1 **Complete genome sequence of *Lactobacillus plantarum* LZ227, a potential**
2 **probiotic strain producing B-group vitamins**

3

4 Ping Li, Qingqing Zhou, Qing Gu *

5

6 Key Laboratory for Food Microbial Technology of Zhejiang Province, College of
7 Food Science and Biotechnology, Zhejiang Gongshang University, Hangzhou 310018,
8 China

9

10 *Corresponding author.

11 E-mail address: guqing2002@hotmail.com (Q. Gu).

12 Tel: +86 571 28008903

13 Fax: +86 571 28008900

14

Abstract:

B-group vitamins play an important role in human metabolism, whose deficiencies are associated with a variety of disorders and diseases. Certain microorganisms such as Lactic acid bacteria (LAB) have been shown to have capacities for B-group vitamin production and thus could potentially replace chemically synthesized vitamins for food fortification. A potential probiotic strain named *Lactobacillus plantarum* LZ227, which was isolated from raw cow milk in this study, exhibits the ability to produce B-group vitamins. Complete genome sequencing of LZ227 was performed to gain insights into the genetic elements involved in B-group vitamin production. The genome of LZ227 contains a circular 3,131,750-bp chromosome, three circular plasmids and two predicted linear plasmids. LZ227 also contains gene clusters for biosynthesis of both riboflavin and folate. This genome sequence provides a basis for further elucidation of its molecular genetics and probiotic functions, and will facilitate its applications as starter cultures in food industry.

Keywords: *L. plantarum* LZ227; Complete genome sequence; B-group vitamins; Folate; Riboflavin

Vitamins are generally classified into two groups, the fat-soluble vitamins and the water-soluble vitamins. B-group vitamins are water-soluble vitamins, including thiamin (B1), riboflavin (B2), niacin (B3), pantothenic acid (B5), pyridoxine (B6), folate (B11) and cobalamin (B12). B-group vitamins play an important role in metabolism, however, humans cannot produce them and have to obtain them exogenously (Capozzi *et al.*, 2012). Deficiencies of B-group vitamins are associated with various disorders and diseases, for example, folate deficiency is linked to the Alzheimer's Disease, neural tube defects and coronary heart diseases *etc.* (Rossi *et al.*, 2011). Certain microorganisms have capacities for B-group vitamin production, and have been applied to replace the costly chemically synthesized vitamins for food fortification, either by producing and purifying vitamins through microbial fermentation or by *in situ* fortification in fermented foods. Lactic acid bacteria (LAB), especially strains belonging to genera *Lactobacillus* and *Bifidobacterium*, offer unique opportunities to improve nutritional values and vitamin contents of food products due to their generally regarded as safe (GRAS) status (Capozzi *et al.*, 2012).

Several *Lactobacillus* strains have been identified for their ability to produce B-group vitamins. For example, we found *L. plantarum* LZ95 produces riboflavin (Li and Gu, 2016) and *L. reuteri* ZJ03 enhances vitamin B12 content in soy-yogurt (Gu *et al.*, 2015), and Wegkamp *et al* found *L. plantarum* strain WCSF1 is able to produce large amounts of folate (Wegkamp *et al.*, 2010) but not riboflavin (Kleerebezem *et al.*, 2003; LeBlanc *et al.*, 2011). Interestingly, we found that, a newly isolated LAB strain LZ227 from raw cow milk, exhibited capability of producing both riboflavin and folate. As shown in Fig. 1, LZ227 grew well on both folate-deficient medium and riboflavin-deficient medium, and could support the growth of auxotroph strain *Lactobacillus casei* subsp. *rhamnosus* ATCC 7469. A phylogenetic tree, constructed by MEGA software (version 6.0) and neighbor-joining method, showed a distinct branch of LZ227 with *L. plantarum* strains, and *L. plantarum* LZ95 was identified as the closest evolutionary relative of LZ227 (Fig. 2).

In order to gain insights into the genetic bases in its B-group vitamin production, we sequenced the whole genome of LZ227. Genomic DNA was extracted using the Qiagen DNA extraction kit. Quantity and quality of extracted genomic DNA were determined by Agilent 2100 Bioanalyzer. Whole genome sequencing of LZ227 was performed using a combined strategy of PacBio RSII sequencing and Illumina paired-end sequencing technology. A total of 96,698 reads with an average length of

14,288 nt were obtained by the PacBio RS II sequencing, which were *de novo* assembled using HGAP 2.3.0, resulting in 6 major contigs with more than 83-fold average coverage (Chin *et al.*, 2013). Gene prediction was carried out with the Glimmer 3.02 (Delcher *et al.*, 1999) and gene annotation was performed using the NCBI Prokaryotic Genome Annotation Pipeline.

The complete *L. plantarum* LZ227 genome was composed of a circular 3.13 Mb chromosome; three circular plasmids named LZ227p1 (74,177 bp), LZ227p2 (52,148 bp) and LZ227p3 (35,461 bp); and two predicted linear plasmids LZ227p4 (93,544 bp) and LZ227p5 (38,212 bp). GC content of the whole genome was 44.71% and that of the above plasmids were 41.33 %, 41.83 %, 39.61 %, 39.89 %, and 39.80%, respectively. Amongst the 3262 total identified CDSs, 2958 were in the chromosome, and 78, 59, 34, 95 and 36 were in plasmids LZ227p1 - LZ227p5, respectively. 16 rRNA operons and 71 tRNAs were found in the chromosome (Table 1). 2815 of the identified genes were classified to functional categories based on clusters of orthologous genes (COG) designation (Tatusov *et al.*, 2000): 148 genes for translation, 268 genes for transcription, 279 genes for replication, recombination and repair, 29 genes for cell cycle control, mitosis and meiosis, 66 genes for defense mechanisms, 89 genes for signal transduction mechanisms, 147 genes for cell wall/membrane biogenesis, 5 genes for cell motility, 24 genes for intracellular trafficking and secretion, 61 genes for posttranslational modification, protein turnover and chaperones, 115 genes for energy production and conversion, 265 genes for carbohydrate transport and metabolism, 263 genes for amino acid transport and metabolism, 89 genes for nucleotide transport and metabolism, 80 genes for coenzyme transport and metabolism, 74 genes for lipid transport and metabolism, 170 genes for inorganic ion transport and metabolism, 39 genes for secondary metabolites biogenesis, transport and catabolism, 210 genes for unknown functions, 394 genes for general functions (Table S1, Fig. 3). *L. plantarum* LZ227 contains most genes that function in replication, recombination and repair, nucleotide transport and metabolism, cell cycle control and defense mechanisms, compared to the three closest evolutionary relatives: *L. plantarum* LZ95 (CP012122) (Li, *et al.*, 2016), *L. plantarum* WCFS1 (NC_004567) (Kleerebezem *et al.*, 2003), and *L. plantarum* JDM1 (NC_012984) (Zhang *et al.*, 2009), likely indicating a strong cell proliferation potential and specific genetic adaptation for residing in the gastrointestinal tract.

Genes involved in biosynthesis of both riboflavin and folate were identified in

genome of *L. plantarum* LZ227. Genetic information for riboflavin biosynthesis is strain specific in LAB. Strains without *rib* genes or with incomplete *rib* structural genes are usually unable to produce riboflavin (Kleerebezem *et al.*, 2003). A complete functional *rib* operon, including genes *ribH* (riboflavin synthase beta subunit, A8704_RS05995), *ribA* (3, 4-dihydroxy-2-butanone 4-phosphate synthase, A8704_RS06000), *ribB* (riboflavin synthase subunit alpha, A8704_RS06005) and *ribG* (riboflavin-specific deaminase and reductase, A8704_RS06010), was found in LZ227 by *in silico* analysis (supplemental Fig. 1). The gene *ribC* (A8704_RS08420 and A8704_RS03660), encoding bifunctional flavokinase/FAD synthetase that catalyzes the conversion of riboflavin to two active forms FMN and FAD, was also identified in LZ227, as that in its closest evolutionary relatives LZ95 and JDM1 (Burgess *et al.*, 2006; Li, *et al.*, 2016).

De novo biosynthesis of folate originates from two precursors, DHPPP (6-hydroxymethyl-7, 8-dihydropterin pyrophosphate) and pABA (para-aminobenzoic acid), and ultimately leads to THF-polyglu (tetrahydrofolate-polyglutamate). Chorismate is required for pABA production (Capozzi *et al.*, 2012; Rossi *et al.*, 2011). *De novo* pABA synthesis is absent from most of the sequenced members of genus *Lactobacillus*, either due to defects of chorismate production or the lack of enzymes for chorismate conversion into pABA (Capozzi *et al.*, 2012). Notably, we identified genes for chorismate biosynthesis in LZ227: *aroB* (3-dehydroquinate synthetase, A8704_RS07355, A8704_RS08455, and A8704_RS09865), *aroD* (3-dehydroquinate dehydratase, A8704_RS12220 and A8704_RS00080), *aroE* (shikimate 5-dehydrogenase, A8704_RS12195, A8704_RS12200 and A8704_RS12215), *aroK* (shikimate kinase, A8704_RS03650), *aroA* (5-enolpyruvylshikimate-3-phosphate synthase, A8704_RS03640), and *aroC* (chorismate synthase, A8704_RS03630) (Table 2). Genes for pABA biosynthesis were also found in LZ227: *pabA* (anthranilate/para-aminobenzoate synthases component II, A8704_RS05225) and *pabB* (anthranilate/para-aminobenzoate synthases component I, A8704_RS05230) (Table 2). However, JDM1, one of the closest evolutionary relatives of LZ227, lacked genes encoding pABA (Table 2). Moreover, we identified genes for biosynthesis of DHPPP in LZ227: *folE* (GTP cyclohydrolase I, A8704_RS12950), *folQ* (dihydroneopterin triphosphate pyrophosphohydrolase, A8704_RS02240), *folB* (dihydroneopterin aldolase, A8704_RS12940) and *folK* (2-amino-4-hydroxy-6-hydroxymethyldihydropteridine pyrophosphokinase,

A8704_RS12945). There were also genes for subsequent biosynthesis of THF-polyglu: *folP* (dihydropteroate synthase, A8704_RS12965), *folC* (folylpolyglutamate synthase, orf03070-2696 and A8704_RS12955) and *dfrA* (dihydrofolate reductase, A8704_RS04325). Notably, LZ227 contains more than one copy of certain genes that are involved in folate biosynthesis, such as *folC*, overexpression of which has been shown to significantly increase folate production (LeBlanc *et al.*, 2010). These results provide clues to explain the production of both folate and riboflavin in LZ227. However, more *in vivo* and *in vitro* studies are required for confirming the genes' functions.

Together, the whole genome data of *L. plantarum* LZ227 will allow us to better understand the molecular basis for its probiotic functions and could facilitate its potential applications as starter cultures in food industry.

Nucleotide sequence accession number. The complete genome sequence of *L. plantarum* LZ227 has been deposited in GenBank under the accession number CP015857-CP015862. This strain has been deposited at the China Center for Type Culture Collection under the accession number CCTCC AB 2016119.

Conflict of interest

The authors declare that they have no conflict of interest.

Acknowledgements

This project was funded by International Science & Technology Cooperation Program of China (No. 2013DFA32330), National Science Foundation of China (No. 31540044, No. 31271821), National High Technology Research and Development Program ("863" Program) of China (No. 2014AA022210-08), National Science Foundation of Zhejiang Province (No. LY16C200002), Science Foundation of the Zhejiang Education Department (No. Y201534497), and National Research Foundation for the Doctoral Program of Higher Education (No. 20133326110005).

References

Burgess, C.M., Smid, E.J., Rutten, G., van Sinderen, D., 2006. A general method for selection of riboflavin-overproducing food grade micro-organisms. *Microb Cell Fact*

169 5, 24.

170 Capozzi, V., Russo, P., Duenas, M.T., Lopez, P., Spano, G., 2012. Lactic acid
 171 bacteria producing B-group vitamins: a great potential for functional cereals products.
 172 Appl. Microbiol. Biotechnol. 96, 1383-1394.

173 Chin, C.S., Alexander, D.H., Marks, P., Klammer, A.A., Drake, J., Heiner, C., Clum,
 174 A., Copeland, A., Huddleston, J., Eichler, E.E., Turner, S.W., Korlach, J., 2013.
 175 Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing
 176 data. Nat. Methods. 10, 563-569.

177 Delcher, A.L., Harmon, D., Kasif, S., White, O., Salzberg, S.L., 1999. Improved
 178 microbial gene identification with GLIMMER. Nucleic. Acids. Res. 27, 4636-4641.

179 Gu, Q., Zhang, C., Song, D., Li, P., Zhu, X., 2015. Enhancing vitamin B12 content
 180 in soy-yogurt by *Lactobacillus reuteri*. Int. J. Food Microbiol. 206, 56-59.

181 Kleerebezem, M., Boekhorst, J., van Kranenburg, R., Molenaar, D., Kuipers, O.P.,
 182 Leer, R., Tarchini, R., Peters, S.A., Sandbrink, H.M., Fiers, M.W., Stiekema, W.,
 183 Lankhorst, R.M., Bron, P.A., Hoffer, S.M., Groot, M.N., Kerkhoven, R., de Vries, M.,
 184 Ursing, B., de Vos, W.M., Siezen, R.J., 2003. Complete genome sequence of
 185 *Lactobacillus plantarum* WCFS1. Proc. Natl. Acad. Sci. U. S. A. 100, 1990-1995.

186 LeBlanc, J.G., Laino, J.E., del Valle, M.J., Vannini, V., van Sinderen, D., Taranto, M.P.,
 187 de Valdez, G.F., de Giori, G.S., Sesma, F., 2011. B-group vitamin production by
 188 lactic acid bacteria--current knowledge and potential applications. J. Appl. Microbiol
 189 111, 1297-1309.

190 LeBlanc, J.G., Sybesma, W., Starrenburg, M., Sesma, F., de Vos, W.M., de Giori, G.S.,
 191 Hugenholtz, J., 2010. Supplementation with engineered *Lactococcus lactis* improves
 192 the folate status in deficient rats. Nutrition. 26, 835-841.

193 Li, P., Gu, Q., 2016. Complete genome sequence of *Lactobacillus plantarum* LZ95,
 194 a potential probiotic strain producing bacteriocins and B-group vitamin riboflavin. J.
 195 Biotechnol. 229, 1-2.

196 Rossi, M., Amaretti, A., Raimondi, S., 2011. Folate production by probiotic bacteria.
 197 Nutrients. 3, 118-134.

198 Tatusov, R.L., Galperin, M.Y., Natale, D.A., Koonin, E.V., 2000. The cog database: A
 199 tool for genome-scale analysis of protein functions and evolution, Nucleic. Acids.
 200 Res. 28, 33-36.

201 Wegkamp, A., Mars, A.E., Faijes, M., Molenaar, D., de Vos, R.C., Klaus, S.M.,
 202 Hanson, A.D., de Vos, W.M., Smid, E.J., 2010. Physiological responses to folate

203 overproduction in *Lactobacillus plantarum* WCFS1. Microb. Cell Fact. 9, 100.
204 Zhang, Z.Y., Liu, C., Zhu, Y.Z., Zhong, Y., Zhu, Y.Q., Zheng, H.J., Zhao, G.P., Wang,
205 S.Y., Guo, X.K., 2009. Complete genome sequence of *lactobacillus plantarum* jdm1,
206 J. Bacteriol. 191, 5020-5021.

207

Figure legends

Fig. 1 Folate (A) and riboflavin (B) production capacity of *L. plantarum* LZ227.

Turbidimetric assay was used to observe growth of *L. plantarum* LZ227, auxotroph strain *Lactobacillus casei* subsp. *rhamnosus* ATCC 7469 and ATCC 7469 in the presence of *L. plantarum* LZ227 supernatant in folate-deficient liquid medium and riboflavin-deficient liquid medium, respectively. 1, *L. plantarum* LZ227; 2, ATCC 7469 in the presence of *L. plantarum* LZ227 supernatant; 3, ATCC 7469. OD₆₀₀ of the strains was measured after incubation in the media for 10 h. All experiments were performed in triplicate.

Fig. 2 Phylogenetic tree of *L. plantarum* LZ227. The tree was constructed using software MEGA 6.0 by neighbour-joining method based on 16S rDNA gene sequences with 1000 replications in bootstrap test.

Fig. 3 Circular genome map of *L. plantarum* LZ227. From the innermost circles, circle (1) illustrates the GC skew (G-C/G+C). The value is plotted as the deviation from the average GC skew of the entire sequence. Circle (2) GC content, the GC content is plotted using a sliding window, as the deviation from the average GC content of the entire sequence. Circle (3) denotes tRNA/rRNA. Circle (4, 5) indicates the CDSs, colored according to COG function categories, 4 is backward strand, 5 is forward strand. Circle (6, 7) demonstrates m4C and m6A sites in CDS/rRNA/tRNA, 6 is backward strand, 7 is forward strand. Circle (8) denotes m4C and m6A sites in intergene regions.

A folate-deficient medium

B riboflavin-deficient medium

231

232

0.01

233

234

235

236

Tables

Table 1

Features of *Lactobacillus plantarum* LZ227 genome.

Attributes	Values	
Genome size (bp)	3,131,750	242
GC content (%)	44.71%	243
Plasmid	5	244
rRNAs	16	245
tRNAs	70	246
Other RNAs	4	247
Total predicted CDSs	3262 (275 in plasmids)	248

Table 2

Pathways for the biosynthesis of folate in *L. plantarum* LZ227 and other *L. plantarum* strains.

strain	Chorismate							pABA		DHPPP				THF-polyglu		
	<i>aroF</i>	<i>aroB</i>	<i>aroD</i>	<i>aroE</i>	<i>aroK</i>	<i>aroA</i>	<i>aroC</i>	<i>pabA</i>	<i>pabB</i>	<i>folE</i>	<i>folQ</i>	<i>folB</i>	<i>folK</i>	<i>folP</i>	<i>folC</i>	<i>dfra</i>
<i>L. plantarum</i> WCFS1	●	●	●	●	●	●	●	○	○	●	●	●	●	□	□	□
<i>L. plantarum</i> LZ227	●	●	●	●	●	●	●	○	○	●	●	●	●	□	□	□
<i>L. plantarum</i> LZ95	●	●	●	●	●	●	●	○	○	●		●	●	□	□	□
<i>L. plantarum</i> JDM1	●	●	●	●	●	●	●			●	●	●	●	□	□	□