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

Blanca de las Rivas, Ángela Marcobal and Rosario Muñoz

Correspondence
Rosario Muñoz
rmunoz@ifi.csic.es

Departamento 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.

Received 31 August 2005
Revised 18 October 2005
Accepted 19 October 2005

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).

Abbreviations: d_N , non-synonymous substitutions; d_S , synonymous substitutions; I_A^S , standardized index of association; ISR, intergenic spacer region; MLST, multilocus sequence typing; RFLP, restriction fragment length polymorphism; ST, sequence typing.

The GenBank/EMBL/DDBJ accession numbers for the sequences reported in this paper are AJ966402–AJ966404 (*pgm*), AJ966405–AJ966409 (*ddl*), AJ966364–AJ966370 (*gyrB*), AJ966371–AJ966378 (*purK1*), AJ966379–AJ966388 (*gdh*), AJ966389–AJ966396 (*mutS*) and AJ966397–AJ966401 (*tkf4*).

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

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 variations of restriction 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 748^T 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

RT, Ribotype; ITS, 16S–23S ISR type; ST, sequence type; ND, no data available.

Strain no.	Strain	RT	ITS	ST	Allele no. at locus:						Source of isolate		
					<i>pgm</i>	<i>ddl</i>	<i>gyrB</i>	<i>purK1</i>	<i>gdh</i>	<i>mutS</i>	Origin	Country	Year
1	WCFS1	ND	ND	1	1	1	1	1	1	1	Human saliva	UK	1956
2	CECT 220 (ATCC 8014)	1	1	2	1	3	1	2	6	5	Corn silage	ND	1948
3	CECT 221 (ATCC 14431)	2	1	3	1	1	6	3	8	1	Grass silage	ND	1960
4	CECT 223	2	1	4	1	5	3	4	7	1	ND	Pamplona, Spain	1987
5	CECT 224	2	1	4	1	5	3	4	7	1	ND	Pamplona, Spain	1987
6	CECT 748 ^T (ATCC 14917)	2	1	5	3	2	1	2	1	1	Pickled cabbage	Denmark	1919
7	CECT 749 (ATCC 10241)	2	1	6	3	2	1	2	3	1	Pickled cabbage	ND	1955
8	CECT 4185 (NCBF 1193)	2	1	7	3	1	2	1	5	1	Silage	ND	1958
9	CECT 4645 (NCBF 965)	4	1	8	2	4	7	8	10	4	Cheese	ND	1958
10	RM28	2	1	9	1	2	1	7	7	3	Wine	Valladolid, Spain	2000
11	RM35	4	1	10	3	2	1	2	2	8	Wine	Toledo, Spain	1998
12	RM38	2	1	11	3	1	4	2	4	2	Wine	Toledo, Spain	1998
13	RM40	3	1	12	1	1	1	5	9	1	Wine	Toledo, Spain	1998
14	RM71	2	1	13	3	1	5	6	5	6	Wine	Valladolid, Spain	2001
15	RM72	2	1	13	3	1	5	6	5	6	Wine	Valladolid, Spain	2001
16	RM73	2	1	14	3	2	1	2	1	7	Wine	Madrid, Spain	2000

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 16S14F 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 WCFS1 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 (*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 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://jwbrown.mbio.ncsu.edu/BioEdit/bioedit.html>) (Hall, 1999) and converted into MEGA and NEXUS files with START (Sequence Type Analysis and Recombinatorial 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 (I_A) (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 d_N/d_S ratio (Nei & Gojobori, 1986). To avoid dependence on the number of loci, we also calculated the standardized I_A (I_A^S) (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 748^T); ribotype 3,

Table 2. Primers used for MLST

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

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 *AluI* (lane 1), *CfoI* (2), *DdeI* (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.

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* strains. A PCR product of approximately 550 bp was obtained for each strain, indicating that the ISR is highly similar in all strains analysed (data not shown). PCR products of the 16S–23S rDNA ISR were digested with *AluI*, *CfoI*, *DdeI* and *TaqI*. All the strains analysed showed identical RFLP patterns for each enzyme (Fig. 1b).

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, d_N) and the proportion 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 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).

Table 3. Sequence variation at six loci

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

*Number of silent polymorphic sites in parentheses.

†Calculated by using the START program expressed as d_N/d_S ratio multiplied by 100.

gyrB					
00000000000000000000000000000000111111111111 6666677777777777888999999999000000001111122 13366011256788914900336777800112486677802 83603217732769896737039058725479401836861 consensus CGATCTTTCTGGTGATCCTTTTTCCTCTGTGCTTAGGA					
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					
000000000000000000000000000000111 mutS 1111111					
6666777777778888888888889999999900 5677888					
66890112568344566778893345568125 6101004					
36485470383109245062840924568160 0159398					
consensus CAAGCTTTGTCTCGCGCATCACATGTGCGCT consensus CATCGTG					
allele 1 (3)T.....T..... allele 1 (8)					
allele 2 (1)T.....T..T..... allele 2 (1)A					
allele 3 (1) T.G....AC.T..A..GC.G..... allele 3 (1) ...T...					
allele 4 (1) T.G...C.AC.T..A.A.GC.G..... allele 4 (1) ...T.C.					
allele 5 (3) T.G...C.AC.T..A..GC.G..... allele 5 (1) ...TAC.					
allele 6 (1) T.G...C.CAC...A..GC.GTGCACG...A allele 6 (2) T...AC.					
allele 7 (3)A.....T.T....A... allele 7 (1) ..A....					
allele 8 (1) .G.....A.....T.T....A... allele 8 (1) .TA....					
allele 9 (1)T.....G..T.....					
allele 10 (1) ...A.....G..T.....TTC					
ddl					
3356777 purK1 555556777789 pgm 566788					
2647389 022798444568 637829					
4363298 022798444568 460951					
consensus GAGCCGG consensus GTCAGATTGCCCT consensus CACGTT					
allele 1 (7) allele 1 (2) allele 1 (8)					
allele 2 (5)A. allele 2 (6)T..... allele 2 (7) T.....					
allele 3 (1)A allele 3 (1)C.... allele 3 (1) .GTACC					
allele 4 (1) CG..T.. allele 4 (2) .C....C....					
allele 5 (2) ..AT... allele 5 (1)CA...					
allele 6 (2)GC....					
allele 7 (1)GGC...G					
allele 8 (1) AC.GAGGC.TTG					

For all genes, multiple strains carried identical alleles. However, only two pairs of strains had identical alleles 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 *gdh* locus. Concatenated *pgm*, *ddl*, *gyrB*, *purK1*, *gdh* and *mutS* gene sequence fragments were analysed and graphically displayed with SPLITSTREE (Fig. 4). The

Fig. 3. UPGMA dendrogram showing the genetic relatedness of the *L. plantarum* strains examined in this study. Phylogenetic and molecular evolutionary analyses were conducted with the MEGA program (version 2.1) by the UPGMA method. Bootstrap confidence intervals are shown.

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^S , 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^S value was 0.139, indicating that the genes investigated are close to linkage equilibrium. This low I_A^S 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)

Fig. 4. Split decomposition analysis based on the allelic profiles of the 16 *L. plantarum* strains examined in this study. The numbering refers to strain numbers.

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 14917^T (CECT 748^T) 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

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.

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,

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.

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

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

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^T and 749 that were identical in five out of the six genes analysed, differing only in the *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. plantarum* strains, which can survive or even grow in different environments. Similar results were found by Molenaar *et al.* (2005) when they explored *L. 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. 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 *et al.*, 2000). In our study, four possible examples of recombination events were found: allele 3 of *gyrB*, allele 6 of *gdh*, allele 8 of *purK1* and allele 3 of *pgm*. The mean divergence for 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 in a possible example of an intergenic recombinatorial event in the *gdh* gene between *Streptococcus pneumoniae* and *Streptococcus mitis* (Enright & Spratt, 1998). The alleles in the other loci showed a divergence of 4.11 versus 1.45 % (allele 6 of *gdh*), 1.90 versus 0.33 % (allele 8 of *purK1*) and 0.29 versus 0.09 % (allele 3 of *pgm*). The wide range of variable sites found in the *L. plantarum* loci analysed (1 % in *ddl* to 7.7 % in *gdh*) reflects variation due to point mutations through to 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. plantarum* strains (Molenaar *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_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 is the location of the loci on the chromosome. With the exception of the *purK1* and *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. 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. plantarum* and plays a major role in generating genetic heterogeneity among strains. A low I_A^S value (0.138) is indicative of a weak 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 larger number of isolates could contribute to improved knowledge about the structure of *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. 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

- Abazinge, M. D. A., Fontenot, J. P., Allen, V. G. & Flick, G. J. (1993). Ensiling characteristics of crab waste and wheat straw treated with different additives. *J Agric Food Chem* **41**, 657–661.
- De las Rivas, B., Marcobal, A. & Muñoz, R. (2004). Allelic diversity and population structure in *Oenococcus oeni* as determined from sequence analysis of housekeeping genes. *Appl Environ Microbiol* **70**, 7210–7219.
- Dingle, K. E., Colles, F. M., Wareing, D. R. A. & 7 other authors (2001). Multilocus sequence typing system for *Campylobacter jejuni*. *J Clin Microbiol* **39**, 14–23.

- Elegado, F. B., Guerra, M. A. R. V., Macayan, R. A., Mendoza, H. A. & Lizaran, M. B. (2004). Spectrum of bacteriocin activity of *Lactobacillus plantarum* BS and fingerprinting by RAPD-PCR. *Int J Food Microbiol* **95**, 11–18.
- Enright, M. C. & Spratt, B. G. (1998). A multilocus sequence typing scheme for *Streptococcus pneumoniae*: identification of clones associated with serious invasive disease. *Microbiology* **144**, 3049–3060.
- Farfán, M., Miñana-Galbis, D., Fusté, M. C. & Lorén, J. G. (2002). Allelic diversity and population structure in *Vibrio cholerae* O139 Bengal based on nucleotide sequence analysis. *J Bacteriol* **184**, 1304–1313.
- Feil, E. J., Enright, M. C. & Spratt, B. G. (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.
- Hall, T. A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98NT. *Nucleic Acids Symp Ser* **41**, 95–98.
- Haubold, B. & Hudson, R. R. (2000). LIAN 3.0: detecting linkage disequilibrium in multilocus data. *Bioinformatics* **14**, 68–73.
- Helgason, E., Tourasse, N. J., Meisal, R., Caugant, D. A. & Kolsto, A. B. (2004). Multilocus sequence typing scheme for bacteria of the *Bacillus cereus* group. *Appl Environ Microbiol* **70**, 191–201.
- Herias, M. V., Hessle, C., Telemo, E., Midtvedt, T., Hanson, L. A. & Wold, E. (1999). Immunomodulatory effects of *Lactobacillus plantarum* colonizing the intestine of gnotobiotic rats. *Clin Exp Immunol* **116**, 283–290.
- Huson, D. H. (1998). SplitsTree: analyzing and visualizing evolutionary data. *Bioinformatics* **14**, 68–73.
- Johansson, M.-L., Quednau, M., Molin, G. & Åhrné, S. (1995). Randomly amplified polymorphic DNA (RAPD) for rapid typing of *Lactobacillus plantarum* strains. *Lett Appl Microbiol* **21**, 155–159.
- Kleerebezem, M., Boekhorst, J., van Kranenburg, R. & 17 other authors (2003). Complete genome sequence of *Lactobacillus plantarum* WCFS1. *Proc Natl Acad Sci U S A* **100**, 1990–1995.
- Kumar, S., Tamura, K. & Nei, M. (1994). MEGA: molecular evolutionary genetics analysis software for microcomputers. *Comput Appl Biosci* **10**, 189–191.
- Manfreda, G., de Cesare, A., Stella, S., Cozzi, M. & Cantoni, C. (2005). Occurrence and ribotypes of *Listeria monocytogenes* in Gorgonzola cheeses. *Int J Food Microbiol* **102**, 287–293.
- Marchesi, J. R., Sato, T., Weightman, A. J., Martin, T. A., Fry, J. C., Hiom, S. J. & Wade, W. G. (1998). Design and evaluation of useful bacterium-specific PCR primers that amplify genes coding for bacterial 16S rRNA. *Appl Environ Microbiol* **64**, 795–799.
- Maynard Smith, J., Smith, N. H., O'Rourke, M. & Spratt, B. G. (1993). How clonal are bacteria? *Proc Natl Acad Sci U S A* **90**, 4384–4388.
- Merry, R. J., Dhanoa, M. S. & Theodorou, M. K. (1995). Use of freshly cultured lactic acid bacteria as silage inoculants. *Grass For Sci* **50**, 112–123.
- Molenaar, D., Bringel, F., Schuren, F. H., de Vos, W. M., Siezen, R. J. & Kleerebezem, M. (2005). Exploring *Lactobacillus plantarum* genome diversity by using microarrays. *J Bacteriol* **187**, 6119–6127.
- Nei, M. & Gojobori, T. (1986). Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol Biol Evol* **3**, 418–426.
- Oneca, M., Irigoyen, A., Ortigosa, M. & Torre, P. (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.
- Rodas, A. M., Ferrer, S. & Pardo, I. (2005). Polyphasic study of wine *Lactobacillus* strains: taxonomic implications. *Int J Syst Evol Microbiol* **55**, 197–207.
- Ruiz-Barba, J. L., Cathcart, D. P., Warner, P. J. & Jiménez-Díaz, R. (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.
- Sánchez, I., Seseña, S. & Palop, L. L. (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.
- Urlings, H. A. P., Bijker, P. G. H. & van Logtestijn, J. G. (1993). Fermentation of raw poultry by-products for animal nutrition. *J Anim Sci* **71**, 2420–2426.
- Urwin, R. & Maiden, M. C. J. (2003). Multi-locus sequence typing: a tool for global epidemiology. *Trends Microbiol* **11**, 479–487.
- Vaquero, I., Marcobal, A. & Muñoz, R. (2004). Tannase activity by lactic acid bacteria isolated from grape must and wine. *Int J Food Microbiol* **96**, 199–204.
- Vescovo, M., Torriani, S., Dellaglio, F. & Bottazi, V. (1993). Basic characteristics, ecology and application of *Lactobacillus plantarum*: a review. *Ann Microbiol Enzymol* **43**, 261–284.
- Yansanjav, A., Svec, P., Sedlacek, I., Hollerova, I. & Nemec, M. (2003). Ribotyping of lactobacilli isolated from spoiled beer. *FEMS Microbiol Lett* **229**, 141–144.
- Zadoks, R. N., Tikofsky, L. L. & Boor, K. J. (2005). Ribotyping of *Streptococcus uberis* from a dairy's environment, bovine feces and milk. *Vet Microbiol* **109**, 257–265.
- Zavaleta, A. I., Martínez-Murcia, A. J. & Rodríguez-Varela, F. (1996). 16S-23S rDNA intergenic sequences indicate that *Leuconostoc oenos* is phylogenetically homogeneous. *Microbiology* **142**, 2105–2114.
- Zhong, W., Millsap, K., Bialkowska-Hobrazanska, H. & Reid, G. (1998). Differentiation of *Lactobacillus* species by molecular typing. *Appl Environ Microbiol* **64**, 2418–2423.