Development of a multilocus sequence typing method for analysis of *Lactobacillus plantarum* strains

BLANCA DE LAS RIVAS, ÁNGELA MARCOBAL, AND ROSARIO MUÑOZ*

Departamento de Microbiología, Instituto de Fermentaciones Industriales, CSIC, Juan de la Cierva 3, 28006 Madrid, Spain

Running Title

DEVELOPMENT OF MLST METHOD FOR *L. PLANTARUM*

* Corresponding author. Mailing address: Departamento de Microbiología. Instituto de Fermentaciones Industriales, CSIC, Juan de la Cierva 3, 28006 Madrid, Spain. Phone: 34 91 562 2900. Fax: 34 91 564 4853. E-mail: rmunoz@ifi.csic.es

ABSTRACT

Lactobacillus plantarum is a species of considerable industrial and medical interest. To date, the lack of reliable molecular methods for the definite identification at strain level has hindered studies of the population biology of this organism. Here, a multilocus sequence typing (MLST) system for this organism is described, which exploits the genetic variation present in six housekeeping loci to determine the genetic relationship among isolates. The MLST system was established using 16 *L. plantarum* strains that were also characterized by ribotyping and restriction fragment length polymorphism (RFLP) analysis of the PCR- amplified 16S-23S rDNA intergenic spacer region (ISR). Ribotyping grouped the strains into four groups; however, the RFLP analysis of the ISRs showed no differences in the strains analyzed. In contrast, MLST had a good discriminatory ability. The sequence analysis of the six genes showed 14 different allelic combinations, with 12 of them represented by only one strain. By using this MLST approach we were able to assume the identity of two strains deposited in the Spanish Type Culture Collection as different strains. Phylogenetic analyses indicated a panmictic population structure of *L. plantarum*. Split decomposition analysis indicate that recombination plays a role in creating genetic heterogeneity of *L. plantarum*. Because MLST allows a precise identification, and an easy comparison and exchange of results obtained in different laboratories, the future application of this new molecular method could be a useful tool for the identification of valuable *L. plantarum* strains.

INTRODUCTION

Food preserving and flavour development is often carried out by lactic acid bacteria (LAB). The specific environmental conditions prevailing in a fermenting food substrate promote the growth of certain of these bacteria. *Lactobacillus plantarum* is predominantly found (also used as a starter) in fermented food and feed products. *L. plantarum* is implicated in processed food for human consumption like sauerkraut, dry fermented sausage, wine, green olive fermentations and cheese making (18, 20) as well as in animal nutrition such as crop preservation (16), fish and crab waste fermentation (1), and poultry by-product fermentation (22).

Identification of these bacteria is essential in both basic and applied research. Intraspecific differentiation is an important preliminary step for the selection of starter cultures, because technological, probiotic, antimicrobial, and sensorial attributes are strain specific and it may help to distinguish strains with particular technological properties. Currently, a great number of mostly molecular techniques are available for the identification of LAB, for industrial processes and food products. For each specific type of research or analysis, a well-considered choice has to be made of the methodology to be applied, in relation to taxonomic resolution, workload and cost. It is important to realize that every technique cannot be used for any purpose. In the course of safety assessments, it is crucial to use techniques working on the strain level in order to obtain a detailed fingerprint of individual isolates. The increasing interest in some *L. plantarum* properties, e.g., probiotic activities of specific strains (10), contributes to the need for a reliable molecular method for the definite identification of *L. plantarum* at strain level. The need for positive identification of different isolates is also acknowledged by research workers in the

field, since many strains from diverse origins are often exchanged between laboratories, and no reliable phenotypic method for certifying their identities is available. The identification of *L. plantarum* at strain level is also important for their industrial use. The biotechnological industry needs tools for monitoring, e. g., the use of patented strains or to distinguish probiotic strains from natural isolates in the host gastrointestinal tract.

Genotypic methods used for *L. plantarum* strain typing are typically PCR based methods (e.g. ramdomly amplified polymorphic DNA (RAPD) (4, 12, 18)) or variation of restriction enzyme analysis (e.g. ribotyping (19, 25), pulse-field gel electrophoresis (PFGE) (21)). Multilocus sequence typing (MLST) has recently been shown to be a powerful technique for bacterial typing. MLST makes use of automated DNA sequencing to characterize the alleles present at different housekeeping genes. Because it is based on nucleotide sequence, it is highly discriminatory and provides unambiguous results that are directly comparable among laboratories. The MLST method was first described in 1998 and since then it has been applied to important bacterial pathogens including several food-borne human pathogens such as *Campylobacter jejuni* (3), *Vibrio cholerae* (6), and *Bacillus cereus* (9); recently, MLST was also applied to the nonpathogenic food production bacteria *Oenococcus oeni* (2).

The present study was undertaken with three goals: (i) to develop an MLST method for *L. plantarum*, (ii) to compare, the discriminatory power of ribotyping, RFLP of the 16S-23S rDNA ISR, and MLST for this species, and (ii) to use MLST to analyze *L. plantarum* population structure.

MATERIALS AND METHODS

Bacterial strains. A total of 16 strains of *L. plantarum* were used in this study. Seven were isolated from must grape or wine of different wine-producing areas of Spain over the period from 1998 to 2001 (Table 1). Eight strains were provided by the Spanish Type Culture Collection (CECT). *L. plantarum* strains were routinely growth in MRS medium (Difco) at 30 ºC without shaking. Chromosomal DNA was prepared as described previously (23).

Ribotyping. Chromosomal DNA was digested with EcoRI (Roche), and the products were separated by electrophoresis in 0.7% agarose gels in 1X Tris-acetate-EDTA buffer. Digested DNA was transferred onto positively charged nylon membranes (Roche) by Southern blot. Probe 16S rDNA was obtained from *L. plantarum* CECT 748^T by PCR by using the eubacterial universal pair of primers 63f and 1387r (14). The 16S rDNA probe was digoxigenin labeled and detected by chemiluminescence by using a DIG-High Prime DNA Labeling and Detection Starter Kit (Roche) according to the manufacturer´s intructions.

RFLP of the PCR-amplified 16S-23S rDNA ISR. Restriction fragment length polymorphism (RFLP) analysis of the IRSs was performed by using the primers 16S14f and 23S1R based on conserved areas of aligned rRNA bacterial sequences (26). These primers amplified a 550-bp fragment in all the *L. plantarum* strains tested. The PCR was performed in a volume of 50 µl as described previously (26). The amplified 16S-23S ISRs from *L. plantarum* strains were digested with the restriction enzymes AluI, CfoI, DdeI, and

TaqI (Roche). The digested products were separated by electrophoresis in 4.5% MS-8 agarose gels (Hispanlab).

Amplification and nucleotide sequencing. The following housekeeping genes were chosen to be analyzed: phosphoglucomutase (*pgm*), D-alanine-D-alanine ligase (*ddl*), B subunit of DNA gyrase (*gyrB*), ATPase subunit of phosphoribosylaminoimidazole carboxylase (*purK1*), glutamate dehydrogenase (*gdh*), DNA mismatch repair protein (*mutS*), and transketolase (*tkt4*). The DNA sequences of these candidate loci were available from GenBank (Table 2). These genes were selected on the criterion that they are widely separated on the chromosome, except *purK1* and *gdh* genes that were only 28.5 kb apart.

PCR was performed to amplify gene fragments from chromosomal DNA of the *L. plantarum* strains by using oligonucleotides described in Table 2. The conditions of PCR, purification and sequencing of DNA fragments were described previously (2).

Data analysis. For each locus, the sequences obtained for all isolates were compared and allele numbers were assigned to each unique sequence. Each isolate was defined by the combination of numbers corresponding to the alleles at the loci analyzed, which is an allele profile or sequence type (ST) (Table 1). Sequences different even at a single nucleotide site were considered distinct alleles.

Sequence alignments and comparison were done with the program BioEdit (http://jwbrown.mbio.ncsu.edu/BioEdit/bioedit.html) (8) and converted into MEGA and NEXUS files with START [Sequence Type Analysis and Recombinatorial Tests; (http://outbreak.ceid.ox.ac.uk/software.html)].

The method of split decomposition was used to assess the degree of tree-like structure present in the alleles found for each locus in the complete set of 16 isolates (11). The sequence alignments were converted to NEXUS files, and the split decomposition was performed with SPLITSTREE 2.0 (http://bibiserv.techfak.uni-bielefeld.de/splits/). Using the START program we performed a recombination test, the index of association (I_A) (15), and a test for selection, the d_N/d_S ratio (17).

Nucleotide sequence accession numbers. The nucleotide sequences of each allele at the seven loci analyzed in the present work have been assigned the following GenBank accession numbers: AJ966402 to AJ966404 (*pgm*), AJ966405 to AJ966409 (*ddl*), AJ966364 to AJ966370 (*gyrB*), AJ966371 to AJ966378 (*purK1*), AJ966379 to AJ966388 (*gdh*), AJ966389-AJ966396 (*mutS*), and AJ966397-AJ966401 (*tkt4*).

RESULTS

Ribotyping and ISRs RFLP analysis. Ribopattern analysis with EcoRI revealed four bands for all *L. plantarum* strains (Fig. 1A). Four different groups of strains were defined on this basis. Ribotype 1, containing *L. plantarum* CECT 220, showed bands of ca. 2.5, 3.4, 3.7 and 7 kb; ribotype 2, including *L. plantarum* type strain (CECT 748^T) presented bands of ca. 2.5, 3.7, 4.0 and 7 kb; ribotype 3, represented only by *L. plantarum* RM40, showed bands of ca. 2.5, 2.9, 3.7 and 7 kb; and in ribotype 4, containing *L. plantarum* CECT 4645 and RM35, bands of ca. 2.5, 3.7, 4.0 and 10 kb appeared. The

assignment of ribotype groups to each strain is listed in Table 1. Eleven out 16 strains belonged to ribotype group 1.

Primers 16S14F and 23S1R, complementary to target sequences at ca. 140 nucleotides from the 3´end of the 16S rRNA gene and at ca. 120 nucleotides from the 5´end of the 23S rRNA gene (26), respectively, were used to amplify the ISR of the *L. plantarum* strains. A PCR product of approximately 550 bp was obtained for each strain, indicating that the ISR is highly similar in all strains analyzed (data not shown). PCR products of the 16S-23S rDNA ISR were digested with AluI, CfoI, NdeI and TaqI. All the strains analyzed showed identical RFLP patterns for each enzyme: one apparent band of 500 pb (AluI), two bands of 120 and 430 pb (CfoI), two bands of 90 and 370 pb (DdeI), and two bands of 200 and 350 pb (TaqI) (Fig. 1B).

Variation at the MLST loci. The sequences of the seven chosen loci were determined for the 16 strains, with exception of *tkt4* locus which could not be amplified from strains CECT 223, 224 and 4645. The alleles defined for the MLST scheme were between 414 bp (*gdh*) and 704 bp (*gyrB*) in length, and between 3 (*pgm*) and 10 (*gdh*) alleles were present per locus (Table 3, Figure 2). The proportion of variable sites present in the MLST alleles ranged from 1.03% (*ddl*) to 7.72% (*gdh*). The average G+C contents of the different gene fragments ranged from 42.7 (*ddl*) to 50.5% (*gdh*). The G+C content of the *L. plantarum* chromosome was 44.5% (13), so the *gdh* allele sequences showed an unexpected high G+C content.

The proportions of nucleotide alterations that changed the amino acid sequence (nonsynonymous substitutions, d_N) and the proportions of silent changes (synonymous substitutions, d_S) were calculated for each gene. With these data, the d_N/d_S ratios were

calculated for all loci and were all <1, indicating that most of the sequence variability identified is selectively neutral (Table 3).

The allele frequencies showed that for most of the loci one allele (in *gyrB*, *mutS* and *purK1*) or two alleles (in *ddl* and *pgm*) (Fig. 2, Table 1) were dominant in the population.

Relationship of *L. plantarum* strains by sequence analysis of housekeeping loci. For all genes, multiple strains carried identical alleles. However, only two pairs of strains had identical sequences for all fragments, and all other strains could be distinguished from each other because they had unique combinations of alleles (Table 1). Allelic profiles (ST) were assigned (Table 1). All STs differed in various loci, except ST-5 and ST-6 that differ only in the *gdh* locus. Concatenated *pgm*, *ddl*, *gyrB*, *purK1*, *gdh*, and *mutS* gene sequences fragments were analyzed and graphically displayed with SplitsTree (Fig. 3). The algorithm used in this software is able to display conflicting results in the phylogenetic descent of sequences. A tree-like structure is created when the descent is clonal, but an interconecting network or a parallelogram will appear when recombination may have been involved in the evolution of the analyzed gene. Figure 4 shows the split graphs for all alleles of the six fragments analyzed. The fit parameter was 100, indicating that all phylogenetic information in the sequences could be visualized in the graphs. We observed parallelograms for two of the six genes examined. The *purK1* and *gdh* loci present network-like or parallelograms structures, indicating the presence of homoplasies, probably evolved by intergenic recombination. The split graphs obtained with *pgm*, *ddl*, *gyrB*, and *mutS* loci showed no evidence of network-like evolution.

Linkage disequilibrium between alleles was estimated with the standardized I_A, I^S_A, an statistic that does not depend on the number of loci analyzed and is expected to be zero

when the alleles are in linkage equilibrium (free recombination). The I^S_A value was 0.139, indicating that the genes investigated are close to linkage equilibrium. This low (I^S_A) value is also indicative of extensive recombination.

DISCUSSION

The identification of strains belonging to the same species is still difficult. Until now, a suitable and precise *L. plantarum* typing method is not available and it is urgent since strain characterization is necessary prior to patenting and release of a valuable *L. plantarum* strain for commercial applications.

Ribotyping is a both intra- and interspecific typing method that has been successfully used for differentiation between strains of the same *Lactobacillus* species. Yansanjav et al. (2003) (25) used ribotype on five *L. plantarum* strains. These five strains formed two ribotype patterns, and they found that strains isolated from different breweries shared the same ribotype. Other studies, based on their ribotypes, apparently identical pairs of *L. plantarum* strains were isolated in very different settings: *L. plantarum* ATCC 14917^T (CECT 748^T) was purchased by the authors from ATCC, and *L. plantarum* UH 2153 isolated from the vagina of a woman in Ontario (27). In our study we found four different ribotypes; however, most of strains, 11 out 15, belonged to the same ribotype 2. These ribotype 2 strains are not time or geographically related, therefore, more specific identification methods, such as the use of MLST method would be preferred.

In addition, the variation in length and sequence of the 16S-23S rDNA spacer regions have been used to type strains. The ISRs usually show variations which made it

possible to discriminate between strains within some species. As shown in Table 1 and Fig. 1, all the *L. plantarum* strains analyzed in this study shared the same RFLP-ISR pattern. Therefore, the RFLP-ISR results support the idea about the necessity of a more discriminating method able to differentiate between *L. plantarum* strains.

As a first step for developing a MLST typing method, we analyzed the sequence diversity of seven housekeeping genes in order to know if they are sufficient to have enough typing discrimination. The internal fragments of six loci that were selected (*pgm*, *ddl*, *gyrB*, *purK1*, *gdh*, and *mutS*) could be amplified from all the strains examined. The amplified internal fragments were sequenced, and from these sequences we were able to use fragments between 414 and 704 bp for analysis (Table 3). However, internal *tkt4* fragments, could not be amplified from strains CECT 223, 224, and 4645. From these results, we can not exclude possibilities such as a *tkt4* gene deletion or the presence of a non-homologous gene copy on these strains. In the strains where the *tkt4* locus could be amplified, we sequenced a 567 bp fragments having six polymorphic sites that generate 5 different alleles (data not shown). Since the *tkt4* locus could not be amplified from all the *L. plantarum* strains, the *tkt4* locus was discarded from the MLST scheme typing.

The number of alleles from the six housekeeping loci ranged from 3 to 10 (Table 3). Our results corroborates the previously described genetic heterogeneity of this species (21, 24). The six loci were polymorphic, and most of the types were represented by a single strain. We only found two exceptions. One of them, are strains CECT 223 and 224 that were deposited in the CECT by D. M. Alvarez Marques in 1987 and isolated in Pamplona, Spain (Table 1). By the three typing methods used in this study, both strains are identical, therefore, it is tempting to assume that both isolates could belong to the same original strain. A similar situation was observed in strains RM71 and RM72, since both strains were

isolated from the same winery in the same year, and they shared identical types by all the typing methods used in this study.

However, the MLST method was able to discriminate between highly similar strains. That is the case of strains CECT 748^T and 749 that were identical at five out six genes analyzed, differing only on the *gdh* locus (Table 1). Moreover, allele 1 (CECT 748^T) and allele 3 (CECT 749) on the *gdh* locus showed multiple nucleotide differences (Fig. 2). The significant nucleotide differences at the *gdh* locus (1 and 3 alleles) must be due to recombination, as it seems inconceivable that mutation alone could give such level of divergence while other genes did not diverge at all. All the genes analyzed encode proteins involved in housekeeping functions, with no reason to expect differences in the level of selection to account for the disparity in sequence variation in the different genes. Therefore, the more likely explanation of these differences could be a recombinational exchange, possible among isolates of the same species, e. g. with alleles 4, 5 or 6 in the *gdh* locus. This example demonstrated the highly discriminating power showed by the MLST scheme proposed.

Examination of the sequences of housekeeping genes from biosynthetic pathways can provide evidence for the significance of recombination, since the variation within these genes is likely to be selectively neutral. Recombination can be detected by, e. g., the appearance of a network of relationships between sequences rather than a bifurcating tree-like structure phylogeny. The split decomposition analysis of the *L. plantarum* strains (Fig. 4) reveals two uncentered edges, suggesting that the evolution of these strains has been initiated by a couple of strains and from these two origins, single branches radiate.

The most simple method to detect recombination in the aligned sequences is the detection of mosaic structures by eye. Significant mosaic structure is indicative of

recombinatorial exchange, usually among isolates of the same species (7). In our study, four possible examples of recombination events were found: allele 3 of *gyrB*, allele 6 (*gdh*), allele 8 (*purK1*), and allele 3 of *pgm*. The mean divergence between allele 3 of gyrB (3.97%) is much higher than the mean diversity within the other *gyrB* alleles (0.57%) (Fig. 2). This divergence is similar to the divergence observed previously in a possible example of an intergenic recombinatorial event in the *gdh* gene between *S. pneumoniae* and *S. mitis* (5)

The utility of MLST for the analysis of the genetic structure of bacterial populations is mainly based on the characteristics of housekeeping genes to have a selectively neutral variability. Analysis of synonymous and nonsynonymous changes in the allele sequences of a locus can be used to determine if it is subject to positive selection, so a d_N/d_S ratio of greater than 1 implies selection for amino acid changes. In our genetic analysis, the six housekeeping loci had d_N/d_S ratios significantly lower than 1 (Table 3). Another important characteristic in relation to this fact is that the location of loci on the chromosome, at exception of the *purK1* and *gdh* loci, was distant enough to make the joint horizontal transfer of two loci unlikely.

Three types of populations structure are known in bacteria: clonal, panmictic and epidemic. Panmictic populations may be so variable that identical strains are only found among isolates from direct contacts. In our study, only strains isolated from the same source, such as the strains CECT 223 and 224, and strains RM71 and 72 share the same ST. The analysis of the *L. plantarum* population structure presented here suggested a panmictic, non-clonal population structure with a substantial extend of recombination. The split decomposition analysis give strong evidence that intraspecies recombination occurs frequently in *L. plantarum* and plays a major role in generating genetic heterogeneity

between strains. A low I^S_A value (0.138) is indicative of a weakly clonal population and also confirms the importance of recombination in *L. plantarum* and supports the estimation that the genes investigated in *L. plantarum* are close to linkage equilibrium. The extension of the present analysis to a higher number of isolates could contribute to a better knowledge of the structure of the *L. plantarum* populations.

In conclusion, the MLST scheme presented here will be a useful new tool for the precise and unambiguous characterization of *L. plantarum* isolates and appears to have sufficient discriminatory power for population investigations and should allow comparison of the isolates of this species by a means much easier and more precise than the typing procedures used at present.

ACKNOWLEDGMENTS

This work was supported by grant RM03-002 from the INIA. We thank R. González and E. García for their critical reading of the manuscript. The technical assistance of M. V. Santamaría is greatly appreciated. B. de las Rivas was a recipient of a postdoctoral fellowship and A. Marcobal of a predoctoral fellowship both from the Comunidad de Madrid.

REFERENCES

1. **Abazinge, M. D. A., J. P. Fontenot, V. G. Allen, and G. J. Flick.** 1993. Ensiling characteristics of crab waste and wheat straw treated with different additives. J. Agric. Food Chem. **41:**657-661.
2. **De las Rivas, B., A. Marcobal, and R. Muñoz.** 2004. Allelic diversity and population structure in *Oenococcus oeni* as determined from sequence analysis of housekeeping genes. Appl. Environ. Microbiol. **70:**7210-7219.
3. **Dingle, K. E., F. M. Colles, D. R. A. Wareing, R. Ure, A. J. Fox, F. E. Bolton, H. J. Bootsma, R. J. L. Willems, R. Urwin, and M. C. J. Maiden.** 2001. Multilocus sequence typing system for *Campylobacter jejuni*. J. Clin. Microbiol. **39:**14-23.
4. **Elegado, F. B., M. A. R. V. Guerra, R. A. Macayan, H. A. Mendoza, and M. B. Lizaran.** 2004. Spectrum of bacteriocin activity of *Lactobacillus plantarum* BS and fingerprinting by RAPD-PCR. Int. J. Food Microbiol. **95:**11-18.
5. **Enright, M. C., and B. G. Spratt**. 1998. A multilocus sequence typing scheme for *Streptococcus pneumoniae*: identification of clones associated with serious invasive disease. Microbiology **144:**3049-3060.
6. **Farfán, M., D. Miñana-Galbis, M. C. Fusté, and J. G. Lorén.** 2002. Allelic diversity and population structure in *Vibrio cholerae* O139 Bengal based on nucleotide sequence analysis. J. Bacteriol. **184:**1304-1313.
7. **Feil, E. J., M. C. Enright, and B. G. Spratt**. 2000. Estimating the relative contributions of mutation and recombination to clonal diversification: a comparison between *Neisseria meningitidis* and *Streptococcus pneumoniae*. Res. Microbiol. **151:**465-469.

8. **Hall, T. A.** 1999. Bio Edit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98NT. Nucleic Acids Symp. Ser. **41:**95-98.
9. **Helgason, E., N. J. Tourasse, R. Meisal, D. A. Caugant, and A. B. Kolsto**. 2004. Multilocus sequence typing scheme for bacteria of the *Bacillus cereus* group. Appl. Environ. Microbiol. **70:**191-201.
10. **Herias, M. V., C. Hessle, E. Telemo, T. Midtvedt, L. A. Hanson, and E. Wold.** 1999. Immunomodulatory effects of *Lactobacillus plantarum* colonizing the intestine of gnotobiotic rats. Clin. Exp. Immunol. **116:**283-290.
11. **Huson, D. H.** 1998. SplitsTree: analyzing and visualizing evolutionary data. Bioinformatics **14:**68-73.
12. **Johansson, M. –L., M. Quednau, G. Molin, and S. Ahrné**. 1995. Randomly amplified polymorphic DNA (RAPD) for rapid typing of *Lactobacillus plantarum* strains. Lett. Appl. Microbiol. **21:**155-159.
13. **Kleerebezem, M., J. Boekhorst, R. van Kranenburg, D. Molenaar, O. P. Kuipers, R. Leer, R. Tarchini, S. A. Peters, H. M. Sandbrink, M. W. E. J. Fiers, W. Stiekema, R. M. K. Lankhorst, P. A. Bron, S. M. Hoffer, M. N. Groot, R. Kerkhoven, M de Vries, B. Ursing, W. M. de Vos, and R. J. Siezen.** (2003). Complete genome sequence of *Lactobacillus plantarum* WCFS1. Proc. Natl. Acad. Sci. USA. **100:**1990-1995.
14. **Marchesi, J. R., T. Sato, A. J. Weightman, T. A. Martin, J. C. Fry, S. J. Hiom, and W. G. Wade.** 1998. Design and evaluation of useful bacterium-specific PCR primers that amplify genes coding for bacterial 16S rRNA, Appl. Environ. Microbiol. **64:**795-799.

15. **Maynard Smith, J., N. H. Smith, M. O´Rourke, and B. G. Spratt**. 1993. How clonal are bacteria?. Proc. Natl. Acad. Sci. USA. **90:**4384-4388.
16. **Merry, R. J., M. S. Dhanoa, and M. K. Theodorou.** 1995. Use of freshly cultured lactic acid bacteria as silage inoculants. Grass For. Sci. **50:**112-123.
17. **Nei, M, and T. Gojobori.** 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol. Biol. Evol. **3:**418-426.
18. **Oneca, M., A. Irigoyen, M. Ortigosa, and P. Torre**. 2003. PCR and RAPD identification of *L. plantarum* strains isolated from ovine milk and cheese. Geographical distribution of strains. FEMS Microbiol. Lett. **227:**271-277.
19. **Rodas, A. M., S. Ferrer, and I. Pardo**. 2005. Polyphasic study of wine *Lactobacillus* strains: taxonomic implications. Int. J. Syst. Evol. Microbiol. **55:**197-207.
20. **Ruiz-Barba, J. L., D. P. Cathcart, P. J. Warner, and R. Jiménez-Díaz.** 1994. Use of *Lactobacillus plantarum* LPCO10, a bacteriocin producer, as a starter culture in spanish-style green olive fermentations. Appl. Environ. Microbiol. **60:**2059-2064.
21. **Sánchez, I., S. Seseña, and L. L. Palop.** 2004. Polyphasic study of the genetic diversity of lactobacilli associated with "Almagro" eggplants spontaneous fermentation, based on combined numerical analysis of randomly amplified polymorphic DNA and pulse-field gel electrophoresis patterns. J. Appl. Microbiol. **97:**446-458.
22. **Urlings, H**. **A. P., P. G. H. Bijker, and J. G. van Logtestijn**. 1993. Fermentation of raw poultry by-products for animal nutrition. J. Anim. Sci. **71:**2420-2426.
23. **Vaquero, I., A. Marcobal, and R. Muñoz.** 2004. Tannase activity by lactic acid bacteria isolated from grape must and wine. Int. J.Food Microbiol. **96:**199-204.

24. **Vescoso, M., S. Torriani, F. Dellaglio, and V. Bottazi**. 1993. Basic characteristics, ecology and application of *Lactobacillus plantarum*: a review. Ann. Microbiol. Enzymol. **43:**261-284.

25. **Yansanjav, A., P. Svec, I. Sedlacek, I. Hollerova, and M. Nemec**. 2003. Ribotyping of lactobacilli isolated from spoiled beer. FEMS Microbiol. Lett. **229:**141-144.

26. **Zavaleta, A**. **I., A. J. Martínez-Murcia, and F. Rodríguez-Varela.** 1996. 16S-23S rDNA intergenic sequences indicate that *Leuconostoc oenos* is phylogenetically homogeneous. Microbiology **142:**2105-2114.

27. **Zhong, W., K. Millsap, H. Bialkowska-Hobrazanska, and G. Reid**. 1998. Differentiation of *Lactobacillus* species by molecular typing. Appl. Environ. Microbiol. **64:**2418-2423.

Legends to Figures

FIG. 1. Typing analysis of the 16 *L. plantarum* strains examined in this study. (A) Ribotyping patterns of EcoRI digest of *L. plantarum* chromosomal DNAs. Ribotype 1 (lane 1), ribotype 2 (lane 2), ribotype 3 (lane 3), and ribotype 4 (lane 4). The molecular sizes (in kilobases) of the labeled fragments are indicated on the left. (B) 16S-23S rDNA ISR patterns obtained after AluI (lane 1), CfoI (lane 2), DdeI (lane 3), and TaqI (lane 4) digestion of the 550-bp DNA fragment PCR amplified with primers 16S14f and 23S1R. The molecular sizes (in kilobases) of some standards of the 50-bp leader marker are indicated on the left.

FIG. 2. Polymorphic nucleotide sites in *L. plantarum* MLST genes. Only the variable sites are shown. The nucleotide at each site is shown for a putative consensus sequence; only those that differ from the nucleotide in the consensus sequence are shown for the alleles. Nucleotide sites are numbered in vertical format from the first nucleotide position of the corresponding gene. The number of strains possessing the allele is indicated in parentheses

FIG. 3. Split decomposition analysis based on the allelic profiles of the 16 *L. plantarum* strains examined in this study. The numbering refers to strain numbers.

FIG. 4. Split decomposition analysis of alleles obtained from 16 *L. plantarum* strains for six loci. The observation that in the *purK1* and *gdh* graphs several alleles in the sample are conected to each other by multiple pathways, forming parallelograms structures, is

suggestive of recombination. All branch lengths are drawn to scale. The numbering refers to allele numbers.

TABLE 1. Properties of *L. plantarum* isolates

Strain no.	Strain	RT[a]	ITS[b]	ST[c]	Allele no. at locus						Source of isolate	
					pgm	*ddl*	*gyrB*	*purK1*	*gdh*	*mutS*	Country	Year
1	WCFS1	ND[d]	ND	1	1	1	1	1	1	1	United Kingdom	1956
2	CECT 220 (ATCC 8014)	1	1	2	1	3	1	2	6	5	ND	1948
3	CECT 221 (ATCC 14431)	2	1	3	1	1	6	3	8	1	ND	1960
4	CECT 223	2	1	4	1	5	3	4	7	1	Pamplona, Spain	1987
5	CECT 224	2	1	4	1	5	3	4	7	1	Pamplona, Spain	1987
6	CECT 748^T (ATCC 14917)	2	1	5	3	2	1	2	1	1	Denmark	1919
7	CECT 749 (ATCC 10241)	2	1	6	3	2	1	2	3	1	ND	1955
8	CECT 4185 (NCBF 1193)	2	1	7	3	1	2	1	5	1	ND	1958
9	CECT 4645 (NCBF 965)	4	1	8	2	4	7	8	10	4	ND	1958
10	RM28	2	1	9	1	2	1	7	7	3	Valladolid, Spain	2000
11	RM35	4	1	10	3	2	1	2	2	8	Toledo, Spain	1998
12	RM38	2	1	11	3	1	4	2	4	2	Toledo, Spain	1998
13	RM40	3	1	12	1	1	1	5	9	1	Toledo, Spain	1998
14	RM71	2	1	13	3	1	5	6	5	6	Valladolid, Spain	2001
15	RM72	2	1	13	3	1	5	6	5	6	Valladolid, Spain	2001
16	RM73	2	1	14	3	2	1	2	1	7	Madrid, Spain	2000

[a] RT, Ribotype
[b] ITS, 16S-23S ISR Type
[c] ST, Sequence Type
[d] ND, no data available

TABLE 2. Primers used for MLST

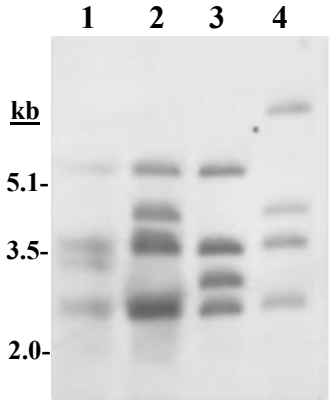
Protein	Accession no.	Gene	Primers	Sequence 5′→3′	5′start position	Accession no.
Phosphoglucomutase	NP_784514.1	*pgm*	PGM1-LP	CTTGCGGCCAACCCCAGAAC	363	AJ966402-AJ966404
			PGM2	CCGTAGGATTCTTCAAAACC	1226	
D-alanine-D-alanine ligase	NP_785815.1	*ddl*	DDL1-LP	AACATGATGTTTCGAAGCG	56	AJ966405-AJ966409
			DDL2-LP	GTTAGTAAAACCAGGTAACG	972	
DNA gyrase, B subunit	NP_783871.1	*gyrB*	GYRB1-LP	GTGGTCTTCACGGGGTCG	368	AJ966364-AJ966370
			GYRB2-LP	TTCGACAATGAACAACAC	1314	
Phosphoribosylaminoimidazole carboxylase, ATPase subunit	NP_786114.1	*purK1*	PURK1	TGACCTACGAGTTTGAAAAC	230	AJ966371-AJ966378
			PURK2	GGTGACATGACCCATCTTGCG	1068	
Glutamate dehydrogenase	NP_784837.1	*gdh*	GDH1	CCTTACAAGGGCGGCTTACG	271	AJ966379-AJ966388
			GDH2	ACGCCACCAGCATTGGCAGC	1130	
DNA mismatch repair protein	NP_785757.1	*mutS*	MUT1	AAGTACGTTCTCATCCCATATG	1170	AJ966389-AJ966396
			MUT2	ATAACGCACACCCCGCAGGTC	2157	
Transketolase	NP_786741.1	*tkt4*	TKT41	GGTGATGGCGACTTAATGG	469	AJ966397-AJ966401
			TKT42	CCCATCCTCGCCGACCGC	1410	

TABLE 3. Sequence variation at six loci

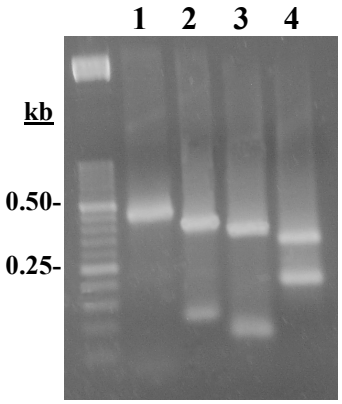
Gene	Fragment size (bp)	Mean G+C content (%)	No. of alleles	No. of polymorphic sites[a]	No. of nucleotide substitutions per nucleotide site	d_N/d_S[b]
pgm	558	43.3	3	6 (6)	0.0004	0
ddl	677	42.7	5	7 (6)	0.0008	4.78
gyrB	704	45.2	7	41 (41)	0.0056	0
purK1	525	48.1	8	12 (8)	0.0023	11.15
gdh	414	50.5	10	32 (29)	0.0072	2.07
mutS	594	47.0	8	7 (5)	0.0010	8.72

[a] Number of silent polymorphic sites in parentheses
[b] Calculated by using the START program expressed as d_N/d_S ratio multiplied by 100

A



B



gyrB	*0000000000000000000000000011111111111111*
	6666677777777778889999999900000001111122
	1336601125678891490033677780011248667780 2
	8360321773276989673703905872547940183686 1
consensus	**CGATCTTTCTGGTGATCCTTTTCCTCTGTCCTTGCTTAGGA**
allele 1 (8)	.....
allele 2 (1)	.....T.....T..
allele 3 (2)	T..CTCCC.CAC.CGC.TG.CA.TATC.CTTCC....T.AG
allele 4 (2)	.A....C.....T..T...
allele 5 (1)	.A....C.....TC.T...
allele 6 (1)	.A....C.....A..C.....T.CT...
allele 7 (1)	......C.T...C.....T.....AT..T...

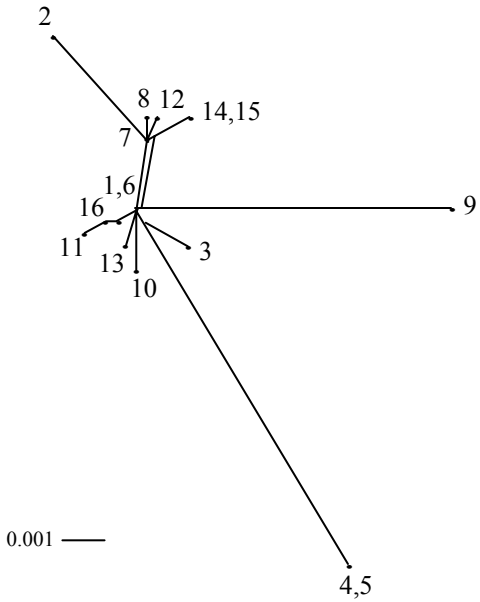
gdh	*000000000000000000000000000111*
	6666777777788888888889999999000
	6689011256834456677889334556812 5
	3648547038310924506284092456816 0
consensus	**CAAGCTTTGTCCTCGCGCATCACATGTCGCCT**
allele 1 (1)	.....T.....T.....
allele 2 (1)	.....T.....T..T.....
allele 3 (1)	T.G.....AC.T..A...GC.G.....
allele 4 (1)	T.G...C.AC.T..A.A.GC.G.....
allele 5 (3)	T.G...C.AC.T..A...GC.G.....
allele 6 (1)	T.G..C.CAC....A...GC.GTGCACG...A
allele 7 (3)	.....A.....T.T.....A...
allele 8 (1)	.G.....A.....T.T.....A...
allele 9 (1)	....T.....G..T.....
allele 10 (1)	...A.....G..T.....T.....TTC

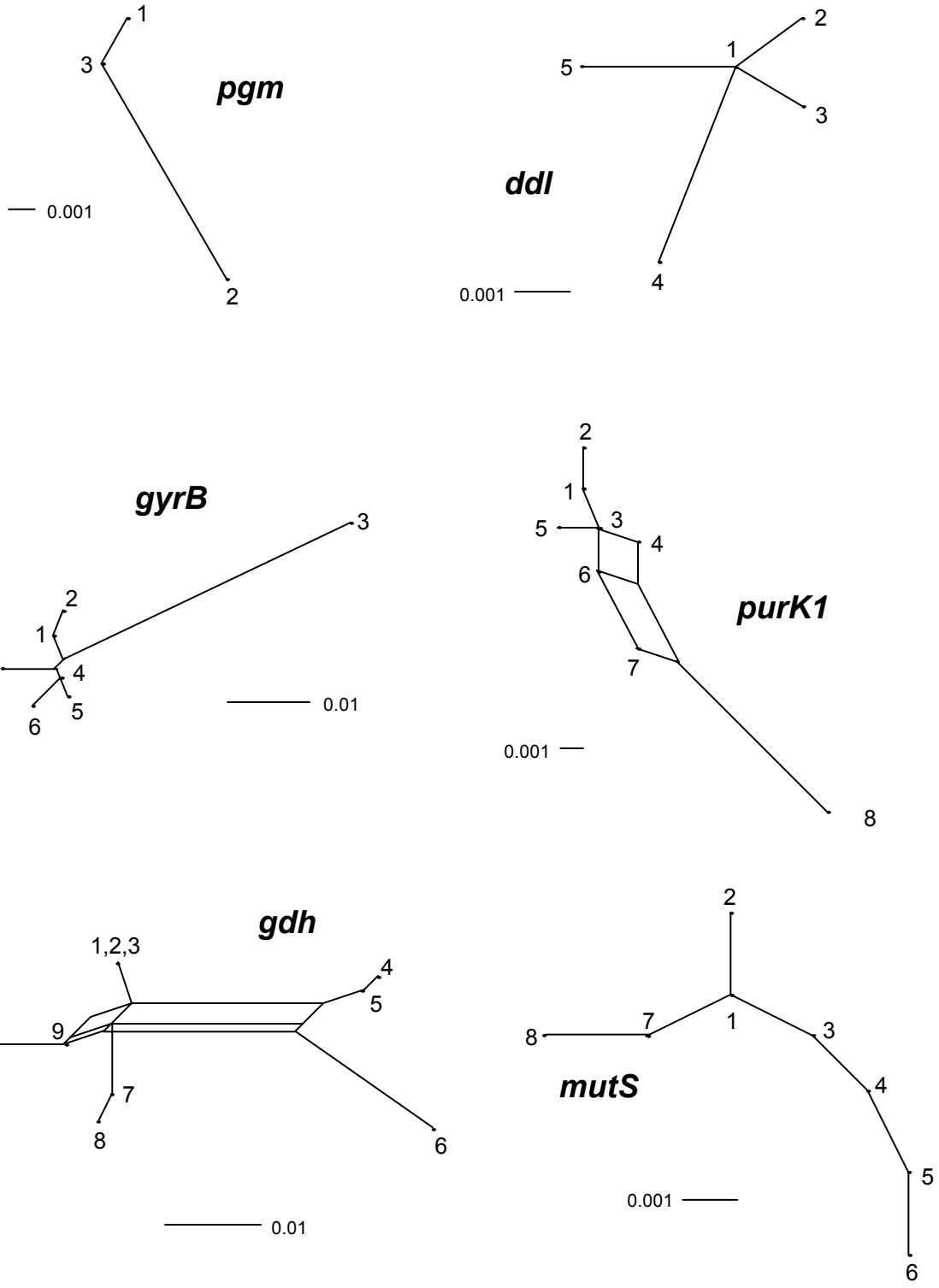
mutS	*1111111*
	5677888
	6101004
	0159398
consensus	**CATCGTG**
allele 1 (8)	.....
allele 2 (1)	.....A
allele 3 (1)	...T...
allele 4 (1)	...T.C.
allele 5 (1)	...TAC.
allele 6 (2)	T...AC.
allele 7 (1)	..A....
allele 8 (1)	.TA....

ddl	*3356777*
	2647389
	4363298
consensus	**GAGCCGG**
allele 1 (7)	.....
allele 2 (5)	.....A.
allele 3 (1)	.....A
allele 4 (1)	CG..T..
allele 5 (2)	..AT...

purK1	*555556777789*
	0227984445 68
	0227984445 68
consensus	**GTCAGATTGCCT**
allele 1 (2)	.....
allele 2 (6)	..T.....
allele 3 (1)	.....C....
allele 4 (2)	.C.....C....
allele 5 (1)	.....CA...
allele 6 (2)	.....GC....
allele 7 (1)	.....GGC...G
allele 8 (1)	AC.GAGGC.TTG

pgm	*566788*
	6378 29
	4609 51
consensus	**CACGTT**
allele 1 (8)	.....
allele 2 (7)	T.....
allele 3 (1)	.GTACC


2
8
12
14,15
7
1,6
16
9
11
13
3
10
4,5
0.001


pgm
1
3
2
0.001
ddl
2
1
5
3
4
0.001
gyrB
3
2
1
7
4
5
6
0.01
purK1
2
1
5
3
4
6
7
8
0.001
gdh
1,2,3
4
5
9
10
7
8
6
0.01
mutS
2
7
1
3
8
4
5
6
0.001