes are presented in Table S2. Genes involved in replication, recombination, and repair (8.49%) were most abundant, followed by amino acid transport and metabolism (7.73%), translation, ribosomal structure, and biogenesis (6.35%), transcription (5.60%), carbohydrate transport and metabolism (5.55%), and cell wall/membrane biogenesis

(5.03%). The functional enrichment of such ORFs in the genome suggests that *L. mucosae* LM1 is able to adapt to harsh environments, such as the human gastrointestinal tract [12-15].

Therefore, we further analyzed three factors that may contribute to probiotic activity: (1) genes involved in stress resistance and maintenance of cell-envelope integrity, (2) genes that mediate signaling and regulatory mechanisms, and (3) adherence and cell-surface factors that mediate colonization and bacterium–host crosstalk [16]. A list of potential probiotic genes present in strain LM1 presented previously [8,11] and updated in this study are also included as supplementary Table S3. Such traits are important for the ability of probiotic microorganisms to compete with harmful bacteria, modulate the host immune system, and regulate the gut microbiota [4].

2.2. L. mucosae LM1 plasmid (pLM1)

The plasmid replicon pLM1 (Figure 1C) has a predicted incomplete 18.8 kb prophage region (at positions 20,613–39,427), which contains phage replication elements such as a recombinase and a phage tail fiber protein (Figure S2). It is found within the *oriC* region (*rnpA-rmpH-dnaA-dnaN-recF-gyrB-rnpA* gene cluster) [17], adjacent to *dnaA* (LBLM1_RS00005). The conserved *dnaA* box clusters, *dnaN* (LBLM1_RS10815) and *recF-gyrB* (LBLM1_RS10805-LBLM1_RS10800), flank the predicted pLM1 sequence, which suggests incorporation of pLM1 within the putative chromosomal *oriC* site. This plasmid also contains prophage elements, suggesting that it is a phage-plasmid. Plasmids that also function as phages have been reported in other bacteria including *Streptomyces* [18], *E. coli* [19,20]. However, whether this is the case in strain LM1 will require further investigation.

Whole sequence BLAST analysis of this plasmid region showed no significant similarity with other plasmids; however, specific sections aligned at an e-value of 4e-92 to chromosomal regions of *L. amylolyticus* L6 complete genome (76% identity), at 2e-15 to chromosomal

regions in the complete genomes of *L. fermentum* strains and *L. oris* J-1 (92-94%), and, likewise, at e-values of 3e-14 to *Leuconostoc mesenteroides* plasmids (78-79%). These regions are annotated as either hypothetical proteins or phage integrases. Annotated ORFs also include mostly genes involved in replication, recombination and repair, such as deoxynucleoside kinases which were deemed important deoxynucleoside salvage pathways for DNA precursor biosynthesis in lactobacilli [21].

2.3. Genomic Islands

The identified genomic islands (GIs) and their genes are summarized in Table S3, with 68 predicted IS elements (File S1). The highest similarity of IS elements was with IS30 (12 hits), which is present in various *Lactobacillus* spp., including *Lactobacillus plantarum* [22]. Furthermore, the presence of an IS30-related element in *L. plantarum* is associated with genome plasticity and environmental adaptation, due to acquisition of putative genes such as those involved in sugar utilization [23], which likely assists gut colonization.

A total of 12 GIs, which are of probable horizontal transfer origin, were identified in the LM1 genome (Figure 1B). Using PHAST, three prophage elements were identified in the chromosome of LM1 (Figure 1B and Figure S1), where GI region 3 (GI-3) incorporates prophage 1 (P1). GI-4 coincides with prophage 2 (P2), and GI-9 and -10 with prophage 3 (P3).

The only intact prophage region (region 2) resembled Bacill_BalMu_1 (NC_030945), which is 32.02 kilobases (kb) in size. Non-phage elements in this region included ORFs involved in putative functions such as copper and potassium homeostasis, lipoprotein biosynthesis, and a phosphate transcriptional regulon may play a role in osmotic adaptation and stress signaling [24,25]. Interestingly, among the *L. mucosae* strains, this prophage region was found only in LM1. Additionally, the prophages detected seem to create some genomic

variations among the *L. mucosae* strains. However, non-phage elements in these regions are hypothetical proteins and are, thus, unknown.

Two exopolysaccharide (EPS) operons were identified within GI-2 and GI-12. GI-2 EPS is complemented by two glycosyl transferases, which are involved in the production of glycosaminoglycan (GAG) and chondroitin sulfate. *L. mucosae* DPC 6426 produces EPS with mannosyl residues [26]. However, whether LM1 produces EPS or not has yet to be determined.

2.4. Self-defense and Immune Systems

Restriction-Modification (RM) systems and Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) associated with CRISPR-associated genes, are some mechanisms by which bacteria resist phage infections [27]. Both are compatible, wherein the CRISPR-Cas system also cleaves invading DNA that were methylated by RM systems [28]. In *L. mucosae* LM1, only a few genes associated to type I and type III restriction endonuclease systems were found; whereas, two CRISPR-Cas loci (type I and type II), were detected in regions 1,305,291-1,315,478 bp, with 29 direct repeats (DRs) and 13 spacers, and 1511616-1514090 bp, with 36 DRs and 37 spacers. The information processing subsystem Cas1 (LBLM1_RS06065) and Cas2 (LBLM1_RS06060) genes are present in the first region along with the cascade complex component Cas3 (LBLM1_RS06095) gene. On the other hand, a putative Cas9 gene (LBLM1_RS07010) is found within the second region.

2.5. Phylogenetic Analysis

The carbohydrate metabolism of LM1 is similar to that of the type strain *L. mucosae* S32^{T} (DSM 13345), in terms of utilization of carbohydrates such as glucose, ribose, D-xylose, galactose, aesculin, melibiose, D-raffinose, and gluconate [5,8]. *L. mucosae* strains belong to the obligately heterofermentative group, together with one of the major *Lactobacillus*

phylogenetic groups, the *L. reuteri* group, based on their 16s rRNA gene sequences (Figure S3). The 16S rRNA sequences of the *L. mucosae* strains clustered with *L. fermentum* IFO 3956, *Lactobacillus pontis* LTH 2587, and *L. reuteri* JCM 1112, in agreement with a previous report [5], but in a different branch from these species, as confirmed by ANI analysis. A discrepancy with closely related species such as *L. fermentum* and *L. reuteri* strains was detected (Figure 2), as evidenced by a < 72% ANI value by 16s rRNA gene sequencing. All *L. mucosae* strains studied here were phylogenetically similar by 16s rRNA gene sequencing. Further, a species cut-off of 95% [29] was confirmed for all *L. mucosae* strains, with the rumen isolate strains KHPC15, KHPX11, and WCC8 showing lower ANI values than those of the type strain *L. mucosae* S32^T and LM1.

According to the ANI values, the complete genome of LM1 was most similar (97.72%) to that of the *L. mucosae* type strain S32^T (DSM 13345), which was also isolated from swine intestine [5], followed by DPC 6426 from the bovine gastrointestinal tract (97.37%) [30], Marseille from a patient with Kwashiorkor (97.32%) [31], and the mannitol-producing strain CRL 573 (96.00%), which was isolated from child feces [32]. Such dissimilarities may be due to host-specific strain diversity and niche adaptation. Next, LM1 was subjected to a comparative genomic analysis with the other *L. mucosae* strains as well as *L. reuteri* and *L. fermentum*.

2.6. Comparative genomic analysis of L. mucosae *LM1*

2.6.1. Comparison of LM1 with other L. mucosae *strains*

Initially, the sequencing quality and completeness of the draft genomes may have contributed to the differences in contig assembly among the *L. mucosae* strains. However, after re-aligning and re-assembling the contigs using MAUVE with the complete genome of LM1, comparison among the strains was possible as demonstrated by the genome-scale dotplots

(Figure S4), with internal conserved blocks and homologous genes considered for comparison among the *L. mucosae* strains.

Analysis of genome-scale dotplots using the complete LM1 genome as a reference (Figure S4) indicated general synteny among the *L. mucosae* strains. However, the *oriC* positions varied among these strains, as shown in the whole genome MAUVE alignment (Figure 3); specifically, only LM1 had a plasmid (pLM1) prophage within the putative *oriC* site. Numerous chromosomal insertions, arrangements, and inversions were also detected by MAUVE alignment (Figure 3). Interestingly, even the *L. mucosae* strains KHPC15 and KHPX11, which originated from the same host (ruminant) and had perfect synteny (ANI value, 100%), showed a strain-specific chromosomal topology; an ~ 81 kb chromosomal inversion was present at positions 724,730–809,682 in KHPC15 and 698,767–783,145 in KHPX11 (Figure S5). The variations were found to be due to the presence of numerous IS elements; indeed, prophage elements were identified in the genome of LM1 (see section 2.3).

Putative integrases related to *Staphylococcus* pathogenicity islands (SaPIs), such as LBLM1_RS02825, LBLM1_RS03880, LBLM1_RS05565, and LBLM1_RS07345, were found in the LM1 genome with specific phage elements. SaPIs reportedly have an intimate relationship with temperate phages, allowing these genomic islands to be encapsidated into small phage heads [33]. Of note, the integrase ORF LBLM1_RS02825 was part of P1, which was predicted by MAUVE alignment to have a conserved region that was likely translocated among the *L. mucosae* strains, with the exception of Marseille and DPC 6426. Further, a chromosomal inversion of the conserved block was present in KHPX11 (Figure 3 and Figure S5). In Figure 3, color blocks 17 to 21 in KHPX11 and KHPC15 show the presence of P1 (block 18) between blocks 17 (blue) and 19, which are adjacent to each other in the other *L. mucosae* strains. The presence of transposases, including LBLM1_RS03940, LBLM1_RS03945, and LBLM1_RS04210 (putative transposases in the *snaA-snaB* intergenic

region), may have a considerable impact on the functionality and expression of the genes involved with these transposable elements in LM1. Interestingly, genes in these blocks are essential for colonization of, and survival in, the gastrointestinal tract. These include a major facilitator superfamily (MFS) transporter, which is implicated in bile tolerance [34]; enzymes involved in the maltose and glycogen metabolic pathways [35] and in the mevalonate pathway [36], which may contribute to sugar utilization adaptation; and cell-surface protein-coding genes, such as choline-binding proteins and penicillin-binding proteins, which are implicated in adhesion [37] and maintaining peptidoglycan under bile stress [38], respectively. Genes in the inverted genome blocks of LM1 are summarized in Table S4.

A rearrangement of a large genomic region at positions 971,200–1,116,880 and 576,650–613,070 in LM1 was detected by comparison with those of S32^T (485,600–628,245 and 980,305–1,019,760 of S32^T, respectively) (Figure 3). Putative genes in this region also include P1 in the latter region, and insertional elements annotated as IS905 transposases with sequence homology to IS256 (LBLM1_RS02185 and LBLM1_RS02260), and this set of transposed regions is also present in S32^T. In the LM1 genome, internal transposases and transposons are also present in the former region, which may increase the plasticity of these conserved genomic blocks. Interestingly, a MUB (LBLM1_RS04610) is present in this region, and may be transported through a proximally-located Sec translocation system (LBLM1_RS04640-LBLM1_RS04660). Further, production of metabolic energy by malolactic fermentation may be mediated by the malate/lactate antiporter (LBLM1_RS04685) in this region, similar to *L. plantarum* [39]. Other orthologous genes involved in tolerance to acid, bile, and oxidative stress were also found in this region. The conserved orthologous genes in *L. mucosae* LM1 also present in the rearranged genome blocks of KHPX11 and KHPC15 are summarized in Table S5.

A prior tandem proteogenomic analysis and reconstruction of the metabolic pathways of LM1 [15] provided insight into the mechanisms by which *L. mucosae* tolerates environmental stress, maintains cellular integrity, and adheres to mucosal components [14,40]. Acidic and oxidative stress result in activation of cell membrane biosynthesis and repair of the peptidoglycan layer. Moreover, upregulation of carbohydrate metabolic enzymes such as glycosyl hydrolase family 65 protein (*GSH65*; LBLM1_RS04255), xylulose-5-phosphate/fructose-6-phosphate phosphoketolase (*Xfp*; LBLM1_RS0385), and 6-phosphogluconate dehydrogenase (*6PGDH*; LBLM1_RS00305), enables use of various carbohydrates, such as maltose and trehalose. GSH65 is adjacent to P1 in block 17, which is involved in maltose and glycogen metabolism.

The absence of ATP-dependent 6-phosphofructokinase (*PfkA*) and FAD-dependent pyridine nucleotide-disulfide oxidoreductase (*DhnA*) from the *L. mucosae* genomes suggests that Xfp activity may be important for redirection of energy production via the phosphoketolase pathway [41]. Use of other carbon sources was suggested by the presence of other enzymes, such as β-glucosidase (LBLM1_17760), glucan 1,3-β-glucosidase (LBLM1_17810), and pectinesterase (LBLM1_04470). Upregulation of CitT (LBLM1_00930) facilitates citrate uptake, which generates malate and succinate, important byproducts that can be used as an energy source by host intestinal cells [15]. These findings provide evidence for the mechanisms that allow LM1 to survive in carbohydrate-depleted environments.

2.6.2. Pan-genomic analysis of LM1 with other L. mucosae *strains and closely related* Lactobacillus *species*

2.6.2.1. Glycogen metabolism

To confer beneficial effects on the host, the viability and prolonged retention of probiotics in the gut are important. Therefore, genes related to stress tolerance and starvation, colonization,

and gut persistence were analyzed. Interestingly, a complete glycogen metabolism (*glg*) operon was present in the LM1 genome (Figure 4). The glycogen biosynthetic pathway is involved in energy production [42], environmental stress tolerance [43,44], biofilm formation [45], and colonization and environmental persistence [35,46-48]. Glycogen is used by bacteria to store carbohydrates and enable colonization of diverse niches [49,50]. Although the complete *glg* operon is also present in some *Lactobacillus* spp. from mammalian and natural environments [51,52], it was incomplete in other *L. mucosae* strains, and absent from *L. reuteri* JCM 1112 and *L. fermentum* IFO 3956 (Figure 3).

Furthermore, activation of the *glg* operon during early stationary phase growth of *L. mucosae* LM1 [15] suggests involvement in mucosa-associated, niche-specific adaptation, which provides ecological and competitive advantages. The reason for the absence of this operon in some isolates from mammalian hosts is unclear, but may be due to their preferences for other carbohydrates for carbon storage.

2.6.2.2. Folate biosynthesis in L. mucosae *species*

Comparison of the genomes of *L. mucosae* LM1 and *L. fermentum* IFO 3956 revealed a folate producer, indicating that LM1 can synthesize folate, which is an essential cofactor in all living cells [53]. LM1 exhibits a complete folate biosynthetic pathway (Figure 5) and uses para-aminobenzoic acid, an intermediate of folate biosynthesis, via the shikimate pathway [54]. This gene set was also found in *L. reuteri* JCM 1112, but this strain lacks dihydrofolate synthase/folylpolyglutamate synthase (*folC*) and dihydrofolate reductase (*dfrA*), which are essential for producing tetrahydrofolate. Folate produced in the intestine by lactic acid bacteria is used by the colonic epithelium, and prevents inflammation and colonic cancer [55]. *Lactobacillus delbrueckii* subsp. *bulgaricus*, another folate producer, is added to dairy products

to increase their folate content [56]. The conditions under which LM1 produces folate should be the focus of further studies.

2.6.3. Neopullulanase and its prebiotic functions

Genome mining of LM1 revealed genes encoding α- and β-galactosidases, β-glucuronidases, β-fructosidases (levansucrases), and α-amylases, which are involved in the production of prebiotic substances. We isolated an α-amylase gene (LBLM1_RS010140) and found that it encoded a neopullulanase. Overexpression of this gene in *E. coli* showed that the neopullulanase can degrade pullulan to generate the prebiotic panose, an isomaltooligosaccharide used by beneficial intestinal bacteria [57]. This was the first neopullulanase cloned from lactic acid bacteria, with others cloned and characteracterised from in *Bacillus* sp. [58,59] and *Bacteroides thetaiotaomicron* 95-1 [60]. Interestingly, BLAST analysis of the LM1 neopullulanase gene aligns with high identity and similarity to the other *L. mucosae* alpha-glycosidases (>97%), but only around 78% query coverage to the neopullulanase found in rumen isolates, but still at a high identity. Significant hits with alpha-glycosidases and neopullulanases in other *Lactobacillus* sp. shows <69% identity, and lower coverage, showing some diversity of this enzyme among LAB. Nevertheless, the ability of LM1 to produce panose, and its presence in many *L. mucosae* strains suggests the utility of *L. mucosae* strains as a prebiotic-producing-probiotic, as most possess this gene [61].

2.6.4. Cell surface adhesins and other adhesion factors that may mediate host-microbe cross talk

Bacterial adhesion to specific glycoprotein receptors in mucus is important because mucus functions as a natural barrier to protect gut epithelial cells. Lee *et al*. [11] identified a MUB homologue in the LM1 genome, and LM1 reportedly has an affinity for glycoreceptors on

porcine intestinal epithelial cells (IPEC-J2) [10]. Although most *L. mucosae* genomes contain a MUB homologue [5,11,31], *L. mucosae* CRL 573 and DPC 6426 possess a truncated form [30,32], which has high homology with a mucin-binding protein (MucBP) domain. Other adhesins may facilitate adhesion to gut epithelial cells; *e.g.*, putative collagen-binding proteins and fibronectin-binding protein in DPC 6426 [30].

BLASTp analysis of the complete genome of LM1 also identified MucBP proteins (LBLM1_RS02150 and LBLM1_RS04140); these are glycosylated streptococcal protein B (GspB) homologues with affinity for sialic acid residues in mucins [62]. The GspB platelet-binding protein from *Streptococcus gordonii* is glycosylated in the serine-rich SSR1 and SSR2 regions with glucose, N-acetylglucosamine, and to a lesser extent galactose and N-acetylgalactosamine [63].

MucBP domain searching revealed the presence of two putative internalin-like proteins (LBLM1_RS07130 and LBLM1_RS06500), which have high homology with mucus-binding domains, suggesting some functional similarities. These internalin-like proteins have an LPXTG motif, and their amino acid sequence is similar to that of the leucine-rich repeat-containing proteins of *Listeria monocytogenes* and the MucBPs of other *Lactobacillus* spp. Von Ossowski *et al*. [64] identified an ORF (LGG_02337), which was previously annotated as Internalin J, in the commercial probiotic *L. rhamnosus* GG. The purified form of this surface protein readily adhered to human intestinal mucus. Further, this protein has an L-type lectin domain, suggesting that it mediates adhesion to glycoreceptors on the surface of host cells.

A 2.2 kb operon that showed similarity and synteny with the mucus adhesion-promoting protein (*mapA*) operon (AJ293860.1) from *L. reuteri* [65] is present in the draft genome sequence of LM1 [8]. MapA is an adhesion protein that is degraded into an antimicrobial

peptide [65]. The ORFs in this operon, which correspond to the *MapA* region, include LBLM1_RS05045 to LBLM1_RS05055.

Lam29, an adhesin-like protein specific to blood group A and B antigens, was originally identified in *L. mucosae* ME-340 [66,67]; homologues of this protein are present in LM1 (LBLM1_RS02430) and other *L. mucosae* strains. These should be the subject of further studies on the interaction of *L. mucosae* strains with the host.

2.6.5. Core genome of L. mucosae *strains and closely related* Lactobacillus *species*

Large areas in the heat map in Figure 6 delineate the orthologous and paralogous differences among the genomes of the species and strains analyzed in this study. The presence of horizontally transferred genomic islands and prophages, such as P1/GI-3 and P3/GI-9 to GI-11, are major differences among the strains and species. However, a core genome functionally rich in various metabolic pathways, such as proteins, amino acids and derivatives, carbohydrates, cofactors, vitamins, prosthetic groups, and pigments (Figure S6) provides a competitive advantage to *L. mucosae* and likely other gut bacteria by enabling production of diverse metabolites. In addition, the common pathways in *L. mucosae* spp., *L. reuteri*, and *L. fermentum* suggest the potential use of *L. mucosae* as a probiotic and as a component of synbiotics.

In conclusion, our comparative analysis of the complete genome of *L. mucosae* LM1, which is the first to be performed on *L. mucosae*, with other *L. mucosae* strains and closely related species enhances our understanding of its role in the gut and the mechanisms by which it colonizes and persists therein. Identification of the folate and glycogen biosynthetic pathways and alternative energy routes suggested the means by which *L. mucosae* survives in the harsh environment of the gut. It is interesting to observe that in comparison to the closest complete genome representative strains of *L. fermentum* sp. and *L. reuteri* sp., *L. fermentum* utilizes

specific pathways present in *L. fermentum* that *L. reuteri* lacks or is incomplete in. In this respect, along with the ANI values, *L. mucosae* seems to be more genotypically closer and functionally similar to *L. fermentum* compared to *L. reuteri*, despite having common essential genes such as the MUB gene. These pathways are present in analyzed *L. mucosae* strains and are, therefore, deemed common among them. Furthermore, genes encoding adhesion and colonization factors likely enhance gut persistence and competitiveness, which also seem common among the *L. mucosae* strains. These factors are associated with the beneficial effect of LM1, which requires establishment in, and symbiosis with, the host.

3. Materials and Methods

3.1. Growth conditions and genomic DNA preparation

L. mucosae LM1 was grown in 5 mL de Man–Rogosa–Sharpe (MRS) broth (Difco, Pont-de-Claix, France) at 37°C under static conditions for 24 h. Genomic DNA was isolated using the Qiagen Genomic DNA Extraction Kit (Hilden, Germany) according to the protocol provided by the manufacturer. DNA quality was verified by agarose gel electrophoresis, and DNA concentration and purity (A_{260}/A_{280}) were measured using an Optizen Nano Q (Mecasys, Daejeon, Korea).

3.2. Genome sequencing and assembly

The draft genome sequence of *L. mucosae* LM1 was determined to enable genome mining of probiotic genes (Table S3) [11]. The whole genome shotgun (WGS) project was deposited in the GenBank database under the WGS sequencing project accession number **AHIT00000000.1** in February 2012. It was superseded by the GenBank accession records, **CP011013.1–CP011014.1**, which were used to generate the comparative genomics data

presented here. Complete genome sequencing was performed by ChunLab, Inc. (Seoul, Korea), and a more detailed and updated analysis of the data is presented in this article.

The complete genome of *L. mucosae* LM1 was sequenced on 454 Titan PE 8K, Illumina 100 PE, and PacBio 10K sequencing platforms, and yielded 16.31× (200,748 reads), 1002.38× (22,807,602 reads), and 149.19× (150,292 reads) coverage, respectively, for a total sequencing depth of 1167.89×. Genome assembly was performed using Roche gsAssembler 2.6, CLC bio CLC Genomics Workbench 6.5, and SMRT Analysis 2.0.0, and gap closing was performed using CodonCode Aligner 3.7. The final assemblies resulted in two contigs with a total length of 2.43 megabase pairs (Mbp; 46.13% GC ratio).

3.3. Genome annotation

Coding sequences (CDS) were predicted by Glimmer 3.02 [68]. Transfer RNAs (tRNAs) were identified by tRNA-Scan-SE [69], and ribosomal RNAs (rRNAs) were searched using HMMER with the EzTaxon-e rRNA profiles [70,71]. The predicted protein CDSs were compared to the Catalytic Families (CatFam) and NCBI Clusters of Orthologous Groups (COGs) databases by rpsBLAST and to the NCBI Reference Sequences (RefSeq) and SEED databases by BLASTp for functional annotation [72-75]. Restriction-modification (R-M) system and the CRISPR-Cas loci were predicted from these annotations. Following, CRISPR direct repeats (DRs) and spacers were identified using CRISPRFinder (http://crispr.i2bc.paris-saclay.fr/) [76-78].

Prophage regions were predicted, and figures were generated, using PHAST (http://phast.wishartlab.com) [79]. As submitted to NCBI, each contig, one for the chromosome and one for the plasmid, was subjected to analysis using PHAST to predict the completeness of the prophage regions, and to identify and map their elements. Insertion sequence (IS) elements were predicted using IS Finder software (http://issaga.biotoul.fr/issaga_index.php)

[80]. Genomic islands were predicted using IslandPath and SIGI-HMM, which is part of IslandViewer4 [81], to identify regions with evidence of horizontal transfer.

The subcellular localization of proteins was predicted using SignalP 4.1 (http://www.cbs.dtu.dk/services/SignalP) [82] and TMHMM 2.0 (http://www.cbs.dtu.dk/services/TMHMM) [83]. Lipoproteins were identified among the latter using LipoP 1.0 (http://www.cbs.dtu.dk/services/LipoP) [84].

3.4. Phylogenetic Inference and Average Nucleotide Identity (ANI) Analysis

Using the 16s rRNAs from the complete genomes of representative strains, a phylogenetic tree was constructed by the neighbor-joining method [85], using molecular evolutionary genetics analysis (MEGA) 7.0 software [86]. The tree included 21 representative *Lactobacillus* strains and *Bifidobacterium breve* as an outgroup, in comparison with the eight *L. mucosae* strains with available genomes in the NCBI database.

Average nucleotide identity (ANI) analysis was performed to determine the relationship of *L. mucosae* species with other species as predicted by 16s rRNA analysis, and to evaluate the species cut-off (< 95%) among the *L. mucosae* strains included in this comparative analysis. ANI values were predicted using the ANI calculator available on the EzGenome website, provided by Chunlab, Inc. [87].

3.5. Comparative Genomics

MAUVE software was used to reassemble the contigs of the draft *L. mucosae* genomes available in the NCBI database using the complete genome of *L. mucosae* LM1 as a reference [88]. The draft genomes included the strains CRL573 (**NZ_JROC00000000.1**), DPC 6426 (**NZ_JSWI00000000.1**), Marseille (**NZ_CVQW00000000.1**), and the type strain S32^{T}, also known as *L. mucosae* DSM 13345 (**NZ_AZEQ01000089.1**) [49]. We also included the newly

available draft genomes of *L. mucosae* strains KHPX11 (**NZ_FOSI00000000.1**), KHPC15 (**NZ_FNZY00000000.1**), and WCC8 (**NZ_FNIH00000000.1**), which were submitted as part of the WGS sequencing project of the Department of Energy Joint Genome Institute. After reassembly, whole-genome sequencing alignments were performed using progressiveMauve [89], annotated, and analyzed. Dot plot matrices were generated using these reassembled contigs against the complete genome of LM1 using Gepard software [90]. The MAUVE and Gepard alignments predict chromosomal rearrangement events such as insertion, inversion, and translocation, and combine them with the data from IS Finder and PHAST. Further, genetic comparisons of the completed LM1 genome with other published *L. mucosae* draft genomes were performed by all-*vs.*-all BLAST and HMMER searching of Pfam domains.

Additionally, comparative genomics was performed based on the bioinformatics pipeline and protocol of Chun *et al*. [91]. The complete genomes of the closely related strains *L. fermentum* IFO 3956 (**NC_010610.1**) and *L. reuteri* JCM 1112 (**NC_010609.1**) [92] were compared with that of *L. mucosae* LM1. Two *L. mucosae* draft genomes, strains CRL 573 and DPC 6426 from bovine gut, were included in the comparison of whole genome sequences with *L. fermentum* and *L. reuteri*. ANI values were used to produce a dendrogram using the unweighted pair group method with arithmetic averages (UPGMA). Homologous regions between the compared genomes were determined using the BLASTN software built-in to CLGenomics™, and aligned using a pairwise global alignment algorithm [93]. From this, the core genome was extracted and re-annotated for common functional groups using the RAST server [94].

Competing Interests

The authors declare that there are no competing interests.

Authors' Contributions

VDV performed comparative genomics and drafted the manuscript. DKK designed research and project outline, and finalized the manuscript. JKO, BBB and HK performed genome analysis. All authors have read and approved the final manuscript.

Acknowledgements

This work was supported by the National Research Foundation of Korea Grant funded by the Korean Government (NRF-2016R1D1A1A09918986).

REFERENCES

1. J. Walters. Ecological role of lactobacilli in the gastrointestinal tract: implications for fundamental and biomedical research. Appl Environ Microbiol. 74 (2008) 4985–4996.
2. E. Mann, S. Schmitz-Esser, Q. Zebeli, M. Wagner, M. Ritzmann, B.U. Metzler-Zebeli. Mucosa-associated bacterial microbiome of the gastrointestinal tract of weaned pigs and dynamics linked to dietary calcium-phosphorus. PLoS One 9 (2014) e86950.
3. S. Delgado, A.M.O. Leite, P. Ruas-Madiedo, B. Mayo. Probiotic and technological properties of *Lactobacillus* spp. strains from the human stomach in the search for potential candidates against gastric microbial dysbiosis. Front Microbiol. 5 (2014) 766.
4. V.D. Valeriano, M.P. Balolong, D-K. Kang. Probiotic roles of *Lactobacillus* spp. in swine: insights from gut microbiota. J Appl Microbiol. 122 (2017) 554-567.
5. S. Roos, F. Karner, L. Axelsson, H. Jonsson. *Lactobacillus mucosae* sp. nov., a new species with *in vitro* mucus-binding activity isolated from pig intestine. Int J Syst Evol Microbiol. 50 Pt.1 (2000) 251-258.
6. S. Fakhry, N. Manzo, E. D'Apuzzo, L. Pietrini, I. Sorrentini, E. Ricca, M. De Felice, L. Baccigalupi. Characterization of intestinal bacteria tightly bound to the human ileal epithelium. Res Microbiol. 160 (2009) 817-823.
7. S.I. Pavlova, A.O. Kilic, S.S. Kilic, J.S. So, M.E. Nader-Macias, J.A. Simoes, L. Tao. Genetic diversity of vaginal lactobacilli from women in different countries based on 16S rRNA gene sequences. J Appl Microbiol. 92 (2002) 451-459.

8. V.D. Valeriano, M.M. Parungao-Balolong, D.-K. Kang. *In vitro* evaluation of the mucin-adhesion ability and probiotic potential of *Lactobacillus mucosae* LM1. J Appl Microbiol. 117 (2014) 485-497.
9. P. Fernández de Palencia, M. Fernández, M.L. Mohedano, V. Ladero, C. Quevedo, M.A. Alvarez, P. Lopez. Role of tyramine synthesis by food-borne Enterococcus durans in adaptation to the gastrointestinal tract environment. Appl Environ Microbiol. 77 (2011) 699–702.
10. V.D. Valeriano, B.B. Bagon, M.P. Balolong, D.-K. Kang. Carbohydrate-binding specificities of potential probiotic *Lactobacillus* strains in porcine jejunal (IPEC-J2) cells and porcine mucin. J Microbiol. 54 (2016) 510-519.
11. J.H. Lee JH, V.D. Valeriano, Y.R. Shin, J.P. Chae, G.B. Kim, J.S. Ham, J. Chun, D.-K. Kang. Genome sequence of *Lactobacillus mucosae* LM1, isolated from piglet feces. J Bacteriol. 194 (2012) 4766.
12. G. Jan, P. Leverrier, V. Pichereau, and P. Boyaval. Changes in Protein Synthesis and Morphology during Acid Adaptation of *Propionibacterium freudenreichii. Appl Environ Microbiol. 67 (*2001*): 2029-2036.*
13. K. Koskenniemi, K. Laakso, J. Koponen, M. Kankainen, D. Greco, P. Auvinen, K. Savijoki K., et al. Proteomics and transcriptomics characterization of bile stress response in probiotic *Lactobacillus rhamnosus* GG. Mol Cell Proteomics. 10 (2011): M110.002741.
14. J.Y. Lee, E.A.B. Pajarillo, M.J. Kim, J.P. Chae, D.-K. Kang. Proteomic and transcriptional analysis of *Lactobacillus johnsonii* PF01 during bile salt exposure by iTRAQ shotgun proteomics and quantitative RT-PCR. J Proteome Res. 12 (2013) 432-443.

15. E.A.B Pajarillo, S.H. Kim, J.Y. Lee, V.D. Valeriano, D.-K. Kang. Quantitative proteogenomics and the reconstruction of the metabolic pathway in *Lactobacillus mucosae* LM1. Korean J Food Sci Anim Resour. 35 (2015) 692-702.

16. S. Lebeer, J. Vanderleyden, S.C. De Keersmaecker. Host interactions of probiotic bacterial surface molecules: comparison with commensals and pathogens. Nat Rev Microbiol. 8 (2010) 171–184.

17. P. Mackiewicz, J. Zakrzewska-Czerwinska, A. Zawilak, M.R. Dudek, S. Cebrat. Where does bacterial replication start? Rules for predicting the oriC region. Nucleic Acids Res. 32 (2004) 3781-3791.

18. L. Zhong, Q. Cheng, X. Tian, L. Zhao L, Z. Qin. Characterization of the replication, transfer, and plasmid/lytic phage cycle of the *Streptomyces* plasmid-phage pZL12. J Bacteriol.192 (2010) 3747-3754.

19. M.B. Łobocka, D.J. Rose, G. Plunkett 3rd, M. Rusin, A. Samojedny, H. Lehnherr H, M.B. Yarmolinsky, F.R. Blattner. Genome of bacteriophage P1. J Bacteriol. 186 (2004) 7032-7068.

20. N.V. Ravin. N15: the linear phage-plasmid. Plasmid 65 (2011) 102-109.

21. D.H. Ives, S. Ikeda. Life on the salvage path: the deoxynucleotide kinases of *Lactobacillus acidophilus* R-26. In: K. Moldave, editor. *Progress in Nucleic Acid Research and Molecular Biology* Volume 59. USA: Academic Press, 1997. pp.205-255.

22. R. Kumar, S. Grover, J.K. Kaushik, V.K. Batish. IS30-related transposon mediated insertional inactivation of bile salt hydrolase (bsh1) gene of *Lactobacillus plantarum* strain Lp20. Microbiol Res. 169 (2014) 553-560.

23. H. Nicoloff, F. Bringel. ISLpl1 is a functional IS30-related insertion element in *Lactobacillus plantarum* that is also found in other lactic acid bacteria. Appl Environ Microbiol. 69 (2003) 6032-6040.

24. Y. Kanesaki, I. Suzuki, S.I. Allakhverdiev, K. Mikami, N. Murata. Salt stress and hyperosmotic stress regulate the expression of different sets of genes in *Synechocystis* sp. PCC 6803. Biochem Biophys Res Commun. 290 (2002) 339–348.

25. S.G. Tetu, B. Brahamsha, D.A. Johnson, V. Tai, K. Phillippy, B. Palenik, I.T. Paulsen. Microarray analysis of phosphate regulation in the marine *Cyanobacterium Synechococcus* sp. WH8102. ISME J. 3 (2009) 835-849.

26. London LE, Price NP, Ryan P, Wang L, Auty MA, Fitzgeral GF, Stanton C, Ross RP. Characterization of a bovine isolate *Lactobacillus mucosae* DPC 6426 which produces an exopolysaccharide predominantly of mannose residues. J Appl Microbiol. 117 (2014) 509-517.

27. Z. Xing, W. Geng, C. Li, Y. Sun, Y. Wang. Comparative genomics of *Lactobacillus kefiranofaciens* ZW3 and related members of *Lactobacillus* spp. reveal adaptations to dairy and gut environments. Sci Rep. 7 (2017) 12827.

28. Dupuis, M-E, Villion M, Magadán, AH, Moineau S. CRISPR-Cas and restriction–modification systems are compatible and increase phage resistance. Nat Comm. 4 (2013) 2087.

29. Richter M, Rosselló-Móra R. Shifting the genomic gold standard for the prokaryotic species definition. Proc Natl Acad Sci USA. 106 (2009) 19126-19131.

30. P.M. Ryan, C.M. Guinane, L.E. London, P.R. Kelleher, G.F. Fitzgerald, N.M. Caplice, R.P. Ross, C. Stanton. Genome sequence of the heteropolysaccharide-producing strain *Lactobacillus mucosae* DPC 6426. Genome Announc. 3 pii (2015) e01350-14.

31. F. Drissi, V. Merhej, C. Blanc-Tailleur, D. Raoult. Draft genome sequence of the *Lactobacillus mucosae* strain Marseille. Genome Announc. 3 pii: (2015) e00841-15.

32. J. Bleckwedel, L.C. Terán, J. Bonacina, L. Saavedra, F. Mozzi F, R.R. Raya. Draft genome sequence of the mannitol-producing strain *Lactobacillus mucosae* CRL573. Genome Announc. 2 pii (2014) e01292-14.

33. M.A. Tormo, M.D. Ferrer, E. Maiques, C. Úbeda, L. Selva, Í. Lasa, J.J. Calvete, R.P. Novick, J.R. Penadés. *Staphylococcus aureus* Pathogenicity Island DNA Is Packaged in Particles Composed of Phage Proteins. J Bacteriol. 190 (2008) 2434–2440.

34. E.A. Pfeiler, T.R. Klaenhammer. Role of transporter proteins in bile tolerance of *Lactobacillus acidophilus*. Appl Environ Microbiol. 75 (2009) 6013-6016.

35. S.A. Jones, M. Jorgensen, F.Z. Chowdhury, R. Rodgers, J. Hartline, M.P. Leatham, C. Struve et al.
Glycogen and maltose utilization by *Escherichia coli* O157:H7 in the mouse intestine. Infect Immun. 76 (2008) 2531-2540.

36. K. Makarova, A. Slesarev, Y. Wolf, A. Sorokin, B. Mirkin, E. Koonin, A. Pavlov, et al. Comparative genomics of the lactic acid bacteria. Proc Natl Acad Sci USA. 103 (2006) 15611–15616.

37. E. Swiatlo, F.R. Champlin, S.C. Holman, W.W. Wilson, J.M. Watt. Contribution of Choline-Binding Proteins to Cell Surface Properties of *Streptococcus pneumoniae.* Infect Immun. 70 (2002) 412–415.

38. M. Hattori, G.A. Torres, N. Tanaka, S. Okada, A. Endo, J. Nakagawa. Detection and analysis of *Lactobacillus paracasei* penicillin-binding proteins revealed the presence of cholate-sensitive penicillin-binding protein 3 and an elongated cell shape in a cholate-sensitive strain. Biosci Microbiota Food Health. 36 (2017) 65–72.

39. E.B. Olsen, J.B. Russell, T. Henick-Kling. Electrogenic L-malate transport by *Lactobacillus plantarum*: a basis for energy derivation from malolactic fermentation. J Bacteriol. 173 (1991) 6199-6206.

40. J.K. Lithgow, E.J. Hayhurst, G. Cohen, Y. Aharonowitz, S.J. Foster. Role of a cysteine synthase in *Staphylococcus aureus*. J Bacteriol. 186 (2004) 1579-1590.

41. E. Årsköld, E. Lohmeier-Vogel, R. Cao, S. Roos, P. Rådström, E.W. van Niel. Phosphoketolase pathway dominates in *Lactobacillus reuteri* ATCC 55730 cont aining dual pathways for glycolysis. J Bacteriol. 190 (2008) 206-212.

42. A.E. Belanger, G.F. Hatfull. Exponential-phase glycogen recycling is essential for growth of *Mycobacterium smegmatis*. J Bacteriol. 181 (1999) 6670-6678.

43. G.M. Seibold, B.J. Eikmanns. The *glgX* gene product of *Corynebacterium glutamicum* is required for glycogen degradation and for fast adaptation to hyperosmotic stress. Microbiology. 153 (2007) 2212-2220.

44. Y.J. Goh, T.R. Klaenhammer. A functional glycogen biosynthesis pathway in *Lactobacillus acidophilus*: expression and analysis of the glg operon. Mol Microbiol. 89 (2013) 1187-200.

45. M.A. Bonafonte, C. Solano, B. Sesma, M. Alvarez, L. Montuenga, D. Garcia-Ros, C. Gamazo. The relationship between glycogen synthesis, biofilm formation and virulence in *Salmonella enteritidis*. FEMS Microbiol Lett. 191 (2000) 31-36.

46. D.E. Chang, D.J. Smalley, D.L. Tucker, M.P. Leatham, W.E. Norris, S.J. Stevenson, A.B. Anderson, J.E. Grissom, D.C. Laux, P.S. Cohen, T. Conway. Carbon nutrition of *Escherichia coli* in the mouse intestine. Proc Natl Acad Sci USA. 101 (2004) 7427-7432.

47. T. Sambou, P. Dinadayala, G. Stadthagen, N. Barilone, Y. Bordat, P. Constant, F. Levillain, et al. Capsular glucan and intracellular glycogen of *Mycobacterium tuberculosis*: biosynthesis and impact on the persistence in mice. Mol Microbiol. 70 (2008) 762-774.

48. L. Bourassa, A. Camilli. Glycogen contributes to the environmental persistence and transmission of *Vibrio cholerae*. Mol Microbiol. 72 (2009) 124-138.

49. Z. Sun, H.M.B. Harris, A. McCann, C. Guo, S. Argimón, W. Zhang, X. Yang, et al. Expanding the biotechnology potential of lactobacilli through comparative genomics of 213 strains and associated genera. Nat Commun. 6 (2015) 8322. doi:10.1038/ncomms9322

50. L. Wang, M.J. Wise. Glycogen with short average chain length enhances bacterial durability. Naturwissenschaften. 98 (2011) 719-729.

51. R. Kant, J. Blom, A. Palva, R.J. Siezen, W.M. de Vos. Comparative genomics of *Lactobacillus*. Microb Biotechnol. 3 (2011) 323-332.

52. Y.J. Goh, T.R. Klaenhammer. Insights into glycogen metabolism in *Lactobacillus acidophilus*: impact on carbohydrate metabolism, stress tolerance and gut retention. Microb Cell Fact. 13 (2014) 94.

53. M. Rossi, A. Amaretti, S. Raimondi. Folate production by probiotic bacteria. Nutrients 3 (2011) 118-134.

54. M. Kuratsu, Y. Hamano, T. Dairi. Analysis of the *Lactobacillus* metabolic pathway. Appl Environ Microbiol. 76 (2010) 7299-7301.

55. J.G. LeBlanc, G.S. de Giori, E.J. Smid, J. Hugenholtz, F. Sesma. Folate production by lactic acid bacteria and other food-grade microorganisms. In: A. Méndez-Vilas A, editor. *Communicating Current Research and Educational Topics and Trends in Applied Microbiology* Volume 1. Spain: Formatex; 2007. pp.329-339.

56. J.E. Laino, E.M. Hebert, G.S. de Giori, J.G. LeBlanc. Draft genome sequence of *Lactobacillus delbrueckii subsp. bulgaricus* CRL871, a folate-producing strain isolated from a northwestern Argentinian yogurt. Genome Announc. 3 (2015) e00693-15.

57. M.P. Balolong, J.P. Chae, D.-K. Kang. Expression and characterisation of neopullulanase from *Lactobacillus mucosae*. Biotechnol Lett. 38 (2016) 1753-1760.

58. T. Kuriki, S. Okada, T. Imanaka. New type of pullulanase from *Bacillus stearothermophilus* and molecular cloning and expression of the gene in *Bacillus subtilis*. J Bacteriol. 170 (1988) 1554–1559.

59. M.J. Yebra, J. Arroyo, P. Sanz, J.A. Priet. Characterization of novel neopullulanase from *Bacillus polymyxa*. Appl Biochem Biotechnol. 68 (1997) 113–120.

60. K.A. Smith, A.A. Salyers. Characterization of a neopullulanase and an a-glucosidase from *Bacteroides thetaiotaomicron* 95-1. J Bacteriol. 173 (1991) 2962–2968

61. H. Mäkeläinen, O. Hasselwander, N. Rautonen, A.C. Ouwehand. Panose, a new prebiotic candidate. Lett Appl Microbiol. 49 (2009) 666–672.

62. D. Takamatsu, B.A. Bensing, P.M. Sullam. Four proteins encoded in the gspB-secY2A2 operon of *Streptococcus gordonii* mediate the intracellular glycosylation of the platelet-binding protein GspB. J Bacteriol. 186 (2004) 7100-7111.

63. B.A. Bensing, B.W. Gibson, P.M. Sullam. The *Streptococcus gordonii* platelet binding protein GspB undergoes glycosylation independently of export. J Bacteriol. 186 (2004) 638-645.

64. I. Von Ossowski, R. Satokari, J. Reunanen, S. Lebeer, S.C.J. De Keersmaecker, et al. Functional characterization of a mucus-specific LPXTG surface adhesin from probiotic *Lactobacillus rhamnosus* GG. Appl Environ Microbiol. 77 (2011) 4465-4472.

65. L.A. Bøhle, D.A. Brede, D.B. Diep, H. Holo, I.F. Nes. Specific degradation of the mucus adhesion-promoting protein (MapA) of *Lactobacillus reuteri* to an antimicrobial peptide. Appl Environ Microbiol. 76 (2010) 7306-7309.

66. M. Watanabe, H. Kinoshita, R. Yukishita, Y. Kawai, K. Kimura, N. Taketomo, Y. Yamazaki, et al. Identification of a new adhesin-like protein from

Lactobacillus mucosae ME-340 with specific affinity to the human blood group A and B antigens. J Appl Microbiol. 109 (2010) 927-935.

67. M. Watanabe, H. Kinoshita, I.N. Huang, K. Eguchi, T. Tsurumi, Y. Kawai, H. Kitazawa, et al. An adhesin-like protein, Lam29, from *Lactobacillus mucosae* ME-340 binds to histone H3 and blood group antigens in human colonic mucus. Biosci Biotechnol Biochem. 76 (2012) 1655-1660.
68. A.L. Delcher, K.A. Bratke, E.C. Powers, S.L. Salzberg. Identifying bacterial genes and endosymbiont DNA with Glimmer. Bioinformatics. 23 (2007) 673-679.
69. T.M. Lowe, S.R. Eddy. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. 25 (1997) 955-964.
70. S.R. Eddy. Profile hidden Markov models. Bioinformatics 14 (1998) 755-763.
71. O.S. Kim, Y.J. Cho, K. Lee, S.H. Yoon, M. Kim, H. Na, S.C. Park, et al. Introducing EzTaxon-e: a prokaryotic 16S rRNA gene sequence database with phylotypes that represent uncultured species. Int J Syst Evol Microbiol. 62 (2012) 716-721.
72. R.L. Tatusov, M.Y. Galperin, D.A. Natale, E.V. Koonin. The COG database: a tool for genome-scale analysis of protein functions and evolution. Nucleic Acids Res. 28 (2000) 33-36.
73. R. Overbeek, T. Begley, R.M. Butler, J.V. Choudhuri, H.Y. Chuang, M. Cohoon, V. de Crecy-Lagard, et al. The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. Nucleic Acids Res. 33 (2005) 5691-5702.

74. K.D. Pruitt, T. Tatusova, T.W. Klimke, D.R. Maglott. NCBI reference sequences: current status, policy and new initiatives. Nucleic Acids Res. 37 (2009) D32-36.

75. C. Yu, N. Zavaljevski, V. Desai, J. Reifman. Genome-wide enzyme annotation with precision control: catalytic families (CatFam) databases. Proteins 74 (2009) 449-460.

76. I. Grissa, G. Vergnaud, C. Pourcel. The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. BMC Bioinformatics 8 (2007a) 172.

77. I. Grissa, G. Vergnaud, C. Pourcel. CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. Nucleic Acids Res. 35 (2007b) W52-7.

78. I. Grissa, G. Vergnaud, C. Pourcel. CRISPRcompar: a website to compare clustered regularly interspaced short palindromic repeats. Nucleic Acids Res. 36 (2008) W145-8.

79. Y. Zhou, Y. Liang, K. Lynch, J.J. Dennis, D.S. Wishart. PHAST: A Fast Phage Search Tool Nucl Acids Res. 39 (2011) W347-W352.

80. A.M. Varani, P. Siguier, E. Gourbeyre, V. Charneau, M. Chandler. ISsaga is an ensemble of web-based methods for high throughput identification and semi-automatic annotation of insertion sequences in prokaryotic genomes. Genome Biol. 12 (2011) R30. doi: 10.1186/gb-2011-12-3-r30.

81. M.G.I. Langille, F.S.L. Brinkman. IslandViewer: an integrated interface for computational identification and visualization of genomic islands. Bioinformatics 25 (2009) 664–665.

82. T.N. Petersen, S. Brunak, G. von Heijne, H. Nielsen. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Methods. 8 (2011) 785-786.

83. A. Krogh, B. Larsson, G. von Heijne, E.L. Sonnhammer. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol. 305 (2001) 567-80.

84. A.S. Juncker, H. Willenbrock, G. von Heijne, H. Nielsen, S. Brunak, A. Krogh. Prediction of lipoprotein signal peptides in Gram- negative bacteria. Protein Sci. 12 (2003) 1652-1662.

85. N. Saitou, M. Nei. The neighbor-joining method: A new method for reconstructing phylogenetic trees. Mol Biol Evol. 4 (1981) 406-425.

86. S. Kumar, G. Stecher, K. Tamura. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. Mol Biol Evol. 33 (2016) 1870-1874.

87. S.H. Yoon, S.M. Ha, S. Kwon, J. Lim, Y. Kim, H. Seo, J. Chun (2017). Introducing EzBioCloud: A taxonomically united database of 16S rRNA and whole genome assemblies. Int J Syst Evol Microbiol. 67 (2017) 1613-1617.

88. A.C.E. Darling, B. Mau, F.R. Blattner, N.T. Perna. Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome res. 14 (2010) 1394-1403.

89. A.E. Darling, B. Mau, N.T. Perna. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. PLoS One. 5 (2010) e11147.

90. J. Krumsiek, R. Arnold, T. Rattei. Gepard: a rapid and sensitive tool for creating dotplots on genome scale. Bioinformatics. 23 (2007) 1026-1028.

91. J. Chun, C.J. Grim, N.A. Hasan, J.H. Lee, S.Y. Choi, B.J. Haley, E. Taviani et al. Comparative genomics reveals mechanism for short-term and long-term

clonal transitions in pandemic *Vibrio cholerae*. Proc Natl Acad Sci USA. 106 (2009) 15442-15447.

92. H. Morita, H. Toh, S. Fukuda, H. Horikawa, K. Oshima, T. Suzuki, M. Murakami et al.

Comparative genome analysis of *Lactobacillus reuteri* and *Lactobacillus fermen tum* reveal a genomic island for reuterin and cobalamin production. DNA Res. 15 (2008) 151-161.

93. E.W. Myers, W. Miller. Optimal alignments in linear space. Comput Appl Biosci. 4 (1988) 11-17.

94. R.K. Aziz, D. Bartels, A.A. Best, M. DeJongh, T. Disz, R.A. Edwards, et al. The RAST Server: Rapid annotations using subsystems technology. BMC Genomics. 9 (2008) 75.

LIST OF FIGURES

Figure 1. Genome maps of *Lactobacillus mucosae* LM1. A) Chromosomal map showing the various features of the complete *L. mucosae* LM1 genome. Circles illustrate the following from outermost to innermost rings: (1) rRNA/tRNA; (2) Reverse CDS; (3) Forwards CDS; (4) GC ratio; and (5) GC Skew. OriC region is located at the 0 Mbp region, where the minimum GC skew and the dnaA gene is found**. B) Genomic islands and prophage regions in the chromosome of *L. mucosae* LM1.** The colored shapes orange blue and red determine the predicted genomic islands as identified by SIGI-HMM (orange) and IslandPath-DIMOB (blue), and red showing the integrated genomic island search results. where In red boxes also show the prophage regions (P1, P2 and P3) identified by PHAST. **C) Plasmid pLM1 depicting incomplete prophage detected by PHAST (pP1).** Circles illustrate the following from outermost to innermost rings: (1) Reverse CDS; (2) Forwards CDS; (3) GC ratio; and (4) GC Skew. Putative prophage region is indicated as pP1.

Figure 2. Heatmap showing relative Average Nucelotide Identity (ANI) between species. Proposed species cut-off boundary is around 95-96%, showing difference with the closest *Lactobacillus* species, *L. fermentum* and *L. reuteri*. Numbered values indicate representative strains of selected Lactobacillus species as follows: 1-8 (*L. mucosae*); 9-11 (*L. fermentum*); 12-14 (*L. reuteri*); 15-16 (*L. gasseri*); 17-19 (*L. johnsonii*); 20 (*L. plantarum*); 21 (*L. rhamnosus*); 22-23 (*L. acidophilus*). Strain names are indicated next to their number on the left. Percentile color values set with 50% as median value.

Figure 3. MAUVE Alignment of all *L. mucosae* genomes available in the NCBI database. Prophage regions 1, 2 and 3 are identified as P1, P2 and P3, respectively, are shown in red boxes, with *L. mucosae* LM1 as the reference genome. Indicated in black boxes are the following: a) chromosomal inversion between strains KHPC15 and KHPX11; b) chromosomal rearrangements in strain S32^T. Numbered blocks identified in specific strains refer to referenced blocks in text.

Figure 4. Overview of the glycogen metabolism pathway in *L. mucosae* LM1. The solid arrow indicates the glycogen synthesis pathway while the dotted arrow indicates the glycogen degradation pathway. Blue indicates the presence of the enzyme in *L. reuteri* JCM 1112, orange indicates the presence of the enzyme in *L. fermentum* IF03956, while green shows the presence of the enzyme in *L. mucosae* LM1. Colourless boxes represent unknown enzymes. The gene name of the enzyme in the complete genome of LM1 is indicated. The pathway was derived from the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways and the review presented by Goh and Klaenhammer (2014).

Figure 5. Overview of the folate biosynthesis pathway present in *L. mucosae* LM1. Blue indicates the presence of the enzyme in *L. reuteri* JCM 1112, orange indicates the presence of the enzyme in *L. fermentum* IF03956, while green shows the presence of the enzyme in *L. mucosae* LM1. Colourless boxes represent unknown enzymes. The gene name of the enzyme in the complete genome of LM1 is indicated. The pathway was derived from KEGG pathways and the review presented by Rossi et al. (2011).

Figure 6. ANI Dendrogram of closely-related lactobacilli strains to *L. mucosae* LM1. Colors represent the relative homology of proteins using BlastP using *L. mucosae* LM1 as reference strain. ANI values are identified in each branch. Strain names are given for

L. mucosae species DPC 6426 and CRL 573, *L. fermentum* IFO 3956 and *L. reuteri* JCM 1112. In the *L. mucosae* LM1 block, genomic islands (orange) and prophage regions (yellow) are indicated. The complete *glg* operon in LM1 is also identified in green.

Figure 1a.

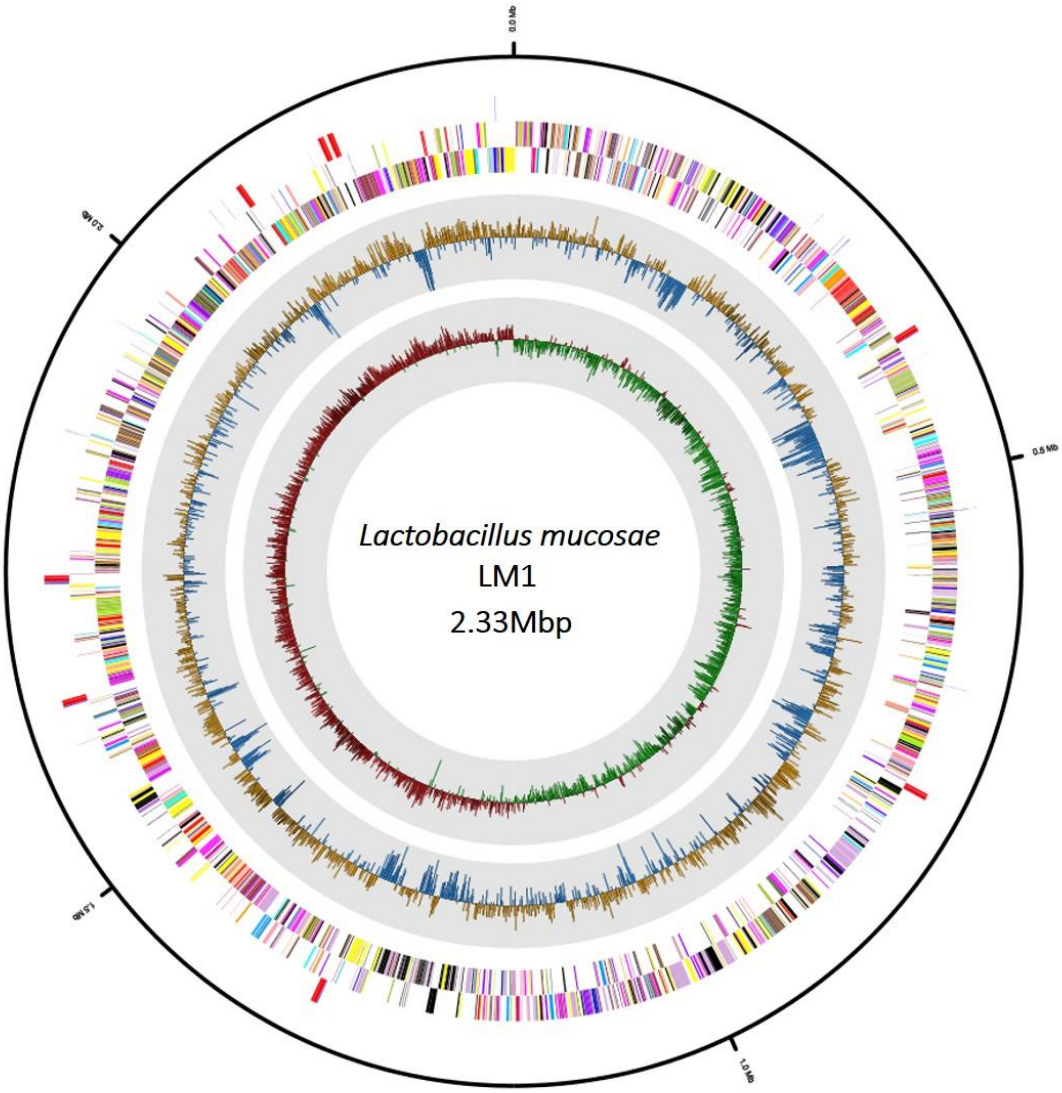


Figure 1b.

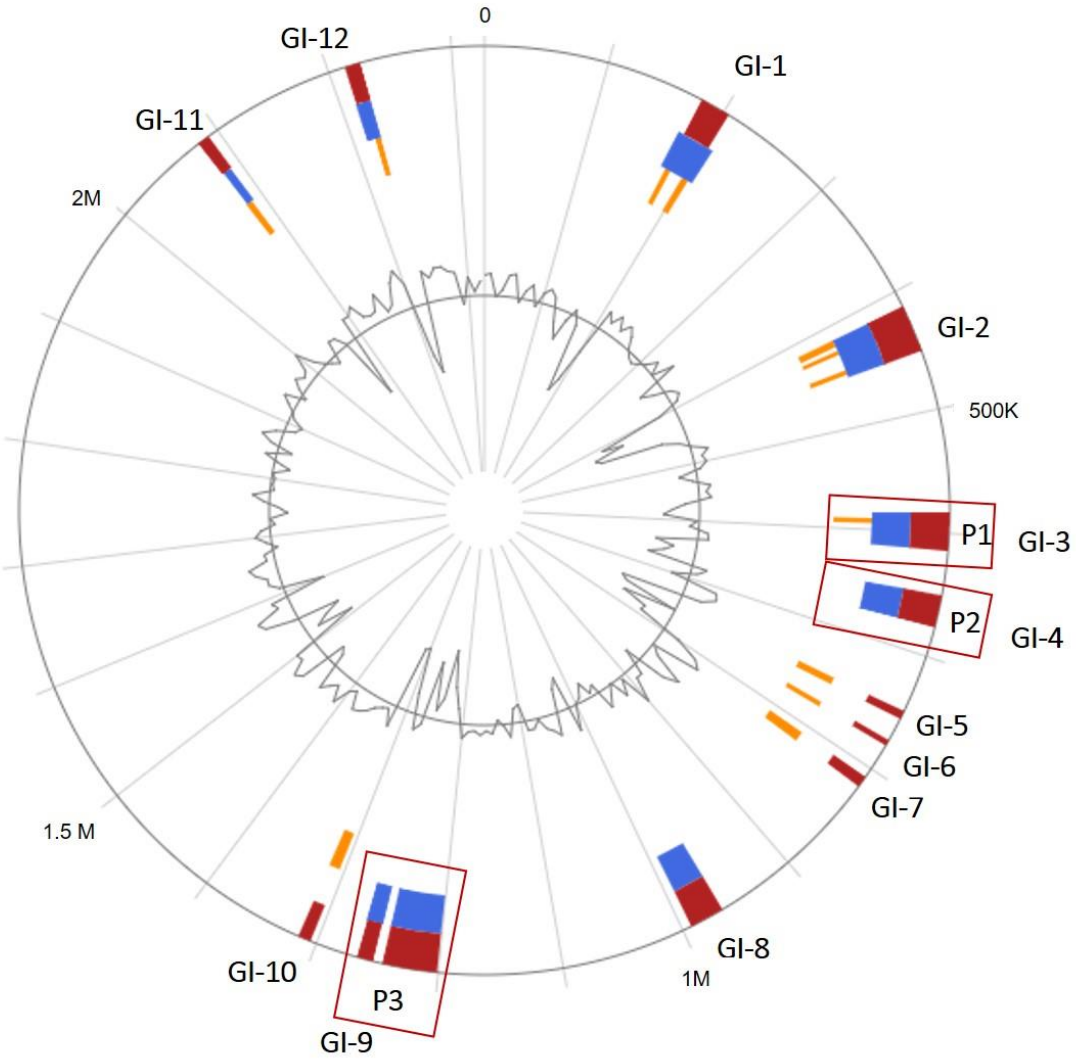

0
GI-12
GI-1
GI-11
2M
GI-2
500K
P1
GI-3
P2
GI-4
GI-5
GI-6
GI-7
1.5 M
GI-10
P3
GI-9
GI-8
1M

Figure 1c.

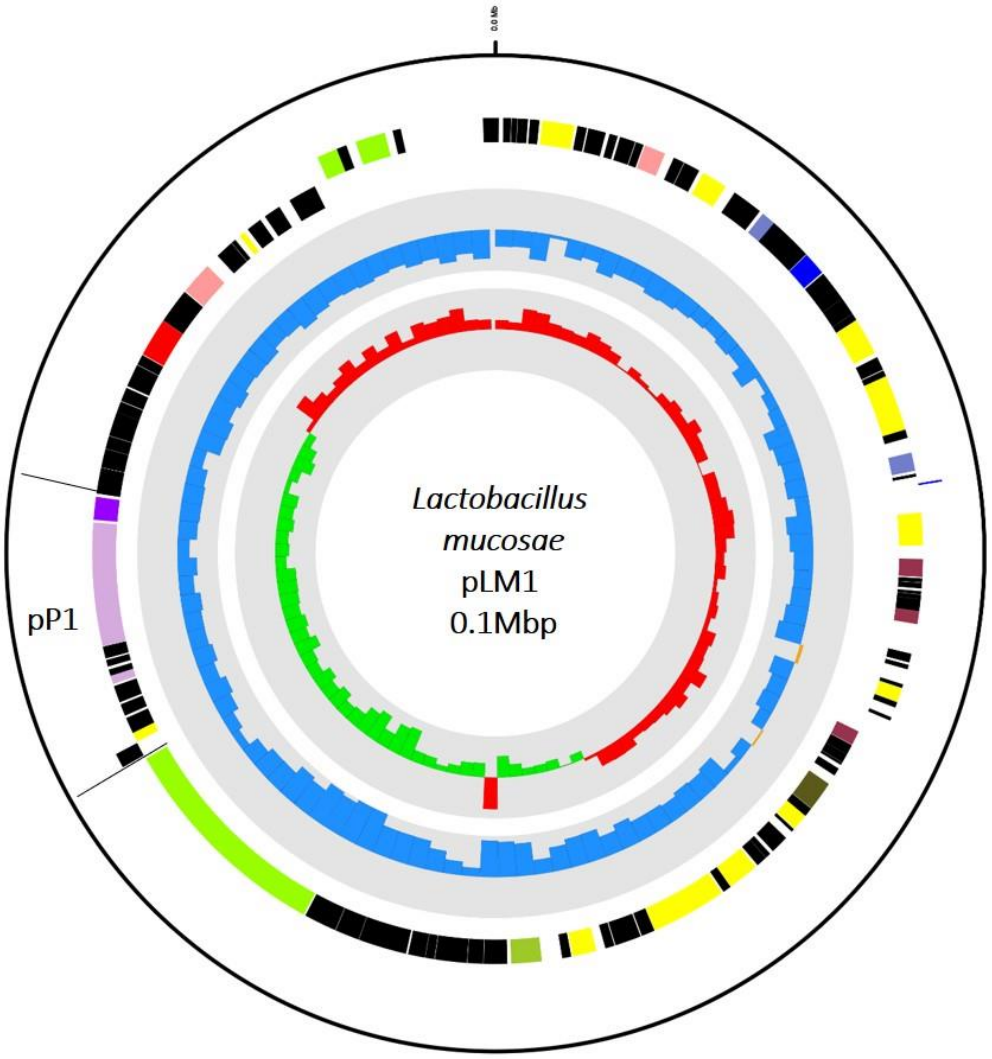


Figure 2.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
1: LM1	X	97.72	97.37	97.32	96.00	95.84	95.80	95.83	71.89	71.26	70.64	70.55	70.92	70.82	67.82	67.41	67.55	67.69	67.80	67.32	66.61	65.62	66.30
2: S32T	97.72	X	97.17	97.17	95.99	95.62	95.49	95.56	70.67	70.71	70.50	69.66	70.30	70.11	66.22	66.52	65.98	66.91	66.68	66.34	65.50	65.95	65.55
3: DPC 6426	97.37	97.17	X	97.36	97.07	96.41	96.35	96.43	70.92	70.77	70.57	70.01	70.39	70.71	68.30	66.24	66.89	66.95	66.12	67.27	66.44	65.69	66.26
4: Marseille	97.32	97.17	97.36	X	96.02	95.87	96	95.87	71.02	71.55	70.39	70.20	70.32	70.37	66.97	67.40	66.41	66.73	66.70	66.50	65.76	65.93	66.48
5: CRL573	96.00	95.99	97.07	96.02	X	96.36	96.15	96.19	71.44	71.47	71.37	70.27	70.69	70.42	66.98	67.35	66.68	66.93	66.41	66.82	65.98	65.98	66.32
6: WCC8	95.84	95.62	96.41	95.87	96.36	X	100.00	100.00	70.53	70.66	70.64	69.66	69.72	70.21	65.69	65.93	65.99	65.59	65.56	66.47	65.67	65.90	65.87
7: KHPC15	95.80	95.49	96.35	95.70	96.15	100.00	X	100.00	70.82	70.46	70.51	69.64	70.20	69.97	65.74	65.99	65.67	65.65	66.14	66.24	65.65	65.87	65.94
8: KHPX11	95.83	95.56	96.43	95.87	96.19	100.00	100.00	X	70.29	70.52	70.36	69.59	69.97	69.76	65.90	66.16	65.70	65.87	65.35	66.19	65.65	66.02	65.79
9: CECT 5716	71.89	70.67	70.92	71.02	71.44	70.53	70.82	70.29	X	99.45	98.92	69.65	69.36	69.84	67.54	67.87	67.63	67.77	66.63	67.51	67.33	65.67	66.79
10: IFO 3956	71.26	70.71	70.77	71.55	71.47	70.66	70.46	70.52	99.45	X	99.02	69.62	69.57	69.62	67.72	68.05	66.69	68.12	66.99	67.63	67.71	65.75	66.39
11: 28-3 CHN	70.64	70.50	70.57	70.39	71.37	70.64	70.51	70.36	98.92	99.02	X	68.94	69.02	69.17	66.37	66.27	66.48	65.98	65.57	67.14	66.03	65.99	65.69
12: 100-23	70.55	69.66	70.01	70.20	70.27	69.66	69.64	69.59	69.65	69.62	68.94	X	96.52	96.07	67.85	67.37	68.28	67.82	68.22	68.14	66.40	68.55	68.63
13: ATCC 53608	70.92	70.30	70.39	70.32	70.69	69.72	70.20	69.97	69.36	69.57	69.02	96.52	X	96.22	67.95	68.13	67.89	68.13	67.99	68.04	66.50	67.32	67.86
14: JCM 1112	70.82	70.11	70.71	70.37	70.42	70.21	69.97	69.76	69.84	69.62	69.17	96.07	96.22	X	68.16	67.32	67.58	68.02	67.82	68.16	66.35	67.95	67.94
15: ATCC 33323	67.82	66.22	68.30	66.97	66.98	65.69	65.74	65.90	67.54	67.72	66.37	67.85	67.95	68.16	X	93.80	85.49	85.58	85.21	66.82	66.38	73.99	73.30
16: 4M13	67.41	66.52	66.24	67.40	67.35	65.93	65.99	66.16	67.87	68.05	66.27	67.37	68.13	67.32	93.80	X	85.28	85.13	85.35	66.65	66.67	74.13	73.54
17: NCC 533	67.55	65.98	66.89	66.41	66.68	65.99	65.67	65.70	67.63	66.69	66.48	68.28	67.89	67.58	85.49	85.28	X	97.29	97.62	67.33	66.10	74.46	73.56
18: DPC 6026	67.69	66.91	66.95	66.73	66.93	65.59	65.65	65.87	67.77	68.12	65.98	67.82	68.13	68.02	85.58	85.13	97.29	X	98.10	66.43	65.54	74.05	73.62
19: PF01	67.80	66.68	66.12	66.70	66.41	65.56	66.14	65.35	66.63	66.99	65.57	68.22	67.99	67.82	85.21	85.35	97.62	98.10	X	66.60	65.54	74.24	73.23
20: WCSF1	67.32	66.34	67.27	66.50	66.82	66.47	66.24	66.19	67.51	67.63	67.14	68.14	68.04	68.16	66.82	66.65	67.33	66.43	66.6	X	66.81	65.32	66.37
21: GG	66.61	65.5	66.44	65.76	65.98	65.67	65.65	65.65	67.33	67.71	66.03	66.40	66.50	66.35	66.38	66.67	66.10	65.54	65.54	66.81	X	65.24	66.18
22: NCFM	65.62	65.95	65.69	65.93	65.98	65.90	65.87	66.02	65.67	65.75	65.99	68.55	67.32	67.95	73.99	74.13	74.46	74.05	74.24	65.32	65.24	X	99.91
23: La-14	66.30	65.55	66.26	66.48	66.32	65.87	65.94	65.79	66.79	66.39	65.69	68.63	67.86	67.94	73.30	73.54	73.56	73.62	73.23	66.37	66.18	99.91	X

Color Legend: Lowest — Percentile Color Values — Highest

Figure 3.

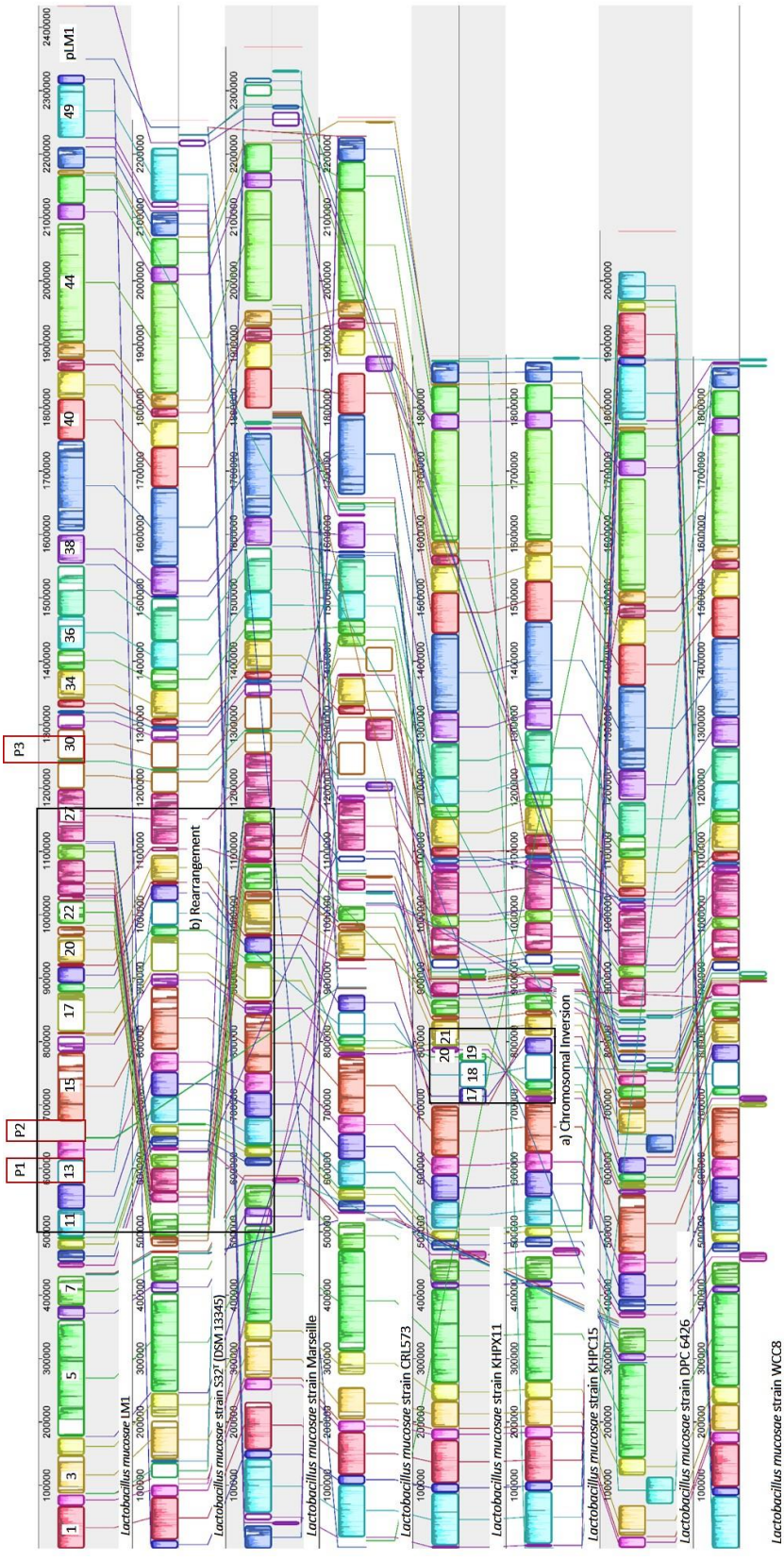


Figure 4.

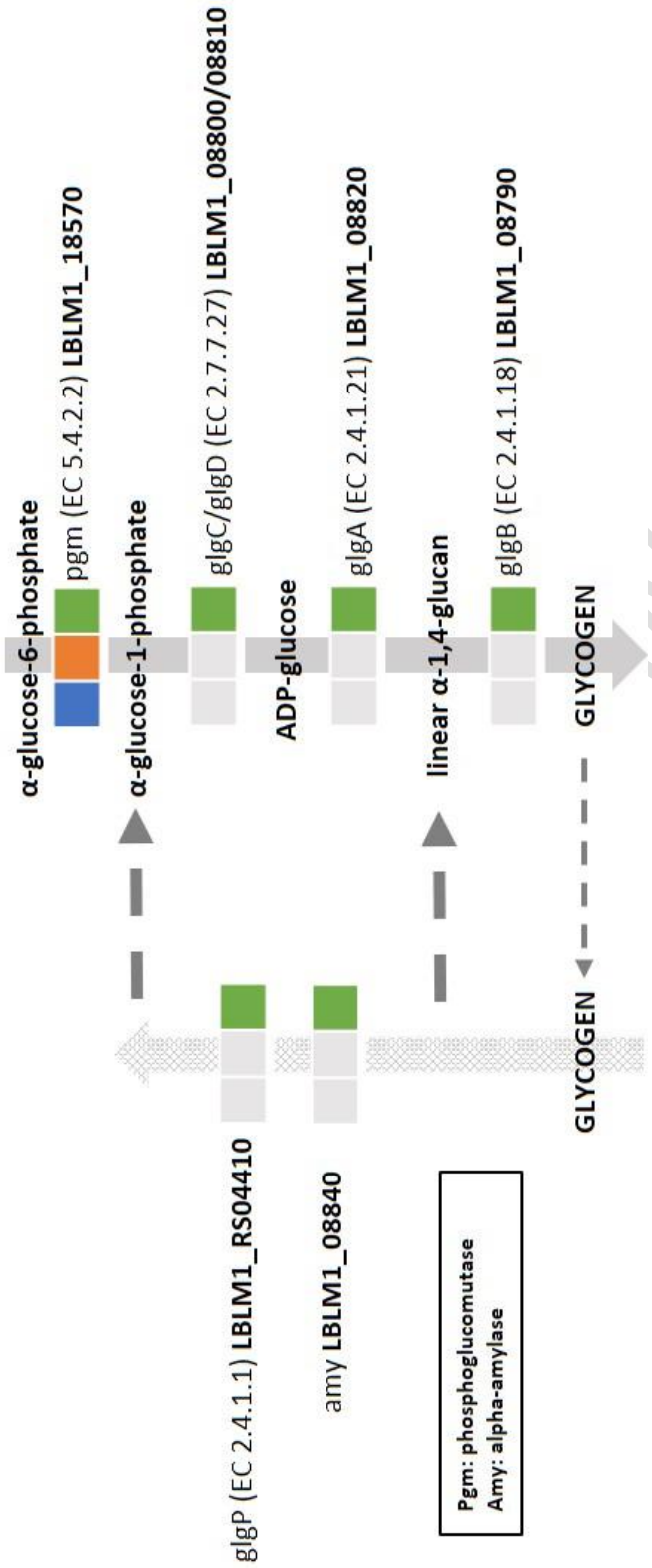


Figure 5.

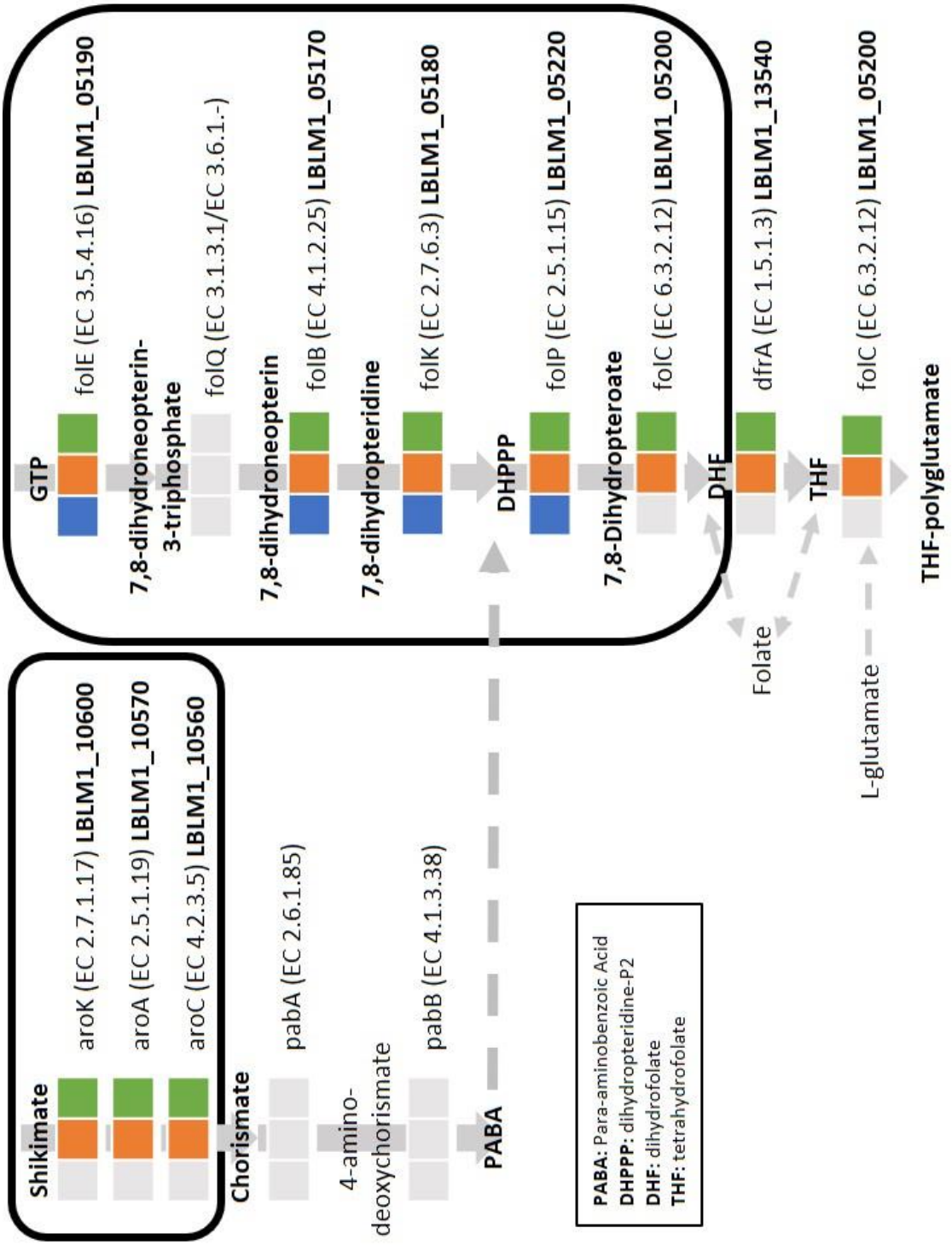
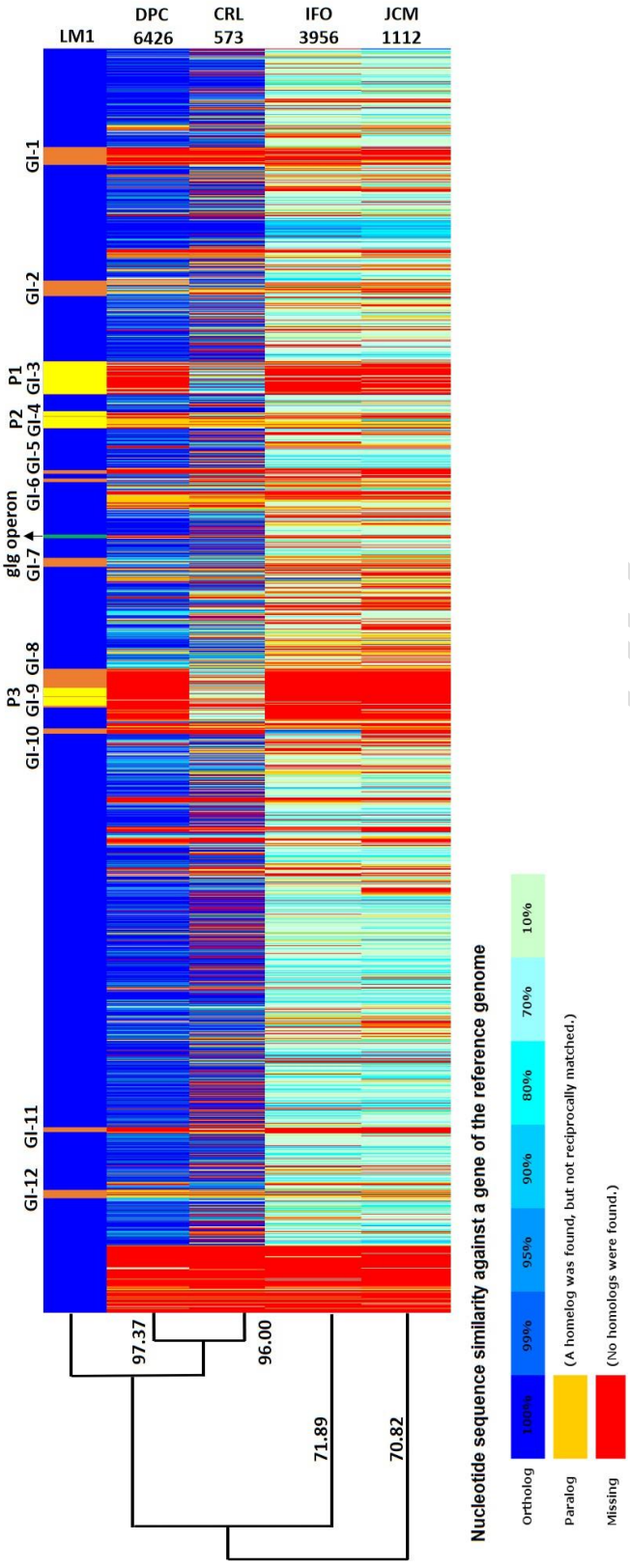

GTP
folE (EC 3.5.4.16) LBLM1_05190
7,8-dihydroneopterin-3-triphosphate
folQ (EC 3.1.3.1/EC 3.6.1.-)
7,8-dihydroneopterin
folB (EC 4.1.2.25) LBLM1_05170
7,8-dihydropteridine
folK (EC 2.7.6.3) LBLM1_05180
DHPPP
folP (EC 2.5.1.15) LBLM1_05220
7,8-Dihydropteroate
folC (EC 6.3.2.12) LBLM1_05200
DHF
dfrA (EC 1.5.1.3) LBLM1_13540
THF
folC (EC 6.3.2.12) LBLM1_05200
THF-polyglutamate
Folate
L-glutamate
Shikimate
aroK (EC 2.7.1.17) LBLM1_10600
aroA (EC 2.5.1.19) LBLM1_10570
aroC (EC 4.2.3.5) LBLM1_10560
Chorismate
pabA (EC 2.6.1.85)
4-amino-deoxychorismate
pabB (EC 4.1.3.38)
PABA
PABA: Para-aminobenzoic Acid
DHPPP: dihydropteridine-P2
DHF: dihydrofolate
THF: tetrahydrofolate

Figure 6.



LIST OF SUPPLEMENTARY TABLES

Table S1. Genome statistics of *L. mucosae* LM1

Table S2. Number of genes associated with general COG functional categories

Table S3. Identified Genomic Island Regions and Genes in the LM1 Chromosome

Table S4. Orthologous genes in *L. mucosae* LM1 involved in inverted regions of *L. mucosae* strain KHPX11

Table S5. Orthologous genes in *L. mucosae* LM1 involved in rearranged regions of *L. mucosae* strain S32^{T}

LIST OF SUPPLEMENTARY FILE

File S1. Putative IS elements sequences in *L. mucosae* LM1