Indian J Microbiol (July–Sept 2016) 56(3):265–276
DOI 10.1007/s12088-016-0605-5

ORIGINAL ARTICLE

Comparative Genomics Reveals Biomarkers to Identify *Lactobacillus* Species

Shikha Koul[1,2] · **Vipin Chandra Kalia**[1,2]

Received: 25 May 2016 / Accepted: 9 June 2016 / Published online: 16 June 2016

Abstract Bacteria possessing multiple copies of 16S rRNA (*rrs*) gene demonstrate high intragenomic heterogeneity. It hinders clear distinction at species level and even leads to overestimation of the bacterial diversity. Fifty completely sequenced genomes belonging to 19 species of *Lactobacillus* species were found to possess 4–9 copies of *rrs* each. Multiple sequence alignment of 268 *rrs* genes from all the 19 species could be classified into 20 groups. *Lactobacillus sanfranciscensis* TMW 1.1304 was the only species where all the 7 copies of *rrs* were exactly similar and thus formed a distinct group. In order to circumvent the problem of high heterogeneity arising due to multiple copies of *rrs*, 19 additional genes (732–3645 nucleotides in size) common to *Lactobacillus* genomes, were selected and digested with 10 Type II restriction endonucleases (RE), under *in silico* conditions. The following unique gene—RE combinations: *recA* (1098 nts)—*HpyCH4 V, CviAII, BfuCI* and *RsaI* were found to be useful in identifying 29 strains representing 17 species. Digestion patterns of genes—*ruvB* (1020 nts), *dnaA* (1368 nts), *purA* (1290 nts), *dnaJ* (1140 nts), and *gyrB* (1944 nts) in combination with REs—*AluI, BfuCI, CviAI, Taq1,* and *Tru9I* allowed clear identification of an additional 14 strains belonging to 8 species. Digestion pattern of genes *recA, ruvB*, *dnaA*, *purA*, *dnaJ* and *gyrB* can be used as biomarkers for identifying different species of *Lactobacillus*.

Keywords *Lactobacillus* · Diagnosis · Biomarkers · Genome · *In silico* · Restriction endonuclease

Introduction

The diversity of microbial world has fascinated researchers since their discovery. Rapid progress in identifying bacteria has been made during the last 3 decades. Prof. Carl R. Woese developed an innovative strategy which brought around a revolutionary change in evolutionary biology and taxonomy. The most widely employed gene for bacterial identification has been 16S rDNA (*rrs*). Almost all laboratories around the world have mastered the technique of sequencing this gene. Ribosomal database project (RDP) (https://rdp.cme.msu.edu/), is a depository of 3,224,600 *rrs* sequences. Using the unique signatures and restriction endonuclease (RE) digestion patterns of *rrs*, it has been possible to re-classify bacteria, which were identified so far only up to genus level to species level [1–4]. In spite of such a roaring success of *rrs*, there have been cases where the gene sequences are quite similar and don't prove effective in distinguishing closely related organisms. Another limitation generally encountered in the usage of *rrs* is in the case of organisms which harbour it in multiple copies: *Clostridium, Yersinia, Vibrio, Staphylococcus, Streptococcus*, etc. Here, high intragenomic heterogeneity leads to mis-identification and even overestimation of bacterial populations [5–10].

Electronic supplementary material The online version of this article (doi:10.1007/s12088-016-0605-5) contains supplementary material, which is available to authorized users.

✉ Vipin Chandra Kalia
vckalia@igib.res.in; vc_kalia@yahoo.co.in

[1] Microbial Biotechnology and Genomics, CSIR - Institute of Genomics and Integrative Biology (IGIB), Delhi University Campus, Mall Road, Delhi 110007, India

[2] Academy of Scientific and Innovative Research (AcSIR), 2, Rafi Marg, Anusandhan Bhawan, Delhi 110001, India

The Trouble with *Lactobacillus* Identification

Conventional methods of identifying *Lactobacillus* involve phenotypic and biochemical characterization. Nevertheless, with the advent of molecular biology, gene based methods were found to be more reliable, precise and consistent. Among the various genes used for identifying bacteria, *rrs* has been the most successful [11, 12]. It has completely revolutionized the way taxonomy has developed. A large number of genomic tools have proved helpful in establishing taxonomic and evolutionary relationships. All molecular methods developed so far have been proven to have some merits vis-a-vis others. The need to develop new genomic tools and find new biomarkers has been associated with the difficulties encountered while carrying out a particular assay or diagnosing a disease. The need for highly precise identification tool is especially connected with the economic importance of the bacteria. In case of food and waterborne contaminants and diseases caused by pathogens rapid diagnosis for prescribing a treatment is necessary. Use of Lactobacillus to control canine intestinal infections is an interesting proposal [13]. *Lactobacillus* is an economically important organism, especially as bioactive molecules, probiotic, preservative, fermentation and maturation of the sausages, milk products [14–18]. The dominance of *Lactobacillus* species such as *L. iners, L. jensenii, L. gasseri,* and *L. crispatus* in the vaginal region acts as a barrier for invasion of pathogenic bacteria and virus. Lactobacillus is an indicator of a healthy vaginal ecosystem, hence their rapid identification can prove helpful in diagnosing bacterial vaginosis [19–21].

Molecular tools used for distinguishing *Lactobacillus* species involve: Random amplified polymorphic DNA, repetitive element PCR, restriction fragment length polymorphism, pulse-field gel electrophoresis, denaturing/temperature gradient gel electrophoresis [22–26]. These techniques have helped to reclassify certain species such as *L. brevis–L. hilgardii* [22]. Multiplex polymerase chain reaction of the region between *rrs* and 23S rRNA was developed to identify seven probiotic *Lactobacillus* species [27, 28]. Amplification of *rrs* gene and its analysis with the help of amplified ribosomal DNA restriction analysis, by employing restriction enzymes (REs) such as: (1) *HaeIII*, *DdeI*, and *HinfI*, (2) *ApaI*, *NotI* and *SmaI*, (3) *HaeIII*, *MspI*, and *HinfI* was used for identifying 17, 24, and 42 lactobacilli, respectively [26, 29, 30]. *In silico* restriction digestion patterns (with 11 REs) were identified for distinguishing *Lactobacillus* only up to the species level. REs—*SphI*, *NcoI* and *NheI* could digest *rrs*, REs—*DraI, EcoRI, HincII, SfuI, SspI*, and *VspI*, and showed digestion patterns in the *rrs*—23S rRNA intergenic region, where as REs—*AvrII* and *HindIII* digested 23S rRNA gene [28]. Matrix-assisted laser desorption/ionization mass spectrometry, based on ribosomal subunit proteins as biomarkers was used to identify *Lactobacillus plantarum* and other species [12, 31]. In order to distinguish very closely related *Lactobacillus* species, *rpoA* gene—encoding for the alpha subunit of RNA polymerase, and *pheS* gene—encoding for the alpha subunit of phenylalanyl-tRNA synthase were sequenced. These two genes provided 10 and 5 % divergence for distinguishing *Lactobacillus* species. The usage of the two genes in combination proved more effective in distinguishing the species in comparison to *rrs* gene [32]. *Lactobacillus plantarum* was found to possess genes—*mub, fbp* and *bsh*, which encode for mucus-binding protein, fibronectin-binding protein and bile salt hydrolase, respectively. These genes were projected as probiotic identification markers [15]. Multiplex PCR using species-specific primers for amplifying *recA* gene of the *L. plantarum* group enabled identification of strains to be *L. paraplantarum* and *L. pentosus* [33]. Comparative genomic hybridization (CGH) has been recognized as a powerful technique for identifying *Lactobacillus*. Genes—*recA*, *pheS*, *pyrG, tuf* and rrs were sequenced to identify strains initially annotated as analyzed to belong to *Lactobacillus taiwanensis*. However, DNA–DNA hybridization assays revealed the five strains to be distinct and more close to *Lactobacillus johnsonii* and *L. gasseri*. CGH with a whole genome DNA microarray of *L. johnsonii* strain NCC533 allowed to place *L. taiwanensis* BL263 independent of *L. johnsonii* ATCC 33200^{T} [34]. Evaluation of lactobacilli present in the cervix of the female genital tract was done initially using (GTG)(5)-PCR fingerprinting. The different clusters were then segregated based on *pheS* gene. This study revealed that although (GTG)(5)-PCR helps in identifying many isolates, however, a supplementary gene information is needed for reliable identification. An interesting finding of this study was the fact that *Lactobacillus acidophilus* is not a common inhabitant of female genital tract [35]. After an initial identification as *Lactobacillus* sp. DMDL 9010 done with *rrs*, the confirmatory test was performed using two fragments flanking the gene *L-ldh*1. The strain was identified as *Lactobacillus pentosus* or *L. plantarum* [36]. Molecular identification using *minD* gene, which encodes for an inhibitor cell division was carried out for *Lactobacillus rhamnosus* and *L. acidophilus.* The gene *minD* was observed to have high homology to that present in *Escherichia coli*, implying common evolution [37]. Horizontal gene transfer in *Lactobacillus* has been reported to be responsible for adapting to changes in lifestyles [38, 39]. Phylogenetic relationship was established through a multilocus sequence typing using house keeping genes, which included: *mutL, polA, ftsZ, pgm, metRS,* and *nrdD* [39, 40].

The genus *Lactobacillus* is composed of 180 species [17]. Sequenced genomes of *Lactobacillus* allow comparative studies: *L. acidophilus, L. johnsonii, L. plantarum, L.*

sakei, and *L. salivarius* [41]. Comparison of information deduced from sequenced genomes of *Lactobacillus ruminis* ATCC 27782 (2.06 Mb) isolated from bovine and *L. ruminis* strain, ATCC 25644 (2.138 Mb), isolated from human with *L. salivarius* core proteins showed that *Lactobacillus* species are categorized into 4 phylogenetically distinguishable groups [42]. Proteomics studies have proved helpful in identifying 6 proteins which can be used as biomarkers for selecting their probiotic potential since they are responsible for responding to bile salt and adaptation in *L. plantarum* [16]. Information on how *Lactobacillus casei* might have undergone diversity to adapt to the environment has been elucidated through completely sequencing its genome [39].

Lactobacillus have evolved from *Bacillus* and lost 600–1200 genes after the divergence. The loss of genes indicated that there has been a shift towards the nutrient rich atmosphere. The genomes of *Lactobacillus* species have sizes in the range of 1.8–3.3 Mb and a G+C content varying from 33 to 51 % [43]. Sequenced genomes of *Lactobacillus* species have revealed the evolution through gene degradation and horizontal gene transfer. Sequenced genomes also show that 55–60 % of proteins of *Lactobacillus bulgaricus* proteins show high homology with those present in *L. acidophilus* and *L. johnsonii*. *L. bulgaricus* has lost genes for mucin-binding proteins and bile salt hydrolase. *L. plantarum* has evolved to utilize diverse carbohydrates. *L. salivarius* has acquired genes for bile salt hydrolase and the pentose phosphate pathway. Comparative genomics done through Differential Blast Analysis (DBA), used for identifying specific genome regions, also elucidated information on regions which were present in some and absent in another set of organisms [18, 21, 43–50]. *Lactobacillus helveticus* DPC4571 show about 75 % homology with the genome of *L. acidophilus* NCFM [51]. This thus poses a major hurdle in identifying genes which may be common to all *Lactobacillus* species. Here, we have identified genes which are present in most of the *Lactobacillus* species. Of the large number of genes which can be potentially used for identifying novel biomarkers, a few were selected and subjected to *in silico* digestion with different type II restriction endonulcease (RE) enyzmes. RE digestion patterns unique to a given genome were identified.

Materials and Methods

Sequence Data and Comparative Genome Analysis

Genome sequences of the 50 strains of 19 species of the genus *Lactobacillus* were downloaded from NCBI (http://www.ncbi.nlm.nih.gov/): *L. acidophilus* (2 strains), *L. amylovorus* (2 strains), *L. brevis* (2 strains), *L. buchneri* (2 strains), *L. casei* (5 strains), *L. delbrueckii* (4 strains), *L. fermentum*, *L. gasseri*, *L. helveticus* (4 strains), *L. johnsonii* (3 strains), *L. kefiranofaciens*, *L. paracasei* (2 strains), *L. plantarum* (6 strains), *L. reuteri* (5 strains), *L. rhamnosus* (6 strains), *L. ruminis L. sakei, L. salivarius, L. sanfranciscensis* (Table S1). Genomes characteristics of the *Lactobacillus* have been presented in Table S1. Genes common to most of the *Lactobacillus* genomes were selected with the help of GenBank data (Table S2). From the common gene pool, 19 genes were selected varying in size from 732 to 3645 nucleotides (nts) (Tables S1 and S2). Gene, *rrs* was used as reference, as it is generally employed for bacterial identification. Sequence analysis and their orientation were done using BioEdit [8].

Restriction Endonuclease Analysis of Common Genes

Based on our previous works 10 Type II REs were considered for generating digestion patterns: (1) 4 base cutters *AluI*, *BfaI*, *BfuCI*, *CviAII*, *HpyCH4 V*, *RsaI*, *TaqI*, *Tru9I*, and (2) 6 base cutters *HaeI*, and *Hin1I* [8]. RE digestion patterns of the 19 common gene sequences along with *rrs* was obtained using Cleaver (http://cleaver.sourceforge.net/) (Table S2). REs producing 5–15 fragments were considered for further analysis using BioEdit [8]. Unique gene-RE combinations were formed the basis for identifying *Lactobacillus* species.

Results

Analysis of *rrs* Gene

Multiple Sequence Alignment

Sequenced genomes of the 50 strains belonging to 19 species of *Lactobacillus* strains had 4–9 copies of *rrs* gene. The multiple sequence alignment of a total of 268 *rrs* copies from all the 50 genomes allowed us to segregate them into 20 groups (Table S3). Six of these 20 groups were represented by 5–61 *rrs* copies (in all 173) belonging to 2–3 *Lactobacillus* species each. These observations indicate high sequence similarity among *rrs* copies of different species. *rrs* copies of all the strains of *L. delbrueckii* (4 strains), *L. reuteri* (5 strains), *L. buchneri* (2 strains) and *L. sanfranciscensis* (1 strain) could be segregated into 5 independent groups. *L. casei*, presented a unique scenario, where *rrs* copies of (1) 3 strains were grouped along with *L. paracasei* (2 strains), *L. rhamnosus* (6 strains) and *L. sakei,* and (2) the rest 5 *rrs* copies in each of two strains were all distinct from each other and formed 10 groups. In summary, only *L. sanfranciscensis* (1 strain)

could be segregated from all other genomes on the basis of its *rrs* gene sequences. Thus MSA proved that there is high level of intragenomic heterogeneity among different species of *Lactobacillus*.

In Silico RE Digestion

In silico RE digestion of all the 268 *rrs* copies belonging to 19 *Lactobacillus* species with 10 different REs revealed a few unique patterns (Table 1, Table S4). The only *rrs*-RE combinations, which could be used to distinguish different strains in an unambiguous manner were the following: (1) *L. acidophilus* La-14 with REs *AluI, BfuCI, HpyCH4 V*; (2) *L. acidophilus* NCFM and *L. casei* LC2 W with RE *AluI;* (3) *L. amylovorus* 30SC, *L. brevis* ATCC 367, and *L. brevis* KB290 with RE *HpyCH4 V*; (4) *L. buchneri* NRRL B-30929 with RE *CviAII*; (5) *L. casei* BL23, *L. casei* LOCK919, and *L. rhamnosus* ATCC 53103 with RE *BfuCI*; (6) *L. fermentum* IFO 3956 and *L. kefiranofaciens* ZW3 with REs *BfuCI* and *Tru9I*; (7) *L. gasseri* ATCC 33323 with RE *Tru9I*; (8) *L. johnsonii* DPC 6026, *L. johnsonii* N6.2 and *L. paracasei* ATCC 334 with RE *CviAII*; (9) *L. plantarum* subsp. plantarum P-8 with REs *AluI* and *Tru9I*; (10) *L. reuteri* TD1 with REs *BfuCI, CviAII,* and *HpyCH4 V;* (11) *L. reuteri* I5007 with RE *BfaI*; (12) *L. rhamnosus* ATCC 53103 with RE *BfuCI*; (13) *L. ruminis* ATCC 27782 with REs *AluI, CviAI, HpyCH4,*and *Tru9I*; (14) *L. sakei* subsp. *sakei* 23 K with REs *AluI, BfuCI, CviAII, HpyCH4 V* and *Tru9I;* (15) *L. salivarius* UCC118 with REs *BfaI*, *BfuCI, CviAII, HpyCH4 V, RsaI* and *Tru9I;* and (16) *L. sanfranciscensis* TMW 1.1304 with REs *BfaI*, *BfuCI, CviAII, HpyCH4 V, TaqI, RsaI* and *Tru9I.*

In summary, only 22 strains belonging to 17 species can be identified without any discrepancy: *L. acidophilus* (2); *L. amylovorus*; *L. brevis* (2); *L. buchneri*; *L. casei* (2); *L. fermentum*; *L. gasseri*; *L. johnsonii* (2); *L. kefiranofaciens*; *L. paracasei*; *L. plantarum*; *L. reuteri* (2); *L. rhamnosus*; *L. ruminis*; *L. sakei*; *L. salivarius*; and *L. sanfranciscensis.*

No unique RE digestion patterns with any of the REs employed were observed in the following strains: *L. amylovorus* GRL1118, *L. casei* str. Zhang, *L. delbrueckii* subsp. *bulgaricus* 2038, *L. helveticus* CNRZ32, *L. helveticus* DPC 4571, *L. helveticus* H10, *L. paracasei* subsp. *paracasei* 8700.2, *L. plantarum* JDM1, *L. plantarum* subsp. *plantarum* ST-III, *L. plantarum* ZJ316, *L. reuteri* DSM 20016, *L. rhamnosus* ATCC 8530, *L. rhamnosus* GG ATCC 53103, *L. rhamnosus* Lc 705, *L. rhamnosus* LOCK900, and *L. rhamnosus* LOCK908.

It may be concluded that with MSA and RE digestions only 22 strains can be identified. For the rest of the 28 strains we may need to resort to other genes.

Analysis of Common Genes

In Silico RE Digestion Analysis of Common Genes

A list of genes chosen for study from the common gene pool from sequenced genomes of *Lactobacillus* strains has been presented in Table S2. Of the 19 selected genes (732–3645 nts in length) only 9 genes (*cysS, dnaA, dnaJ, dnaK, gyrB, polA, pyrB, pyrG,* and *recA*) were present in all the 47–50 genomes, where as the rest of the 10 genes were present in 44–46 genomes. All the 19 genes were present in single copies. *In silico* RE digestion of all the 19 genes present in *Lactobacillus* species with 10 different REs revealed a few unique digestion patterns.

Analysis of recA Gene

Among the genes which were considered as suitable for identifying *Lactobacillus* strains, RE digestion patterns of *recA* (1098 nts) have been presented in Tables 2 and 3. *In silico* RE digestion of *recA*, revealed a few unique patterns in 29 strains representing 17 species. Unique RE digestion patterns could not be identified for *L. amylovorus* (2 strains) and *L. johnsonii* (3 strains). RE—*HpyCH4 V,* proved to be the most effective by generating unique digestion patterns in 19 strains belonging to 14 species. On the other hand, REs—*CviAII, BfuCI* and *RsaI* were found to be useful in identifying 12–15 strains representing 12–14 species each (Tables 2, 3). Hence, through the use of complementary REs, most of the strains can be identified in an unambiguous manner.

Analysis of Other Common Genes

There were 21 strains, which could not be distinguished on the basis of RE digestion pattern of *recA* gene. Analysis of digestion patterns of genes—*ruvB* (1020 nts), *dnaA* (1368 nts), *purA* (1290 nts), *dnaJ* (1140 nts), and *gyrB* (1944 nts) in combination with REs—*AluI, BfuCI, CviAI, Taq1,*and *Tru9I* proved instrumental in providing information for clear identification of 16 strains belonging to 8 species.

Gene *ruvB* in combination with different REs was effective in segregating different phylogenetically close strains (Table 4): (1) RE *AluI* helped to segregate *L. amylovorus* 30SC, *L. amylovorus* GRL1118, *L. johnsonii* DPC 6026, *L. johnsonii* N6, and *L. johnsonii* NCC 533, (2) RE *BfuCI* could distinguish *L. paracasei* subsp. *paracasei* 8700.2 from *L. plantarum* 16, and (3) RE TaqI provided unique digestion pattern for *L. rhamnosus* Lc 705. In addition, the following unique gene—RE combinations can also be used in the following cases: (1) *dnaA- AluI* for *L. delbrueckii* subsp. *bulgaricus* 2038; (2) *purA- Tru9I* for *L.*

Table 1 *In silico* restriction endonuclease digestion patterns (5′–3′) of *rrs* gene of *Lactobacillus* strains

Lactobacillus spp.	GenBank ID	Copies of *rrs* (U/T)[a]	Unique restriction patterns[b]
AluI			
L. plantarum WCFS1	AL935263	2/5	222·51·615·105·102·207·269
		3/5[c]	273·615·105·102·207·269
L. fermentum IFO 3956	AP008937	5	273·186·429·105·102·473
L. acidophilus NCFM	CP000033	4	69·146·20·33·186·429·207·221·44·217
L. salivarius UCC118	CP000233	7	843·207·207·248
L. casei LC2 W	CP002616	1/5	1093·338·20·117
		1/5	307·129·41·245·278·40·316·132·80
		1/5	459·139·39·708·223
		1/5	13·136·344·898·17·160
		1/5[c]	25·1222·321
L. casei BD-II	CP002618	1/5	499·218·793·58
		1/5	268·286·41·55·204·21·24·544·125
		1/5	452·414·235·15·452
		1/5	114·42·53·366·19·92·131·45·364·342
		1/5	140·25·383·452·133·361·74
L. reuteri SD2112	CP002844	2/6	216·57·186·429·105·102·193·14·262
		4/6[c]	273·186·429·105·102·193·14·262
L. ruminis ATCC 27782	CP003032	6	54·37·775·207·207·247
L. helveticus R0052	CP003799	1/4	215·20·33·186·429·206·220·44·217
		1/4	215·20·33·186·430·208·221·44·217
		2/4	205·20·33·186·428·207·221·44·217
L. acidophilus La-14	CP005926	4	63·146·20·33·186·429·207·221·44·210
L. plantarum subsp. plantarum P-8	CP005942	5	63·201·615·105·102·207·262
L. reuteri TD1	CP006603	1/6	281·186·184·245·105·102·193·14·287
		5/6	281·615·105·102·193·14·287
L. sakei subsp. *sakei* 23 K	CR936503	3/7	85·801·160·47·207·271
		4/7	665·160·47·207·271
BfaI			
L. salivarius UCC118	CP000233	7	229·578·518·180
L. casei BD-II	CP002618	1/5	570·268·581·149
		4/5	Not segregated
L. delbrueckii subsp. *bulgaricus* NDO2	CP002341	2/9	104·159·578·323·195·201
		7/9[c]	104·159·578·185·333·202
L. sanfranciscensis TMW 1.1304	CP002461	7	242·33·578·323·195·199
L. reuteri I5007	CP006011	5/6	274·245·333·296·27·195·194
		1/6[c]	274·578·296·27·195·194
BfuCI			
L. reuteri JCM 1112	AP007281	6	7·98·225·892·159·153
L. rhamnosus ATCC 53103	AP011548	5	15·315·699·12·340·177·16
L. salivarius UCC118	CP000233	2/7	162·123·119·1101
		5/7	174·8·115·119·1101
L. sanfranciscensis TMW 1.1304	CP002461	2/7	7·98·1277·175·13
		5/7	7·98·1277·176·13
L. johnsonii DPC 6026	CP002464	1/4	7·319·974·77·176·15
		3/4[c]	45·319·974·77·176·60
L. kefiranofaciens ZW3	CP002764	4	16·312·1051·174·21

Table 1 continued

Lactobacillus spp.	GenBank ID	Copies of *rrs* (U/T)[a]	Unique restriction patterns[b]
L. helveticus R0052	CP003799	1/4	13·188·8·235·934·174·22
		3/4	13·188·8·235·930·174·22
L. casei LOCK919	CP005486	5	16·67·116·132·699·12·340·167
L. acidophilus La-14	CP005926	4	7·188·8·116·119·932·174·15
L. reuteri TD1	CP006603	6	15·98·225·892·159·175·33
L. sakei subsp. *sakei* 23 K	CR936503	3/7	8·439·46·886·176·16
		4/7	226·46·886·176·16
L. casei BL23	FM177140	1/4	15·183·132·699·12·340·177·21
		3/4[c]	15·67·116·132·699·12·340·177·21
CviAII			
L. salivarius UCC118	CP000233	7	15·153·493·269·106·148·125·34·50·112
L. delbrueckii subsp. *bulgaricus* ATCC BAA-365	CP000412	8/9	49·143·9·494·126·143·106·148·209·134
		1/9[c]	49·143·9·494·90·36·143·106·148·209·134
L. paracasei ATCC 334	CP000423	5	50·53·104·494·90·36·143·106·148·209·135
L. sanfranciscensis TMW 1.1304	CP002461	7	47·164·496·90·36·143·106·148·125·34·50·131
L. johnsonii DPC 6026	CP002464	1/4	47·162·493·90·36·143·106·148·209·134
L. buchneri NRRL B-30929	CP002652	5	577·38·113·102·106·148·108·95·212·22·47
L. reuteri SD2112	CP002844	2/6	47·659·90·36·143·106·148·125·34·5·45·125
L. ruminis ATCC 27782	CP003032	3/6	37·142·11·9·485·269·106·148·125·34·50·111
		3/6	39·142·11·267·227·269·106·148·125·34·50·109
L. reuteri TD1	CP006603	6	55·165·494·90·36·143·106·148·159·50·151
L. johnsonii N6.2	CP006811	4	53·162·493·90·36·143·106·148·209·140
L. sakei subsp. *sakei* 23 K	CR936503	3/7	48·162·494·90·36·143·106·148·125·34·50·135
		4/7	483·90·36·143·106·148·125·34·50·135
HpyCH4 V			
L. brevis KB290	AP012167	5	58·35·565·25·201·195·262·234
L. salivarius UCC118	CP000233	1/7	40·35·562·25·396·262·208
		6/7	18·35·561·25·396·262·208
L. brevis ATCC 367	CP000416	5	52·35·565·25·201·195·262·228
L. sanfranciscensis TMW 1.1304	CP002461	7	50·39·571·25·396·255·7·11·216
L. amylovorus 30SC	CP002559	4	59·206·43·349·25·88·113·108·37·50·255·7·11·96·128
L. ruminis ATCC 27782	CP003032	6	40·35·562·25·201·195·255·7·11·196
L. helveticus R0052	CP003799	2/4	46·206·43·349·24·88·113·195·255·7·11·96·128
		2/4[c]	56·206·43·350·25·88·113·196·255·7·11·96·128
L. acidophilus La-14	CP005926	4	50·206·43·349·25·88·113·145·50·255·7·11·96·121
L. reuteri TD1	CP006603	6	58·39·178·43·349·25·88·308·262·247
L. sakei subsp. *sakei* 23 K	CR936503	3/7	51·214·43·349·25·396·262·231
		4/7	44·43·349·25·396·262·231
RsaI			
L. salivarius UCC118	CP000233	1/7	896·253·102·146·131
		6/7	873·355·146·131
TaqI			
L. sanfranciscensis TMW 1.1304	CP002461	6/7	55·735·199·359·222
		1/7[c]	55·734·199·575
L. helveticus R0052	CP003799	4	51·722·199·359·230
L. plantarum 16	CP006033	1/5	65·725·159·41·582
		4/5[c]	64·724·199·583

Table 1 continued

Lactobacillus spp.	GenBank ID	Copies of *rrs* (U/T)[a]	Unique restriction patterns[b]
Tru9I			
L. fermentum IFO 3956	AP008937	5	223·245·22·25·130
L. salivarius UCC118	CP000233	7	225·349·26·338·134·44·150·239
L. gasseri ATCC 33323	CP000413	6	232·672·86·134·44·411
L. sanfranciscensis TMW 1.1304	CP002461	7	620·26·252·86·134·44·150·258
L. buchneri NRRL B-30929	CP002652	4/5	102·122·421·252·86·134·194·252
		1/5[c]	70·32·122·421·252·86·134·194·252
L. kefiranofaciens ZW3	CP002764	4	153·69·421·252·86·134·44·415
L. reuteri SD2112	CP002844	1/6	224·421·252·87·134·44·150·253
		5/6[c]	645·252·86·134·44·150·253
L. ruminis ATCC 27782	CP003032	6	248·349·26·252·86·134·44·150·238
L. buchneri CD034	CP003043	3/5	70·32·122·421·252·86·53·81·194·252
		2/5[c]	70·32·122·421·252·86·134·194·252
L. helveticus R0052	CP003799	1/4	640·252·85·177·416
		3/4[c]	630·251·86·134·44·416
L. plantarum subsp. plantarum P-8	CP005942	5	506·104·278·86·134·44·150·253
L. plantarum 16	CP006033	1/5	1·513·104·278·86·134·44·150·261
		4/5[c]	3·466·47·104·278·86·134·44·150·259
L. sakei subsp. *sakei* 23 K	CR936503	3/7	617·278·86·134·194·262
		4/7	396·278·86·134·44·150·262

Symbol (·) indicates RE site in the gene sequences

[a] Unique/total

[b] Values represent restriction fragments (nucleotides)

[c] This pattern is not unique. It has been presented to indicate the RE digestion pattern of the rest of the *rrs* copies

plantarum subsp. *plantarum* ST-III, and *purA-AluI* for *L. plantarum* JDM1; (3) *dnaJ- AluI* for *L. casei* LOCK919; *dnaJ- BfuCI* for *L. delbrueckii* subsp. *bulgaricus* ATCC BAA-365; *dnaJ- CviAII* for *L. plantarum* WCFS1; and (4) *gyrB- TaqI* for *L. reuteri* DSM 20016; *gyrB- AluI* for *L. reuteri JCM* 1112.

With the available set of gene-RE combinations in this study, 4 strains could not be distinguished. Here, we may have to resort to additional gene-RE combinations.

Discussion

Bacterial identification with the help of *rrs* gene is practiced around the world. However, in cases where bacterial genomes contain multiple copies of *rrs*, such as in *Clostridium, Staphylococcus, Streptococcus, Vibrio,* and *Yersinia* species, a high level of heterogeneity is seen. It is difficult to identify these bacterial species on the basis of their *rrs* gene alone [5–10]. The same problem has been faced in the case of *Lactobacillus* as well. A large number of functional genes have been employed for identifying *Lactobacillus* sp., however, no consensus has been reached as yet. Hence, we identified genes which were common to almost all the *Lactobacillus* species. Of the 19 common genes which were processed for RE digestion patterns, *pheS, polA, pyrG, recA,* and *rpoA* have been reported in literature as markers for *Lactobacillus* [32–34, 39, 40]. Of these five genes, we could use only *recA* with effectiveness; where as the other 4 genes didn't prove helpful, because they generated a large number of fragments on digestion with REs. In our case, the following additional genes could be used to distinguish very closely related species: *ruvB* (1020 nts), *dnaJ* (1140 nts), *purA* (1290 nts), *dnaA* (1368 nts), and *gyrB* (1944 nts) in combination with REs—*AluI, BfuCI, CviAI, Taq1,*and *Tru9I*. Our comparison of *recA* from *Lactobacillus* and *Staphylococcus* gave very distinct RE digestion patterns, allowing easy and reliable distinction [8]. It may be proposed here that the biomarkers identified in this study

Table 2 Unique *in silico* restriction endonuclease digestion patterns (5′-3′) of *recA* gene of *Lactobacillus* strains

Lactobacillus spp.	Restriction endonucleases			
	CviAII	*BfuCI*	*HpyCH4 V*	*RsaI*
L. acidophilus La-14	299·255·226·148·78·42·38·6	565·297·120·87·23	–	–
L. acidophilus NCFM	299·255·226·148·78·42·38·11·6	565·308·120·87·23	–	–
L. brevis ATCC 367	–[a]	–	639·228·120·79·68	555·215·132·116·116
L. brevis KB290	–	633·315·207·9	639·258·120·79·68	555·215·146·132·116
L. buchneri CD034	–	–	389·256·162·145·132·56	–
L. buchneri NRRL B-30929	–	–	645·162·145·132·56	–
L. casei BL23	–	302·297·250·90·87·36	–	–
L. casei str. Zhang	299·255·228·200·42·38	–	568·249·111·83·33·18	564·260·195·43
L. delbrueckii subsp. *bulgaricus* ATCC 11842	309·216·183·168·78·42	186·162·108·107·106·90·72·72·54·39	618·231·129·18	790·81·74·51
L. delbrueckii subsp. *bulgaricus* NDO2	–	220·186·162·108·107·90·72·72·54·39	–	–
L. fermentum IFO 3956	708·159·120·75	265·223·165·150·98·90·65·6	–	758·163·114·27
L. gasseri ATCC 33323	–	392·297·210·185·11	686·241·107·30·24·7	449·341·156·127·22
L. helveticus CNRZ32	–	–	–	–
L. helveticus DPC 4571	–	–	–	–
L. helveticus H10	299·232·232·231·80·24	–	–	–
L. helveticus R0052	–	–	–	331·276·248·227·16
L. kefiranofaciens ZW3	379·255·244·232	380·348·202·132·24·24	612·237·86·41·34·33·31·21·15	250·239·185·129·119·91·81·16
L. paracasei ATCC 334	–	–	564·206·111·83·43·37·18	–
L. plantarum subsp. plantarum P-8	420·411·155·82·75	366·306·174·115·93·63·26	539·469·126·9	–
L. plantarum ZJ316	–	–	–	–
L. reuteri I5007	–	–	–	–
L. reuteri SD2112	379·255·229·147·79	–	426·232·116·111·75·75·48·6	–
L. reuteri TD1	–	546·250·245·48	537·232·116·81·75·48	–
L. rhamnosus ATCC 53103	433·359·144·93·24	250·237·166·111·88·72·31·28·27·27·16	296·242·228·202·52·33	296·244·190·156·91·76
L. rhamnosus GG ATCC 53103	354·232·180·115·68·51·38·11·4	–	299·270·165·136·84·83·12·4	–
L. rhamnosus LOCK900	–	–	–	–
L. ruminis ATCC 27782	371·229·183·159·147	392·392·294·11	452·223·188·142·42·33·5·4	–
L. sakei subsp. *sakei* 23 K	554·285·252·45	633·201·118·109·75	808·259·61·8	467·341·124·112·81·11
L. salivarius UCC118	299·262·255·182·148	819·190·107·30	562·111·93·89·81·69·64·56·12·9	467·294·185·109·91
L. sanfranciscensis TMW 1.1304	489·226·216·99·47·36	–	452·231·136·126·126·42	607·221·198·87

Symbol (·) indicates RE site in the gene sequences

[a] No unique pattern was observed

Table 3 Unique *in silico* restriction endonuclease digestion patterns (5′-3′) of *recA* gene of *Lactobacillus* strains

Lactobacillus spp.	Restriction endonucleases			
	AluI	*TaqI*	*Tru9I*	*BfaI*
L. acidophilus La-14	–[a]	–	–	–
L. acidophilus NCFM	–	–	–	–
L. brevis ATCC 367	–	720·313·70·31	366·323·192·181·48·24	–
L. brevis KB290	–	720·313·100·31	366·337·192·181·48·24·16	–
L. buchneri CD034	–	501·453·159·22·5	535·292·117·87·48·37·24	–
L. buchneri NRRL B-30929	453·432·170·63·22	501·258·195·159·22·5	535·379·117·48·37·24	–
L. casei BL23	–	–	–	–
L. casei str. Zhang	–	580·242·181·59	504·381·110·67	–
L. delbrueckii subsp. *bulgaricus* ATCC 11842	290·241·241·224	315·166·136·98·81·72·54·36·15·14·9	–	–
L. delbrueckii subsp. *bulgaricus* NDO2	–	–	–	–
L. fermentum IFO 3956	–	360·215·205·117·70·51·18·14·12	–	–
L. gasseri ATCC 33323	425·138·136·132·130·98·36	–	300·244·168·87·82·68·67·40·24·15	–
L. helveticus CNRZ32	–	–	508·334·87·69·67·33	–
L. helveticus DPC 4571	270·261·144·144·97·95·87	–	–	–
L. helveticus H10	–	–	–	–
L. helveticus R0052	–	454·435·159·50	–	–
L. kefiranofaciens ZW3	366·330·162·136·90·18·8	784·228·50·48	353·259·184·114·111·41·33·15	–
L. paracasei ATCC 334	–	–	–	–
L. plantarum subsp. plantarum P-8	353·256·216·108·72·47·39·28·24	–	308·270·258·153·53·43·33·21·4	–
L. plantarum ZJ316	–	–	–	725·267·102·25·24
L. reuteri I5007	368·268·231·193·29	–	–	–
L. reuteri SD2112	–	567·247·228·47	–	–
L. reuteri TD1	–	–	–	–
L. rhamnosus ATCC 53103	466·166·129·119·100·73	370·266·235·103·40·27·12	383·240·216·95·70·49	541·422·73·17
L. rhamnosus GG ATCC 53103	–	–	535·289·165·40·24	–
L. rhamnosus LOCK900	478·339·133·54·49	824·124·56·49	–	–
L. ruminis ATCC 27782	491·249·174·94·81	437·227·220·109·96	–	–
L. sakei subsp. *sakei* 23 K	–	372·298·172·150·108·36	352·329·219·165·30·24·14·3	–
L. salivarius UCC118	420·344·207·63·55·45·12	–	535·225·160·117·49·39·21	–
L. sanfranciscensis TMW 1.1304	903·74·46·43·42·5	266·206·193·157·123·118·50	344·184·174·156·138·61·43·10·3	–

Additional unique patterns were also recorded for *L. fermentum* IFO 3956 - *Hin1I* - 552·189·124·108·89

Symbol (·) indicates RE site in the gene sequences

[a] No unique pattern was observed

Table 4 Unique *in silico* restriction endonuclease digestion patterns (5′-3′) of common genes of *Lactobacillus* strains (other than *recA*)

Lactobacillus spp.	Gene	RE	RE digestion patterns
L. amylovorus 30SC	*ruvB*	*AluI*	423·252·220·92·30
L. amylovorus GRL1118			408·252·220·92·30·15
L. johnsonii DPC 6026			561·169·141·108·41
L. johnsonii N6			366·195·169·141·108·41
L. johnsonii NCC 533			561·169·108·80·61·41
L. paracasei subsp. *paracasei* 8700.2		*BfuCI*	351·185·128·93·88·63·56·24·18·11
L. plantarum 16			546·262·169·34
L. rhamnosus Lc 705		*Taq1*	344·271·184·171·50·27
L. delbrueckii subsp. *bulgaricus* 2038	*dnaA*	*AluI*	29·386·228·320·96·150·156
L. plantarum subsp. *plantarum* ST-III	*purA*	*Tru9I*	348·216·173·134·106·86·73·69·52·33
L. plantarum JDM1		*AluI*	574·297·197·196·26
L. casei LOCK919	*dnaJ*	*AluI*	84·268·66·330·109·287·20
L. delbrueckii subsp. *bulgaricus* ATCC BAA-365		*BfuCI*	491·66·147·52·110·141·130
L. plantarum WCFS1		*CviAII*	112·408·78·545
L. reuteri DSM 20016	*gyrB*	*TaqI*	776·461·294·234·190·70
L. reuteri JCM 1112		*AluI*	1041·391·280·238

Symbol (·) indicates RE site in the gene sequences

can be used for developing rapid and reliable protocol for diagnostic purposes.

Acknowledgments We are thankful to the Director of CSIR-Institute of Genomics and Integrative Biology (IGIB), and CSIR projects GENESIS (BSC0121) and INDEPTH (BSC0111) for providing the necessary funds, facilities and moral support. Authors are also thankful to Academy of Scientific & Innovative Research (AcSIR), New Delhi.

References

1. Porwal S, Lal S, Cheema S, Kalia VC (2009) Phylogeny in aid of the present and novel microbial lineages: diversity in *Bacillus*. PLoS One 4:e4438. doi:10.1371/journal.pone.0004438
2. Kalia VC, Mukherjee T, Bhushan A, Joshi J, Shankar P, Huma N (2011) Analysis of the unexplored features of *rrs* (16S rDNA) of the genus *Clostridium*. BMC Genomics 12:18. doi:10.1186/1471-2164-12-18
3. Bhushan A, Joshi J, Shankar P, Kushwah J, Raju SC, Purohit HJ, Kalia VC (2013) Development of genomic tools for the identification of certain *Pseudomonas* up to species level. Indian J Microbiol 53:253–263. doi:10.1007/s12088-013-0412-1
4. Kalia VC (2015) Let's explore the latent features of genes to identify bacteria. J Mol Genet Med 9:e105. doi:10.4172/1747-0862.1000E105
5. Kekre A, Bhushan A, Kumar P, Kalia VC (2015) Genome wide analysis for searching novel markers to rapidly identify *Clostridium* strains. Indian J Microbiol 55:250–257. doi:10.1007/s12088-015-0535-7
6. Kalia VC, Kumar P (2015) Genome wide search for biomarkers to diagnose *Yersinia* infections. Indian J Microbiol 55:366–374. doi:10.1007/s12088-015-0552-6
7. Koul S, Kumar P, Kalia VC (2015) A unique genome wide approach to search novel markers for rapid identification of bacterial pathogens. J Mol Genet Med 9:194. doi:10.4172/1747-0862.1000194
8. Kumar R, Koul S, Kumar P, Kalia VC (2016) Searching biomarkers in the sequenced genomes of *Staphylococcus* for their rapid identification. Indian J Microbiol 56:64–71. doi:10.1007/s12088-016-0565-9
9. Kalia VC, Kumar P, Kumar R, Mishra A, Koul S (2015) Genome wide analysis for rapid identification of *Vibrio* species. Indian J Microbiol 55:375–383. doi:10.1007/s12088-015-0553-5
10. Kalia VC, Kumar R, Kumar P, Koul S (2016) A genome-wide profiling strategy as an aid for searching unique identification biomarkers for *Streptococcus*. Indian J Microbiol 56:46–58. doi:10.1007/s12088-015-0561-5
11. Liu Q, Wang S, Zhi J-F, Ming H, Teng D (2013) Efficient production of lactic acid from sweet sorghum juice by a newly isolated *Lactobacillus salivarius* CGMCC 7.75. Indian J Microbiol 53:332–336. doi:10.1007/s12088-013-0377-0
12. Anderson AC, Sanunu M, Schneider C, Clad A, Karygianni L, Hellwig E, Al-Ahmad A (2014) Rapid species-level identification of vaginal and oral lactobacilli using MALDI-TOF MS analysis and 16S rDNA sequencing. BMC Microbiol 14:312. doi:10.1186/s12866-014-0312-5
13. McCoy S, Gilliland SE (2007) Isolation and characterization of *Lactobacillus* species having potential for use as probiotic cultures for dogs. J Food Sci 72:M94–M97
14. Kaushik JK, Kumar A, Duary RK, Mohanty AK, Grover S, Batish VK (2009) Functional and probiotic attributes of an indigenous isolate of *Lactobacillus plantarum*. PLoS One 4:e8099. doi:10.1371/journal.pone.0008099
15. Hamon E, Horvatovich P, Izquierdo E, Bringel F, Marchioni E, Aoudé-Werner D, Ennahar S (2011) Comparative proteomic analysis of *Lactobacillus plantarum* for the identification of key proteins in bile tolerance. BMC Microbiol 11:63. doi:10.1186/1471-2180-11-63
16. Herbel SR, Vahjen W, Wieler LH, Guenther S (2013) Timely approaches to identify probiotic species of the genus *Lactobacillus*. Gut Pathog 5:27–40. doi:10.1186/1757-4749-5-27
17. Drissi F, Merhej V, Angelakis E, El Kaoutari A, Carrière F, Henrissat B, Raoult D (2014) Comparative genomics analysis of

Lactobacillus species associated with weight gain or weight protection. Nutr Diabetes 4:e109. doi:10.1038/nutd.2014.6
18. Moroeanu VI, Vamanu E, Paun G, Neagu E, Ungureanu OR, Eremia SAV, Radu GL, Ionescu R, Pelinescu DR (2015) Probiotic strains influence on infant microbiota in the in vitro colonic fermentation model GIS1. Indian J Microbiol 55:423–429. doi:10.1007/s12088-015-0542-8
19. Douillard FP, de Vos WM (2014) Functional genomics of lactic acid bacteria: from food to health. Microb Cell Fact 13:S8. doi:10.1186/1475-2859-13-S1-S8
20. Mendes-Soares H, Suzuki H, Hickey RJ, Forney LJ (2014) Comparative functional genomics of *Lactobacillus* spp. reveals possible mechanisms for specialization of vaginal lactobacilli to their environment. J Bacteriol 196:1458–1470. doi:10.1128/JB.01439-13
21. Petrova MI, Lievens E, Malik S, Imholz N, Lebeer S (2015) *Lactobacillus* species as biomarkers and agents that can promote various aspects of vaginal health. Front Physiol 6:81. doi:10.3389/fphys.2015.00081
22. Sohier D, Coulon J, Lonvaud-Funel A (1999) Molecular identification of *Lactobacillus hilgardii* and genetic relatedness with *Lactobacillus brevis*. Int J Syst Bacteriol 49:1075–1081
23. Rantsiou K, Drosinos EH, Gialitaki M, Urso R, Krommer J, Gasparik-Reichardt J, Toth S, Metaxopoulos I, Comi G, Cocolin L (2005) Molecular characterization of *Lactobacillus* species isolated from naturally fermented sausages produced in Greece, Hungary and Italy. Food Microbiol 22:19–28. doi:10.1016/j.fm.2004.05.001
24. Weiss A, Lettner HP, Kramer W, Mayer HK, Kneifel W (2005) Molecular methods used for the identification of potentially probiotic *Lactobacillus reuteri* strains. Food Technol Biotechnol 43:295–300
25. Singh S, Goswami P, Singh R, Heller KJ (2009) Application of molecular identification tools for *Lactobacillus*, with a focus on discrimination between closely related species: a review. LWT Food Sci Technol 42:448–457. doi:10.1016/j.lwt.2008.05.019
26. Markiewicz LH, Biedrzycka E, Wasilewska E, Bielecka M (2010) Rapid molecular identification and characteristics of *Lactobacillus* strains. Folia Microbiol (Praha) 55:481–488. doi:10.1007/s12223-010-0080-z
27. Kwon HS, Yang EH, Yeon SW, Kang BH, Kim TY (2004) Rapid identification of probiotic *Lactobacillus* species by multiplex PCR using species-specific primers based on the region extending from 16S rRNA through 23S rRNA. FEMS Microbiol Lett 239:267–275. doi:10.1016/j.femsle.2004.08.049
28. Moreira JLS, Mota RM, Horta MF, Teixeira SMR, Neumann E, Nicoli JR, Nunes AC (2005) Identification to the species level of *Lactobacillus* isolated in probiotic prospecting studies of human, animal or food origin by 16S-23S rRNA restriction profiling. BMC Microbiol 5:15. doi:10.1186/1471-2180-5-15
29. Delfederico L, Hollmann A, Martínez M, Iglesias NG, De Antoni G, Semorile L (2005) Molecular identification and typing of lactobacilli isolated from kefir grains. J Dairy Res 73:20–27. doi:10.1017/S00022029905001408
30. Soto LP, Frizzo LS, Bertozzi E, Avataneo E, Sequeira GJ, Rosmini MR (2010) Molecular microbial analysis of *Lactobacillus* strains isolated from the gut of calves for potential probiotic use. Vet Med Int 274987:7. doi:10.4061/2010/274987
31. Sun L, Teramoto K, Sato H, Torimura M, Tao H, Shintani T (2006) Characterization of ribosomal proteins as biomarkers for matrix-assisted laser desorption/ionization mass spectral identification of *Lactobacillus plantarum*. Rapid Commun Mass Spectrom 20:3789–3798. doi:10.1002/rcm.2801
32. Naser SM, Dawyndt P, Hoste B, Gevers D, Vandemeulebroecke K, Cleenwerck I, Vancanneyt M, Swings J (2007) Identification of lactobacilli by *pheS* and *rpoA* gene sequence analysis. Int J Syst Evol Microbiol 57:2777–2789. doi:10.1099/ijs.0.64711-0
33. Kingston JJ, Radhika M, Roshini PT, Raksha MA, Raksha HS, Batra HV (2010) Molecular characterization of lactic acid bacteria recovered from natural fermentation of beet root and carrot Kanji. Indian J Microbiol 50:292–298. doi:10.1007/s12088-010-0022-0
34. Sarmiento-Rubiano LA, Berger B, Moine D, Zúñiga M, Pérez-Martínez G, Yebra MJ (2010) Characterization of a novel *Lactobacillus* species closely related to *Lactobacillus johnsonii* using a combination of molecular and comparative genomics methods. BMC Genomics 11:504. doi:10.1186/1471-2164-11-504
35. Švec P, Sedláček I, Chrápavá M, Vandamme P (2011) (GTG)(5)-PCR fingerprinting of lactobacilli isolated from cervix of healthy women. Folia Microbiol 56:80–83. doi:10.1007/s12223-011-0006-4
36. Y-t Fei, D-m Liu, T-h Luo, Chen G, Wu H, Li L, Y-g Yu (2014) Molecular characterization of *Lactobacillus plantarum* DMDL 9010, a strain with efficient nitrite degradation capacity. PLoS One 9:e113792. doi:10.1371/journal.pone.0113792
37. Nguyen THK, Doan VTT, Ha LD, Nguyen HN (2013) Molecular cloning, expression of *minD* gene from *Lactobacillus acidophilus* VTCC-B-871 and analyses to identify *Lactobacillus rhamnosus* PN04 from Vietnam *Hottuynia cordata* Thunb. Indian J Microbiol 53:385–390. doi:10.1007/s12088-013-0384-1
38. Cai H, Thompson R, Budinich MF, Broadbent JR, Steele JL (2009) Genome sequence and comparative genome analysis of *Lactobacillus casei*: insights into their niche-associated evolution. Genome Biol Evol 1:239–257. doi:10.1093/gbe/evp019
39. Yu S, Peng Y, Zheng Y, Chen W (2015) Comparative genome analysis of *Lactobacillus casei*: insights into genomic diversification for niche expansion. Indian J Microbiol 55:102–107. doi:10.1007/s12088-014-0496-2
40. Cai H, Rodriguez BT, Zhang W, Broadbent JR, Steele JL (2007) Genotypic and phenotypic characterization of *Lactobacillus casei* strains isolated from different ecological niches suggests frequent recombination and niche specificity. Microbiology 153:2655–2665. doi:10.1099/mic.0.2007/006452-0
41. Canchaya C, Claesson MJ, Fitzgerald GF, van Sinderen D, O'Toole PW (2006) Diversity of the genus *Lactobacillus* revealed by comparative genomics of five species. Microbiology 152:3185–3196. doi:10.1099/mic.0.29140-0
42. Forde BM, Neville BA, O'Donnell MM, Riboulet-Bisson E, Claesson MJ, Coghlan A, Ross RP, O'Toole PW (2011) Genome sequences and comparative genomics of two *Lactobacillus ruminis* strains from the bovine and human intestinal tracts. Microb Cell Fact 10:S13. doi:10.1186/1475-2859-10-S1-S13
43. Kant R, Blom J, Palva A, Siezen RJ, de Vos WM (2011) Comparative genomics of *Lactobacillus*. Microb Biotechnol 4:323–332. doi:10.1111/j.1751-7915.2010.00215.x
44. Klaenhammer TR, Altermann E, Pfeiler E, Buck BL, Goh YJ, O'Flaherty S, Barrangou R, Duong T (2008) Functional genomics of probiotic lactobacilli. J Clin Gastroenterol 42:S160–S162. doi:10.1097/MCG.0b013e31817da140
45. Douillard FP, Ribbera A, , Kant R, Pietilä TE, Järvinen HM, Messing M, Randazzo CL, Paulin L, Laine P, Ritari J, Caggia C, Lähteinen T, Brouns SJ, Satokari R, von Ossowski I, Reunanen J, Palva A, de Vos WM (2013) Comparative genomic and functional analysis of 100 *Lactobacillus rhamnosus* strains and their comparison with strain GG. PLoS Genet 9:e1003683. doi:10.1371/journal.pgen.1003683
46. Douillard FP, Kant R, Ritari J, Paulin L, Palva A, de Vos WM (2013) Comparative genome analysis of *Lactobacillus casei* strains isolated from Actimel and Yakult products reveals marked similarities and points to a common origin. Microb Biotechnol 6:576–587. doi:10.1111/1751-7915.12062

47. Nadkarni MA, Chen Z, Wilkins MR, Hunter N (2014) Comparative genome analysis of *Lactobacillus rhamnosus* clinical isolates from initial stages of dental pulp infection: identification of a new exopolysaccharide cluster. PLoS One 9:e90643. doi:10.1371/journal.pone.0090643
48. Ojala T, Kankainen M, Castro J, Cerca N, Edelman S, Westerlund-Wikström B, Paulin L, Holm L, Auvinen P (2014) Comparative genomics of *Lactobacillus crispatus* suggests novel mechanisms for the competitive exclusion of *Gardnerella vaginalis*. BMC Genomics 15:1070. doi:10.1186/1471-2164-15-1070
49. Raftis EJ, Forde BM, Claesson MJ, O'Toole PW (2014) Unusual genome complexity in *Lactobacillus salivarius* JCM1046. BMC Genomics 15:771. doi:10.1186/1471-2164-15-771
50. Illeghems K, De Vuyst L, Weckx S (2015) Comparative genome analysis of the candidate functional starter culture strains *Lactobacillus fermentum* 222 and *Lactobacillus plantarum* 80 for controlled cocoa bean fermentation processes. BMC Genomics 16:766. doi:10.1186/s12864-015-1927-0
51. O'Sullivan O, O'Callaghan J, Sangrador-Vegas A, McAuliffe O, Slattery L, Kaleta P, Callanan M, Fitzgerald GF, Ross RP, Beresford T (2009) Comparative genomics of lactic acid bacteria reveals a niche-specific gene set. BMC Microbiol 9:50. doi:10.1186/1471-2180-9-50