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

Blanca de las Rivas, Ángela Marcobal and Rosario Muñoz

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

*Lactobacillus plantarum* is a species of considerable industrial and medical interest. To date, the lack of reliable molecular methods for 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, RFLP analysis of the ISRs showed no differences in the strains analysed. 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 confirm the identity of two strains deposited in the Spanish Type Culture Collection as different strains. Phylogenetic analysis indicated a panmictic population structure of *L. plantarum* and split decomposition analysis indicated that recombination plays a role in creating genetic heterogeneity in *L. plantarum*. As MLST allows precise identification, and easy comparison and exchange of results obtained in different laboratories, the future application of this new molecular method could be useful for the identification of valuable *L. plantarum* strains.

INTRODUCTION

Lactic acid bacteria (LAB) are often responsible for food preservation and flavour development. The specific environmental conditions prevailing in a fermenting food substrate promote the growth of some of these bacteria. *Lactobacillus plantarum* is predominantly found in fermented food and feed products, and is implicated in the processing of food for human consumption, such as sauerkraut, dry fermented sausage, wine and green olive fermentations and cheese making (Ruiz-Barba et al., 1994; Oneca et al., 2003), as well as in animal nutrition, such as crop preservation (Merry et al., 1995), fish and crab waste fermentation (Abazinge et al., 1993) and poultry by-product fermentation (Urlings et al., 1993).

Identification of *L. plantarum* strains 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 regarding the methodology to be applied in relation to taxonomic resolution, workload and cost. It is important to bear in mind that not every technique can be used for every purpose. In the course of safety assessments, it is crucial to use techniques working at the strain level to obtain a detailed fingerprint of individual isolates. The increasing interest in some *L. plantarum* properties, for example probiotic activities of specific strains (Herias et al., 1999), contributes to the need for a reliable molecular method for the definite identification of *L. plantarum* at the 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

**Abbreviations:** $d^n$ non-synonymous substitutions; $d^s$ synonymous substitutions; $I^a$, standardized index of association; ISR, intergenic spacer region; MLST, multilocus sequence typing; RFLP, restriction fragment length polymorphism; ST, sequence typing.

identities is available. The identification of *L. plantarum* at the strain level is also important for industrial use. The biotechnological industry needs tools for monitoring, for example, 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, for example randomly amplified polymorphic DNA (Johansson *et al.*, 1995; Oneca *et al.*, 2003; Elegado *et al.*, 2004) or restrictions enzyme analysis, for example ribotyping (Yansanjav *et al.*, 2003; Rodas *et al.*, 2005) and pulsed-field gel electrophoresis (Sánchez *et al.*, 2004). 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 gene loci. As it is based on nucleotide sequence, it is highly discriminatory and provides unambiguous results that are directly comparable between 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* (Dingle *et al.*, 2001), *Vibrio cholerae* (Farfán *et al.*, 2002) and *Bacillus cereus* (Helgason *et al.*, 2004); recently, MLST was also applied to the non-pathogenic food production bacterium *Oenococcus oeni* (De las Rivas *et al.*, 2004).

The aims of the present study were to: (i) develop an MLST method for *L. plantarum*; (ii) compare the discriminatory power of ribotyping, restriction fragment length polymorphism (RFLP) of the 16S–23S rDNA intergenic spacer region (ISR) and MLST in this species; and (ii) use MLST to analyse *L. plantarum* population structure.

**METHODS**

**Bacterial strains.** A total of 16 strains of *L. plantarum* were used in this study. The sources of the strains are listed in Table 1. Eight strains were provided by the Spanish Type Culture Collection (CECT). *L. plantarum* strains were routinely grown in MRS medium at 30°C without shaking. Chromosomal DNA was prepared as described previously (Vaquero *et al.*, 2004).

**Ribotyping.** Chromosomal DNA was digested with EcoRI and the products were separated by electrophoresis in 0.7% agarose gels in 1× Tris/acetate/EDTA buffer. Digested DNA was transferred onto positively charged nylon membranes by Southern blotting. Probe 16S rDNA was obtained from *L. plantarum* CECT 748T by PCR by using the eubacterial universal primer pair 63f and 1387r (Marchesi *et al.*, 1998). The 16S rDNA probe was digoxigenin-labelled and

**Table 1. Properties of *L. plantarum* isolates**

<table>
<thead>
<tr>
<th>Strain no.</th>
<th>Strain</th>
<th>RT</th>
<th>ITS</th>
<th>ST</th>
<th>Allele no. at locus:</th>
<th>Source of isolate</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>pgm ddi gyrB purK1 gdh mutS</td>
<td>Origin</td>
</tr>
<tr>
<td>1</td>
<td>WCFS1</td>
<td>ND</td>
<td>ND</td>
<td>1 1 1 1 1 1 1</td>
<td>Human saliva</td>
<td>UK</td>
</tr>
<tr>
<td>2</td>
<td>CECT 220 (ATCC 8014)</td>
<td>1</td>
<td>1</td>
<td>2 1 3 1 2 6 5</td>
<td>Corn silage</td>
<td>ND</td>
</tr>
<tr>
<td>3</td>
<td>CECT 221 (ATCC 14431)</td>
<td>2</td>
<td>1</td>
<td>3 1 1 6 3 8 1</td>
<td>Grass silage</td>
<td>ND</td>
</tr>
<tr>
<td>4</td>
<td>CECT 223</td>
<td>2</td>
<td>1</td>
<td>4 1 3 5 3 4 7 1</td>
<td>ND</td>
<td>Pamplona, Spain</td>
</tr>
<tr>
<td>5</td>
<td>CECT 224</td>
<td>2</td>
<td>1</td>
<td>4 1 3 5 3 4 7 1</td>
<td>ND</td>
<td>Pamplona, Spain</td>
</tr>
<tr>
<td>6</td>
<td>CECT 748T (ATCC 14917)</td>
<td>2</td>
<td>1</td>
<td>5 3 2 1 2 1 1</td>
<td>Pickled cabbage</td>
<td>Denmark</td>
</tr>
<tr>
<td>7</td>
<td>CECT 749 (ATCC 10241)</td>
<td>2</td>
<td>1</td>
<td>6 3 2 1 2 3 1</td>
<td>Pickled cabbage</td>
<td>ND</td>
</tr>
<tr>
<td>8</td>
<td>CECT 4185 (NCBF 1193)</td>
<td>2</td>
<td>1</td>
<td>7 3 1 2 1 5 1</td>
<td>Silage</td>
<td>ND</td>
</tr>
<tr>
<td>9</td>
<td>CECT 4645 (NCBF 965)</td>
<td>4</td>
<td>1</td>
<td>8 2 4 7 8 10 4</td>
<td>Cheese</td>
<td>ND</td>
</tr>
<tr>
<td>10</td>
<td>RM28</td>
<td>2</td>
<td>1</td>
<td>9 1 2 1 7 7 3</td>
<td>Wine</td>
<td>Valladolid, Spain</td>
</tr>
<tr>
<td>11</td>
<td>RM35</td>
<td>4</td>
<td>1</td>
<td>10 3 2 1 2 2 8</td>
<td>Wine</td>
<td>Toledo, Spain</td>
</tr>
<tr>
<td>12</td>
<td>RM38</td>
<td>2</td>
<td>1</td>
<td>11 3 1 4 2 4 2</td>
<td>Wine</td>
<td>Toledo, Spain</td>
</tr>
<tr>
<td>13</td>
<td>RM40</td>
<td>3</td>
<td>1</td>
<td>12 1 1 1 5 9 1</td>
<td>Wine</td>
<td>Toledo, Spain</td>
</tr>
<tr>
<td>14</td>
<td>RM71</td>
<td>2</td>
<td>1</td>
<td>13 3 1 5 6 5 6</td>
<td>Wine</td>
<td>Valladolid, Spain</td>
</tr>
<tr>
<td>15</td>
<td>RM72</td>
<td>2</td>
<td>1</td>
<td>13 3 1 5 6 5 6</td>
<td>Wine</td>
<td>Valladolid, Spain</td>
</tr>
<tr>
<td>16</td>
<td>RM73</td>
<td>2</td>
<td>1</td>
<td>14 3 2 1 2 1 7</td>
<td>Wine</td>
<td>Madrid, Spain</td>
</tr>
</tbody>
</table>
detected by chemiluminescence by using a DIG-High Prime DNA Labelling and Detection Starter Kit (Roche), according to the manufacturer’s instructions.

RFLP of the PCR-amplified 16S–23S rDNA ISR. RFLP analysis of the ISRs was performed by using primers 16S14P and 23S1R based on conserved areas of aligned rRNA bacterial sequences (Zavaleta et al., 1996). These primers amplified a 550 bp fragment in all the L. plantarum strains tested. PCR was performed in a volume of 50 μl as described by Zavaleta et al. (1996). The amplified 16S–23S ISRs from L. plantarum strains were digested with restriction enzymes AluI, CfoI, DdeI and TaqI, since, by searching in the L. plantarum strain WCP1 sequence (NC_004567), recognition sites for these enzymes were found in the 16S–23S ISR. The digested products were separated by electrophoresis in 4.5 % MS-8 agarose gels.

Amplification and nucleotide sequencing. The housekeeping genes encoding the following proteins were chosen for analysis: phosphoglucomutase (pgm), D-alanine-D-alanine ligase (dil), B subunit of DNA gyrase (gyrB), ATPase subunit of phosphoribosylaminomimidazole carboxylase (purK1), glutamate dehydrogenase (gdh), DNA mismatch repair protein (mutS) and transketolase (tkr). The DNA sequences of these candidate loci are available from GenBank (Table 2). These housekeeping genes, which are not subject to any unusual selective forces and which diversify slowly by the random accumulation of neutral or nearly neutral variation, should provide reliable information about the relationships among isolates. To avoid the chance association of alleles, these genes were also selected on the criterion that they are widely separated on the chromosome, except purK1 and gdh which are only 28.5 kb apart (Kleerebezem et al., 2003).

PCR was performed to amplify gene fragments from chromosomal DNA of the L. plantarum strains by using oligonucleotides listed in Table 2. The conditions of PCR, purification and sequencing of DNA fragments were described previously (De las Rivas et al., 2004).

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 an allele profile or sequence type (ST) derived from the combination of numbers corresponding to the alleles at the loci analysed (Table 1). Sequences which were different even at a single nucleotide site were considered distinct alleles.

Sequence alignments and comparison were done with the program BioEdit (http://jebrown.mbio.ncsu.edu/BioEdit/bioedit.html) (Hall, 1999) and converted into MEGA and NEXUS files with START (Sequence Type Analysis and Recombinatioral Tests; http://outbreak.ceid.ox.ac.uk/software.html). Phylogenetic trees were compiled with MEGA version 2.1 software (http://www.megasoftware.net). Dendrograms were constructed by the unweighted pair group method with arithmetic averages (UPGMA) with the Kimura two-parameter model. The percentage of bootstrap confidence levels for internal branches, as defined by the MEGA program (Kumar et al., 1994), was calculated from 1000 random resamplings.

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 (Huson, 1998). 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 two recombination tests, the index of association (Ia) (Maynard Smith et al., 1993) and the Sawyer’s run test, and a test for selection, the ratio of mean non-synonymous substitutions per non-synonymous site/mean synonymous substitutions per synonymous site, the dN/dS ratio (Nei & Gojobori, 1986). To avoid dependence on the number of loci, we also calculated the standardized Ia (Ia) (Haubold & Hudson, 2000).

RESULTS

Ribotyping and ISR RFLP analysis

Ribopattern analysis with EcoRI revealed four bands for all L. plantarum strains. The results are shown in Fig. 1(a). Four different groups of strains were defined on this basis: ribotype 1, containing L. plantarum CECT 220; ribotype 2, including L. plantarum type strain (CECT 7485); ribotype 3,ribotype 4.
represented only by *L. plantarum* RM40; and ribotype 4, containing *L. plantarum* CECT 4645 and RM35. The assignment of ribotype groups to each strain is listed in Table 1. Eleven out of 16 strains belonged to ribotype group 2.

Primers 16S14F and 23S1R, complementary to target sequences at about 140 nt from the 3' end of the 16S rRNA gene and about 120 nt from the 5' end of the 23S rRNA gene (Zavaleta et al., 1996), respectively, were used to amplify the ISR of the *L. plantarum* chromosome. Lanes: 1, ribotype 1; 2, ribotype 3; 4, ribotype 4. The sizes of the labelled fragments are indicated on the left. (b) 16S–23S rDNA ISR patterns obtained after *Alu* (lane 1), *Cfo* (2), *Dde* (3) and *TaqI* (4) digestion of the 550 bp DNA fragment PCR-amplified with primers 16S14f and 23S1R. The sizes of some fragments in the 50 bp molecular mass marker are indicated on the left.

**Fig. 1.** Typing analysis of the 16 *L. plantarum* strains examined in this study. (a) Ribotyping patterns of an *EcoRI* digest of *L. plantarum* chromosomal DNAs. Lanes: 1, ribotype 1; 2, ribotype 2; 3, ribotype 3; 4, ribotype 4. The sizes of the labelled fragments are indicated on the left. (b) 16S–23S rDNA ISR patterns obtained after *Alu* (lane 1), *Cfo* (2), *Dde* (3) and *TaqI* (4) digestion of the 550 bp DNA fragment PCR-amplified with primers 16S14f and 23S1R. The sizes of some fragments in the 50 bp molecular mass marker are indicated on the left.

**Table 3.** Sequence variation at six loci

<table>
<thead>
<tr>
<th>Gene</th>
<th>Fragment size (bp)</th>
<th>Mean G+C content (mol%)</th>
<th>No. of alleles</th>
<th>No. of polymorphic sites*</th>
<th>No. of nucleotide substitutions per nucleotide site</th>
<th>dS/dS†</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>pgm</em></td>
<td>558</td>
<td>43.3</td>
<td>3</td>
<td>6 (6)</td>
<td>0·0004</td>
<td>0</td>
</tr>
<tr>
<td><em>ddl</em></td>
<td>677</td>
<td>42.7</td>
<td>5</td>
<td>7 (6)</td>
<td>0·0008</td>
<td>4·78</td>
</tr>
<tr>
<td><em>gyrB</em></td>
<td>704</td>
<td>45.2</td>
<td>7</td>
<td>41 (41)</td>
<td>0·0056</td>
<td>0</td>
</tr>
<tr>
<td><em>purK1</em></td>
<td>525</td>
<td>48.1</td>
<td>8</td>
<td>12 (8)</td>
<td>0·0023</td>
<td>11·15</td>
</tr>
<tr>
<td><em>gdh</em></td>
<td>414</td>
<td>50.5</td>
<td>10</td>
<td>32 (29)</td>
<td>0·0072</td>
<td>2·07</td>
</tr>
<tr>
<td><em>mutS</em></td>
<td>594</td>
<td>47.0</td>
<td>8</td>
<td>7 (5)</td>
<td>0·0010</td>
<td>8·72</td>
</tr>
</tbody>
</table>

*Number of silent polymorphic sites in parentheses.
†Calculated by using the START program expressed as dS/dS ratio multiplied by 100.

**Variation at the MLST loci**

The sequences of the seven chosen loci were determined for the 16 strains, with the exception of the *tkt4* locus which could not be amplified from strains CECT 223, 224 and 4645. In the strains where the *tkt4* locus could be amplified, we sequenced a 567 bp fragment having six polymorphic sites that generate five different alleles (data not shown). Since the *tkt4* locus could not be amplified from all the *L. plantarum* strains, it was discarded from the MLST scheme.

The MLST scheme defined alleles between 414 (*gdh*) and 704 bp (*gyrB*) in length (Table 3). Between 3 (*pgm*) and 10 (*gdh*) alleles were found per locus (Table 3, Fig. 2). The sequence differences in each allele are shown in Fig. 2. The proportion of variable sites present in the MLST alleles ranged from 1·03 (*ddl*) to 7·72% (*gdh*). The mean G+ C content of the different gene fragments varied from 42·7 (*ddl*) to 50·5 mol% (*gdh*). Since the G+C content of the *L. plantarum* chromosome is 44·5 mol% (Kleerebezem et al., 2003), the *gdh* allele sequences showed an unexpectedly high G+C content.

The proportion of nucleotide alterations that changed the amino acid sequence (non-synonymous substitutions, dS) and the proportion of silent changes (synonymous substitutions, dS) were calculated for each gene. With these data, the dS/dS ratios were calculated for all loci and all were <1 (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.

A UPGMA dendrogram showing the genetic relatedness among *L. plantarum* strains investigated in this study is shown in Fig. 3. There were no significant clusters that correlated with the geographic origin of the strains. All strains differed in various loci, except strains RM71 and RM72, and CECT 223 and CECT 224, that shared the same ST (Table 1, Fig. 3).
Relationship of \textit{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 differed only in the \textit{gdh} locus. Concatenated \textit{pgm}, \textit{ddl}, \textit{gyrB}, \textit{purK1}, \textit{gdh} and \textit{mutS} gene sequence fragments were analysed and graphically displayed with SPLITSTREE (Fig. 4). 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 interconnecting network or a parallelogram will appear whenever recombination has been involved in the evolution of the analysed gene. Fig. 5 shows the split graphs for all alleles of the six fragments analysed. 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 \textit{purK1} and \textit{gdh} loci present network-like or parallelogram 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.

Sawyer’s test revealed no detectable cases of intragenic recombination in the sample, except in the case of the gdh locus in which there was evidence against the null hypothesis of no recombination ($P = 0.02$) (Table 4). This evidence disappeared when the maximum condensed fragment value ($P = 1.00$) was considered.

Linkage disequilibrium between alleles was estimated with $I^A$, a statistic that is expected to be zero when the alleles are in linkage equilibrium (free recombination). The latter probably means that the distribution of alleles occurs independently from each other. The $I^A$ value was 0.139, indicating that the genes investigated are close to linkage equilibrium. This low $I^A$ value is also indicative of extensive recombination.

DISCUSSION

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

Ribotyping is a both an inter- and an intraspecific typing method (Manfreda et al., 2005; Zadoks et al., 2005) that has been successfully used to differentiate between strains of the same Lactobacillus species. Yansanjav et al. (2003) used ribotyping in a study of five L. plantarum strains. These five strains formed two ribotype patterns and the authors found that strains isolated from different breweries shared the same ribotype. In other studies, based on these ribotypes, apparently identical pairs of L. plantarum strains were isolated in very different settings: L. plantarum ATCC 14917T (CECT 748T) was purchased by the authors from ATCC and L. plantarum UH 2153 was isolated from the vagina of a woman in Ontario (Zhong et al., 1998). In our study we found four different ribotypes; however, most of the strains (11/15) belonged to the same ribotype, ribotype 2. These ribotype 2 strains are not source-, time- or geographically related; therefore, more specific identification methods, such as MLST would be preferable.

In addition, the variation in length and sequence of the 16S–23S rDNA ISRs has been used to type strains. ISRs usually show variations which make it possible to discriminate among strains within some species. As shown in Table 1 and Fig. 1, all the L. plantarum strains analysed in this study shared the same RFLP ISR pattern. Therefore, the RFLP ISR results support the idea of the need for a more discriminating method able to differentiate L. plantarum strains.

As a first step for developing an MLST typing method, we analysed the sequence diversity of seven housekeeping genes to ascertain if they were sufficient to provide 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 data, we cannot exclude the possibility of a tkt4 gene deletion or the presence of a non-homologous gene copy in these strains.

The number of alleles of the six housekeeping loci ranged from 3 to 10 (Table 3). Our results corroborate the previously described genetic heterogeneity of this species (Vescovo et al., 1993; Sánchez et al., 2004; Molenaar et al., 2005). The six loci were polymorphic and most of the types were represented by a single strain. We only found two exceptions. One of them was found in 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). In the three typing methods used in this study, both isolates were found to be 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 in all the typing methods used in this study. The number of loci used can be increased to improve resolution, but there comes a point when it is not worth the reward because little additional information will be obtained, for epidemiological

Table 4. Sawyer’s test analysis for evidence of intragenic recombination

The Sawyer’s test analysis was carried out by using the START program. Results were obtained from 10,000 random resamplings. SSCF, Sum of the square of condensed fragments; MCF, maximum condensed fragment.

<table>
<thead>
<tr>
<th>Locus</th>
<th>SSCF (P)</th>
<th>MCF (P)</th>
</tr>
</thead>
<tbody>
<tr>
<td>pgm</td>
<td>26 (0.67)</td>
<td>5 (0.67)</td>
</tr>
<tr>
<td>ddl</td>
<td>98 (0.50)</td>
<td>5 (0.83)</td>
</tr>
<tr>
<td>gyrB</td>
<td>7127 (0.33)</td>
<td>35 (0.52)</td>
</tr>
<tr>
<td>purK1</td>
<td>937 (0.80)</td>
<td>8 (1.00)</td>
</tr>
<tr>
<td>gdh</td>
<td>8269 (0.02)</td>
<td>29 (1.00)</td>
</tr>
<tr>
<td>mutS</td>
<td>62 (1.00)</td>
<td>4 (1.00)</td>
</tr>
</tbody>
</table>

Fig. 5. 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 connected to each other by multiple pathways, forming parallelogram structures, is suggestive of recombination. All branch lengths are drawn to scale. The numbering refers to allele numbers.
purposes, compared to the cost and effort involved (Urwin & Maiden, 2003). The proposed MLST scheme showed highly discriminating power since it was able to differentiate between highly similar strains. This was the case for strains CECT 748 and 749 that were identical in five out of the six genes analysed, differing only in the \( \text{gdh} \) locus (Table 1).

Using this typing method we could not find an association between ST and the source from which the different strains were isolated (Table 1). The absence of such an association may be due to the versatility of individual \( L. \) \textit{plantarum} strains, which can survive or even grow in different environments. Similar results were found by Molenaar \textit{et al.} (2005) when they explored \( L. \) \textit{plantarum} genome diversity by using microarrays.

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, for example, the appearance of a network of relationships among sequences rather than a bifurcating tree-like phylogeny. The split decomposition analysis of \( L. \) \textit{plantarum} strains (Fig. 4) reveals two uncentred edges, suggesting that the evolution of these strains stems from a couple of strains from which single branches radiate.

The most simple method to detect recombination in aligned sequences is to look for mosaic structures by eye. Significant mosaic structure is indicative of recombinatorial exchange, usually among isolates of the same species (Feil \textit{et al.}, 2000). In our study, four possible examples of recombination events were found: allele 3 of \( \text{gyrB} \), allele 6 of \( \text{gdh} \), allele 8 of \( \text{purK1} \) and allele 3 of \( \text{pgm} \). The mean divergence for allele 3 of \( \text{gyrB} \) (3-97%) is much higher than the mean diversity within the other \( \text{gyrB} \) alleles (0-57%) (Fig. 2). This divergence is similar to the divergence observed in a possible example of an intergenic recombinatorial event in the \( \text{gdh} \) gene between \( \text{Streptococcus pneumoniae} \) and \( \text{Streptococcus mitis} \) (Enright & Spratt, 1998). The alleles in the other loci showed a divergence of 4-11 versus 1-45% (allele 6 of \( \text{gdh} \)), 1-90 versus 0-33% (allele 8 of \( \text{purK1} \)) and 0-29 versus 0-09% (allele 3 of \( \text{pgm} \)). The wide range of variable sites found in the \( L. \) \textit{plantarum} loci analysed (1% in \( \text{dld} \) to 7-7% in \( \text{gdh} \)) reflects variation due to point mutations through possible horizontal transfer gene events. Recently, and mainly based on unusual base composition, horizontal transfer events from closely related organisms or even the same species have been proposed for \( L. \) \textit{plantarum} strains (Molenaar \textit{et al.}, 2005). The same authors have postulated that lifestyle adaptation would encompass the frequent acquisition and loss of genes involved in adaptation to niche functions.

The utility of MLST for the analysis of the genetic structure of bacterial populations is mainly based on the characteristic of housekeeping genes to have selectively neutral variability. Analysis of synonymous and non-synonymous changes in the allele sequences of a locus can be used to determine if they are subject to positive selection, so a \( d_S/d_K \) ratio of greater than 1 implies selection for amino acid changes. In our genetic analysis, the six housekeeping loci had \( d_S/d_K \) ratios significantly lower than 1 (Table 3). Another important characteristic in relation to this is the location of the loci on the chromosome. With the exception of the \( \text{purK1} \) and \( \text{gdh} \) loci, the loci in this study were distant enough to make 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 strains CECT 223 and 224, and strains RM71 and 72 share the same ST. The analysis of the \( L. \) \textit{plantarum} population structure presented here suggests a panmictic, non-clonal population structure with substantial recombination. The split decomposition analysis provides strong evidence that intraspecies recombination occurs frequently in \( L. \) \textit{plantarum} and plays a major role in generating genetic heterogeneity among strains. A low \( I_{S/S} \) value (0-138) is indicative of a weak clonal population and also confirms the importance of recombination in \( L. \) \textit{plantarum} and supports the estimation that the genes investigated in \( L. \) \textit{plantarum} are close to linkage equilibrium. The extension of the present analysis to a larger number of isolates could contribute to improved knowledge about the structure of \( L. \) \textit{plantarum} populations.

In conclusion, the MLST scheme presented here will be a useful new tool for the precise and unambiguous characterization of \( L. \) \textit{plantarum} isolates and appears to have sufficient discriminatory power for population investigations. It should allow comparison of the isolates of this species by a precise typing procedure.

**ACKNOWLEDGEMENTS**

This work has been supported by grants AGL2005-00470 and RM03-002. We thank R. González and E. García for their critical reading of the manuscript. The technical assistance of M. V. Santamaría is greatly appreciated. We also thank A. Hexter for correcting the English version. B. de las Rivas is a recipient of a postdoctoral fellowship and A. Marcobal of a predoctoral fellowship both from the Comunidad de Madrid.

**REFERENCES**


Development of MLST method for *L. plantarum*


