# Complete nucleotide sequence and structural organization of pPB1, a small *Lactobacillus plantarum* cryptic plasmid that originated by modular exchange

B. de las Rivas, A. Marcobal, and R. Muñoz*

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

*Corresponding author. Tel.:+34-91-562 2900; fax:+34-91-564 4853

E-mail address: rmunoz@ifi.csic.es (R. Muñoz)

**Abstract**

A small cryptic plasmid designated pPB1 was isolated from *Lactobacillus plantarum* BIFI-38 and its complete 2,899 bp nucleotide sequence was determined. Sequence analysis revealed four putative open reading frames. Based on sequence analysis two modules could be identified. First, the replication module consisted of a sequence coding for a replication protein (RepB) and its corresponding target site, and two putative repressor proteins (RepA and RepC). Sequence analysis indicated the possible synthesis of an antisense RNA that might regulate RepB production. A putative lagging-strand initiation site was also found, suggesting that pPB1 replicates via a rolling circle mechanism. The second module of pPB1 consisted of a sequence coding for a putative mobilization protein and its corresponding *oriT* site. Since the nucleotide sequence of the replication module showed 94.5% identity to the similar region on the *Leuconostoc lactis* plasmid pCI411, and the nucleotide sequence of the mobilization module had 97.5% identity to *L. plantarum* plasmid pLB4, it is concluded that pPB1 originated by modular exchange between two such plasmids by homologous recombination. Putative recombination sites where crossover might have taken place were also identified.

***Keywords:*** cryptic plasmid; *Lactobacillus plantarum*; modular exchange; replication; mobilization; homologous recombination

## 1.Introduction

*Lactobacillus plantarum* is a species of considerable industrial and medical interest. It is widely used as starter microorganism to stimulate lactic acid fermentation in grass silage, meat, and vegetables as well as to enhance malolactic fermentation in wine (Buckenhüskes, 1993). Lactobacilli have been used in food production for centuries and, because of their industrial significance, have become the focus of a rapidly increasing number of genetic studies, especially the application of recombinant DNA techniques for the genetic modification of *Lactobacillus* strains.

Only a small fraction of the *Lactobacillus* plasmids that have been described in the literature have been sequenced. These sequences provide insights into plasmid-borne functions and genetic exchange in lactobacilli, and identify replicons that can potentially be used for the development of genetic tools (Alpert et al., 2003). Plasmids (1.5 to 60 kb in size) are frequently found in various strains of *Lactobacillus* and the mode of replication of some of them has been studied in detail. DNA sequence analysis of small multicopy replicons from *L. plantarum* (Bates and Gilbert, 1989; Daming et al., 2003), *L. pentosus* (Leer et al., 1992), *L. acidophilus* (Sano et al., 1997), *L. sakei* (Alpert et al., 2003), and *L. hilgardii* (Josson et al., 1990) has shown that they belong to a family of highly interrelated plasmids replicating via a similar mechanism, termed the rolling circle (RC) method, which is observed in a number of gram-positive multicopy plasmids (del Solar et al., 1993; Khan et al., 1997).

In this communication we present the nucleotide sequence of pPB1. By sequence analysis, we have identified the genetic determinants required for plasmid replication and

mobilization. Moreover, the complete sequence will facilitate determination of gene function by experimental methods and should be an important reference for comparative analysis with other plasmids from lactobacilli.

## 2.Materials and methods

### *2.1.Bacterial strains, plasmids and growth conditions*

*Lactobacillus plantarum* strains were grown in static MRS liquid cultures at 30 ºC. All the *L. plantarum* strains analysed belong to the bacterial culture collection from the Instituto de Fermentaciones Industriales (BIFI), except the *L. plantarum* type strain (*L. plantarum* ATCC 14917$^T$), which was purchased from the Spanish Type Culture Collection (CECT). The *L. plantarum* BIFI strains (*L. plantarum* BIFI-28, BIFI-31, BIFI-34, BIFI-35, BIFI-38, BIFI-39, BIFI-40, BIFI-41, BIFI-71. BIFI-72, BIFI-73) were isolated from grape must and wine (Moreno-Arribas et al., 2003). *L. plantarum* BIFI-38 has been described previously and contained plasmid pPB1.

The *Escherichia coli* strain DH5α (Sambrook et al., 1989) was used as host for recombinant plasmids and preparation of sequencing templates. *E .coli* was grown in Luria-Bertani medium at 37 ºC by vigorous shaking (Sambrook et al., 1989). The respective media were solidified by the addition of 1.5% agar for plating. Where appropiate, ampicillin was added at 100 μg/ml.

Plasmid pUC19 (New England Biolabs, Inc.) was used for subcloning pPB1 fragments. Chromosomal DNA, plasmid purification and transformation of *E.coli* were carried out as described elsewhere (Muñoz et al., 1998).

*2.2.DNA manipulations and hybridization*

Restriction endonucleases, T4 DNA ligase and the Klenow fragment of DNA polymerase were obtained commercially and used according to the recommendations of the suppliers. Gel electrophoresis of plasmids, restriction fragments, and PCR products were carried out in agarose gels as described (Sambrook et al., 1989). DNA was digoxigenin-labeled and chemiluminescently detected by using the DIG High pure DNA labelling and detection Starter Kit (Roche) according to the manufacturer´s instructions.

PCR amplifications were performed as previously described by using *Pfu* DNA polymerase (Stratagene) (Muñoz et al. 1998). The oligonucleotide primers mentioned in the text were: 76 (5´-CTAGTCACCGCCCTTTGTTCC), 78 (5´-GGATAGAAAGCCCACTGCATG), 103 (5´-AGGCTTATGCCGATTAGAGCG) and 104 (5´-GAGCCGGGCCAGGCTCGCGTG).

*2.3.Sequence analysis of pPB1*

A 1.3 kb *Hin*dIII fragment of pPB1 was subcloned into pUC19 and used as the starting point for sequence determination. Based on this sequence, oligonucleotide primers 76 and 78 were designed and used to amplify the remaining plasmid sequence by PCR. The primer-walking method completed the sequence of pPB1. Sequencing reactions were

performed with the ABI dye terminator sequencing reagents by using an ABI Prism 377™ DNA sequencer (Applied Biosystems, Inc.).

Sequence similarity searches were carried out using BLAST (Altschul et al., 1977) on the EMBL/GenBank databases. Computer promoter predictions were carried out at the Internet site http://www.fruitfly.org/seq_tools/promoter.html. Signatures were analysed on the EXPASY (http://www.expasy.ch) site and multiple alignment was done using CLUSTAL W at the EBI site (http://www.ebi.ac.uk) after retrieval of sequences from GenBank and Swiss-Prot. Phylogenetic trees and RNA secondary structure predictions were carried out by the GeneBee program (http://www.genebee.msu.su/genebee.html).

The complete DNA sequence of pPB1 has been deposited in DDBJ/GenBank/EMBL under accession number **AJ716330**.

## 3.Results and discussion

### *3.1.DNA sequence and organization of pPB1*

The plasmid profile of *L. plantarum* BIFI-38 revealed the presence of a small plasmid, designated pPB1. Its size was estimated to be approx. 3 kb by restriction fragment analysis. This was confirmed by the determination of the complete circular sequence of 2,899 nucleotides. Overall, the GC content of the plasmid (37.4%) was lower than that of the chromosome of *L. plantarum* (44.5 %) (Kleerebezem et al., 2003) but was within the range reported for other small plasmids from *L. plantarum* (Bates et al., 1989; Daming et al., 2003). A low G+C content was also reported when *L. plantarum* WCFS1 was

sequenced; this strain contained two small cryptic plasmids and a large plasmid having a slightly lower G+C content than the overall genome (Kleerebezem et al., 2003).

In pPB1 four putative ORFs covering 70% of the plasmid were found. All the ORFs were located in the same orientation.

To attribute a function to the deduced products, they were compared to the databases (Table 1). We could identify a region (*repA*, *repB* and *repC*) that we assumed was involved in plasmid replication by a RC mechanism. Highest similarities for these gene products were observed with plasmids from lactic acid bacteria. The G+C content of these putative replication genes was lower (from 27.33 to 33.17%) than the other parts of the pPB1 plasmid (Table 1). A second region was identified, comprising the *mob* gene, and was assumed to be involved in plasmid mobilization.

*3.2.Replication module*

Analysis of the nucleotide sequence suggested that the *repABC* genes were organized as a single operon, with typical ribosome binding sites located upstream of each gene. A putative promoter was detectable upstream of *repA* (nt 397-424). This promoter regions lies within a stem-loop structure with a calculated $\Delta G$ of –25 kcal/mol. It has been proposed that inverted repeat structures overlapping Rep promoters could constitute a target site for auto-regulation of Rep protein expression in the theta replicating plasmid pIP501 (Brantl and Behnke, 1992). Two putative transcription terminators followed the TAA stop codon of *repC*.

The molecular sizes of the putative RepA, RepB and RepC proteins were 6.2 kDa, 22.6 kDa and 5.3 kDa, respectively. Comparison of these proteins with the databases

revealed significant similarity to proteins from Gram-positive replicons. RepA showed significant identity to proteins involved in the regulation of plasmid transcription (Table 1). These proteins share several features: (i) they are small proteins (45-55 amino acid residues), (ii) they have two conserved K residues at the N-terminal end, and (iii) they show a putative α-helix turn α-helix (HTH) motif, typical of many DNA-binding proteins.

RepB showed 97% identity to the replication initiation protein RepB from the *Leuconostoc lactis* plasmid pCI411 (Coffey et al., 1994) (Table 1). An alignment showing the RepB protein from pCI411 and related proteins is represented in a previous study (Takamatsu et al., 2000) and also in RCR Group II on the RCR Replicon Database (http://www.essex.ac.uk/bs/staff/osborn/DPR/DPR_RCRIIalign.htm).

RepC showed significant identity to RepC of the *Leuconostoc lactis* plasmid pCI411 and ORFD of the *Lactococcus lactis* plasmid pWVO1 (Table 1). ORFD of pWVO1 is believed to encode the repressor protein regulating synthesis of the replication initiation protein (Leenhouts et al., 1991). There was no apparent similarity between RepC and other protein sequences in the databases.

RC plasmids can be classified into a least five families based on sequence similarity of their Rep proteins and leading- or double-strand origins (*dso*) (reviewed by Khan, 1997). On this basis, pPB1 belongs to Group II of the RC plasmids (http://www.essex.ac.uk/bs/staff/osborn/DPR/DPR_RCRIIalign.htm). Replication of the leading strand of RC plasmids initiates from its *dso* and proceeds in a unidirectional manner. The 22-bp sequence, 5′-GGGGGGTACTACGACACCCCCC-3′, which was homologous to the *dso* of several RC plasmids as pCI411, pE194, pLS1, and pWVO1, was identified 200 bp upstream of the *repA* start codon. It is postulated to be the *dso* of pPB1, the inverted repeat where the RepB protein introduces the nick to initiate replication.

Replication of the lagging strand of RC plasmids initiates from their single-strand origins (*sso*). The *sso* is generally located a short distance upstream of the *dso* (Khan et al, 1997). Plamid pPB1 presents a palindromic structure of 116 bp in length at position 2837 to 54, with the potential to form hairpin secondary structure with a $\Delta G$ of –29.5 Kcal/mol, suggesting it to be the putative *sso*. Two additional stem-loop structures could be found at positions 69 to107 and 132 to 152. This 215 bp putative *sso* was aligned, using FASTA, with sequences available from the GenBank database. A high degree of identity (96.7%) was found with a region of the *L. plantarum* plasmid pLB4, suggesting the presence of a putative *sso* lagging-strand origin in this plasmid. In addition, the *Leuconoctoc lactis* plasmid pCI411 had 78.5% identity in a 215 nt overlap (Coffey et al, 1994), a small plasmid from *L. curvatus* had 85.9% identity in a 135 nt region (Klein et al, 1993), and *L. helveticus* plasmid pLH2 showed a 78.1% identity in a 178 nt overlap (Pridmore et al., 1994). Note that *sso* lagging-strand origins usually have a very narrow range of hosts, however, the ability of an *sso* to function in various hosts may contribute to plasmid promiscuity and horizontal transfer among related bacteria (Khan et al, 1997).

Bacterial plasmids exert a tight control over their replicative events to enable their stable maintenance. In the pMV158 plasmid family, control of replication results from the combined repression exerted by two gene products, both acting within the same regulatory circuit. First, synthesis of *rep* mRNA is repressed at the transcriptional level by the binding of Cop protein to its operator, which is located within or in the vicinity of the *rep* promoter (del Solar et al., 1993). As already described, the pPB1 RepA protein may be a transcriptional repressor protein. The second level of control is exerted at the translational level. The synthesis of RepB is regulated by direct pairing of an antisense RNA molecule with the polycistronic *rep* mRNA (del Solar et al., 1993). In pPB1, we identified the

sequence (nt 601-669) that was complementary to the leader region of *repABC* mRNA and showed similarity to the antisense RNA of related plasmids. Computer promoter predictions revealed a sequence (nt 675-703) that could function as a promoter for the synthesis of this molecule. Downstream from the promoter region, inverted repeats followed by a strech of Ts were present, forming a possible transcription terminator.

These observations suggest that replication of pPB1 is controlled by the RepA and an antisense RNA. However, an additional gene, *repC* was identified downstream of *repB*. The RepC protein showed significant identity to a pWVO1 protein that is believed to encode a repressor protein that regulates synthesis of the replication initiation protein. This gene could be inactivated without affecting plasmid replication (Leenhouts et al., 1991). It is clear that the existence of this type of putative regulatory protein is not a general feature on RC plasmids, since they had been only found in two other plasmids, pCI411 (*Leuconostoc lactis*) and pWVO1 (*Lactococcus lactis*). Additional research is needed to attribute a precise function to this protein.

*3.3.Mobilization module*

Transmissible plasmids can be classified as being conjugative (self-transmissible) or mobilizable (transmissible only in the presence of additional conjugative functions). Mobilizable plasmids usually carry a mobilization gene (*mob*) encoding a specific relaxase and the origin of transfer (*oriT*) (Francia et al., 2004).

The *mob* gene of pPB1 encoded a putative 38.6 kDa protein that had high level identity to mobilization proteins encoded by pLB4 (*L. plantarum*) and pLAB1000 (*L. hilgardii*) (Table 1), which both belong in the pMV158-superfamily (Francia et al., 2004).

The organization of the mobilization region is similar in all superfamily members. It is composed one of a single gene, encoding the mobilization protein or relaxase, and an upstream *oriT* site that overlaps the relaxase promoter. Examination of the DNA sequence 5´to the *mob* gene in pPB1 for possible promoter sequences, reveals a potential –10 promoter site (nt 1662-1667), but lacks a recognizable –35 box. As suggested previously for pMV158 and pUB110, it may be that these promoters lack a –35 RNA polymerase-binding site (Priebe and Lacks, 1989).

The *oriT* sites in many of the pMV158 superfamily members display broad sequence conservation, their potential nick sites (Francia et al., 2004). All these potential *oriT* regions have an inverted repeat, the loop of which carries the nick site. Plasmid pPB1 contains an inverted repeat (located at nucleotide 1642 to 1680) having a $\Delta G$ -3.5 Kcal/mol. A putative consensus nick site (AGTGG/GTT) was located in this inverted repeat.

*3.4.Modular exchange as origin of pPB1*

Comparison of the pPB1-encoded proteins to with the databases revealed two different patterns of similarity. The RepA, RepB and RepC proteins showed had similarity to similar to proteins from the *Leuconostoc lactis* plasmid pCI411. By contrast, the Mob protein showed 96% identity to PreA from the *L. plantarum* plasmid pLB4 (Table 1). These data suggest that pPB1 was derived from two sources. To confirm the relationships between the various RC plasmid proteins we carried out phylogenetic analysis of each pPB1 protein. Figure 1 shows the phylograms based on phylogenetic trees obtained by the GeneBee program. As previously observed, the pCI411 proteins were the most similar to those of

pPB1, with exception of the Mob protein. Thus, phylogenetic analysis suggests that pPB1 and pCI411 are derived from a common plasmid ancestor.

By looking at DNA sequence similarities, we found that pPB1 showed 94.5% identity from nucleotide 2863 to 1460 to a region of plasmid pCI411 coding for the RepA and RepB proteins. Similarly, pPB1 showed 97.5% identity from nucleotide 1427 to 278 to the PreA-encoding region of pLB4. Figure 2 shows a schematic diagram of the sequence conservation found between pPB1 and pLB4 and pCI411. We have identified an unnamed ORF upstream of ORF1 that was not identified when pCI411 was originally described (Coffey et al., 1994). This protein appears to be a transcriptional repressor similar to the pPB1 RepA protein. Derived from a homologous recombination event between plasmid pCI411- and pLB4-like plasmids. In the three plasmids we identified two homologous segments of 46 bp and 343 bp, located at nucleotide positions 1427-1460 and 2863-278 in plasmid pPB1, respectively (Fig.3). The larger segment comprises the putative *sso* and *dso* regions of these plasmids. The small segment comprises the last codons of the *repC* gene in pPB1 and of ORF2 in plasmid pCI411, and an intergenic region upstream of *oriT* in plasmid pLB4. Since these segments are not identical, putative crossover points could be identified (Fig. 3).

In summary, pPB1 appears to be a chimera consisting of modules from two previously described plasmids, pCI411 and pLB4. This putative origin follows the Lawrence´s model of selfish evolution which explain the grouping of frequently horizontally transferred genes involved in a biological function in a module (Lawrence, 1999). The *rep* module from a pCI411-like plasmid and can be defined as the DNA fragment containing the *rep* genes, the target site of the Rep protein (*dso*) and the lagging-strand initiation site (*sso*). Besides the *rep* functions, pLB1 contains another module, the

*mob* module from a pLB4-like plasmid and which, encodes a mobilization protein and contains the origin of transfer (*oriT* ).

Thus, pPB1 seems to have originated through macroevolution by the transfer of large DNA fragments containing the *mob* and *rep* modules. It then, evolved through microevolution, as reflected by the few sequence differences found in the proteins and their corresponding target sites. Evolution by accumulation of modular units has been also proposed for some staphylococcal and lactococcal conjugative plasmids, and its well established for RC replication plasmids (Francia el at., 2004).

The data suggest an exchange of plasmid replicons by horizontal transfer between different species of lactic acid bacteria. Although *Lactobacillus plantarum* and *Leuconostoc lactis* belong to two quite different taxonomic groups, the fact that both organisms contained related plasmid sequences is not surprising. In that regard, the coexistence of *Lactobacillus* and *Leuconostoc* in wine fermentation processes may facilitate such a horizontal transfer.

*3.5.Presence of pPB1 derivatives in* L. plantarum *strains*

In previous studies, several *L. plantarum* plasmids that cross-hybridized with the *mob* region of pLAB1000 were identified (Josson et al., 1990). Since pLAB100 Mob protein is 95% identical to the mobilization protein of pPB1, and pLB4 was isolated from a *L. plantarum* strain, we checked the eleven *L. plantarum* strains included in the wine bacterial collection from the Instituto de Fermentaciones Industriales (BIFI) for the presence of one of the two pPB1 modules. Plasmid DNA extracted from these *L. plantarum* strains was Southern blotted and hybridized with two different probes. The *rep* probe was constructed

by PCR using oligonucleotides 103 and 78, it amplified almost the complete *repB* gene. The *mob* probe was constructed similarly by using oligonucleotides 104 and 76, and comprised the *mob* gene. No hydridization was found for any of the probes used (data not shown). Although some of these *L. plantarum* strains were isolated from the same winery or from a close geographical location, none of them harboured plasmids containing modules similar to the pPB1-derived *rep* or *mob* modules.

**Acknowledgments**

This work was supported by grant 07G/0035/2003 from the Comunidad de Madrid and RM03-002 from the Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria (INIA). We thank R. González and A.V. Carrascosa for their advice and critical reading of the manuscript. 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


Altschul, S. F., Madden, T. L., Schaffer, A., Zhang, J., Zhang, Z., Miller, W., Lipman, D. J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucl. Acids Res. 25, 3389-3402.

Alpert, C. A., Crutz-Le Coq, A.-M., Malleret, C., and Zagorec, M., 2003. Characterization of a theta-type plasmid from *Lactobacillus sakei*: a potential basis for low-copy-number vectors in lactobacilli. Appl. Envirom. Microbiol. 69, 5574-5584.

Bates, E. E. M., and Gilbert, H. J., 1989. Characterization of a cryptic plasmid from *Lactobacillus plantarum*. Gene 85:253-258.

Brantl, S., and Behnke, D., 1992. Copy number control of the streptococcal plasmid pIP501 occurs at three levels. Nucleic Acids Res. 20, 395-400.

Buckenhüskes, H. J., 1993. Selection criteria for lactic acid bacteria to be used as starter cultures for various food commodities. FEMS Microbiol. Rev. 12, 253-272.

Coffey, A., Harrington, A., Kearney, K., Daly, C., and Fitzgerald, G., 1994. Nucleotide sequence and structural organization of the small, broad-host-range plasmid pCI411 from *Leuconostoc lactis* 533. Microbiology 140, 2263-2269.

Daming, R., Yinyu, W., Zilai, W., Jun, C., Hekui, L., and Jingye, Z., 2003. Complete DNA sequence and analysis of two cryptic plasmids isolated from *Lactobacillus plantarum*. Plasmid 50, 70-73.

del Solar, G., Moscoso, M., and Espinosa, M., 1993. Rolling circle-replicating plasmids from Gram-positive and Gram-negative bacteria: a wall falls. Mol. Microbiol. 8, 789-796.

Francia, M. V., Varsaki, A., Garcillán-Barcia, M. P., Latorre, A., Drainas, C., and de la Cruz, F., 2004. A classification scheme for mobilization regions of bacterial plasmids. FEMS Microbiol. Rev. 28:79-100.

Josson, K., Soetaert, P., Michiels, F., Joos, H., and Mahillon, J., 1990. *Lactobacillus hilgardii* plasmid pLAB1000 consists of two functional modules commonly found in other gram-positive organims. J. Bacteriol. 172:3089-3099.

Khan, S. A., 1997. Rolling-circle replication of bacterial plasmids. Microbiol. Mol. Biol. Rev. 61, 442-455.

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

Klein, J. R., Ulrich, C., and Plapp, R., 1993. Characterization and sequence analysis of a small cryptic plasmid from *Lactobacillus curvatus* LTH683 and its use for construction of new *Lactobacillus* cloning vectors. Plasmid 30, 14-29.

Lawrence, J., 1999. Selfish operons: the evolutionary impact of gene clustering in prokaryotes and eukaryotes. Curr. Opin. Genet. Dev. 9, 642-648.

Leenhouts, K. J., Tolner, B., Bron, S., Kok, J., Venema, G., and Seegers, J. F. M. S., 1991. Nucleotide sequence and characterization of the broad-host-range lactococcal plasmid pWVO1. Plasmid 26, 55-66.

Leer, R. J., van Luijk, N., Posno, M., and Powels, P. H., 1992. Structural and functional analysis of two cryptic plasmids from *Lactobacillus pentosus* MD353 and *Lactobacillus plantarum* ATCC 8014. Mol. Gen. Genet. 234, 265-274.

Moreno-Arribas, M. V., Polo, M. C., Jorganes, F., and Muñoz, R., 2003. Screening of biogenic amine production by lactic acid bacteria isolated from grape must and wine. Int. J.Food Microbiol. 84, 117-123.

Muñoz, R., López, R, and López, E., 1998. Characterization of IS*1515*, a functional insertion sequence in *Streptococcus pneumoniae*. J. Bacteriol. 180, 1381-1388.

Pridmore, D., Stefanove, T., and Mollet, B., 1994. Cryptic plasmids from *Lactobacillus helveticus* and their evolutionary relationship. FEMS Microbiol. Lett. 124, 301-305.

Priebe, S. D., and Lacks, S. A., 1989. Region of the streptococcal plasmid pMV158 required for conjugative mobilization. J. Bacteriol. 171, 4778-4784.

Sambrook, J., Fritsch, E. F., and Maniatis, T., 1989. Molecular Cloning: a Laboratory Manual, 2nd Edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Sano, K., Otani, M., Okada, Y., Kawamura, R., Umesaki, M., Ohi, Y., Umezawa, C., and Kawamura, R., 1997. Identification of the replication region of the *Lactobacillus acidophilus* plasmid pLA106. FEMS Microbiol. Lett. 148, 223-226.

Takamatsu, D., Osaki, M., and Sekizaki, T., 2000. Sequence analysis of a small cryptic plasmid isolated from *Streptococcus suis* serotype 2. Curr. Microbiol. 40, 61-66.

**Figure legends**

Fig. 1. Phylogenetic analysis of the deduced pPB1 proteins RepA, RepB, RepC and Mob. The unrooted trees were based on the matrix of pair distances between sequences by using the program GeneBee. The amino acid sequences were obtained from the following accession number entries: RepA, **Q04136** (pWVO1), **Q48820** (pPSC22), **Q9ZN82** (pSB02), **Q48550** (pLA106), **P13920** (pMV158), **Q9S5E1** (pSSU1), **Q933S9** (pSMQ172), **Q939P3** (pSH72) and **P20044** (pLB4). RepB, **Q52222** (pCI411), **O32745** (pLA106), **P03858** (pE194), **P20045** (pLB4), **Q939P2** (pSH72), **Q9WW80** (pADB201), **Q9F870** (pBG7AU), **Q93K27** (pRSQ700), and **Q48612** (pCL2.1). RepC, **Q52223** (pCI411) and **Q48695** (pWVO1). Mob, **P35856** (pLAB1000), **P20046** (pLB4), **Q48534** (pLC88), **O32791** (pK214), **Q54160** (pVA380-1), **P13925** (pMV158), **Q9S5D7** (pSSU1), **Q93PW6** (pSMQ172), **Q93SL7** (pER13), and **Q7WVB3** (pSCFS1).

Fig. 2. Schematic overview of the sequence conservation between pPB1 and pLB4 and pCI411. Genes are represented by arrows with names shown below and with arrowheads indicating the direction of transcription. Genes encoding similar proteins are linked by similar shading; the *sso* and the *dso* regions are indicated by a box. Black boxes represent regions containing the putative crossover points. The 0 position on the scale bar represents nucleotide position 1436 in pPB1, 298 in pLB4 and 2395 in pCI411.

Fig. 3. Sequence alignment showing putative recombination cross-over points between pLB4 and pCI411. Nucleotide identities are shown by asterisks. The sequences at which pPB1 changes to being identical from pCI411 to pLB4 are shown in black boxes. White

boxes indicate the pPB1 sequence regions almost identical to pLB4 or to pCI411, that are separated from the putative cross over points. The location of the sequences are indicated by numbers which correspond to the published nucleotide numbers for the respective plasmids (A) Extended segment in the putative *sso-dso* region. (B) Small segment, the boundaries are not as sharply defined and internal mismatched based vary from plasmid to plasmid.

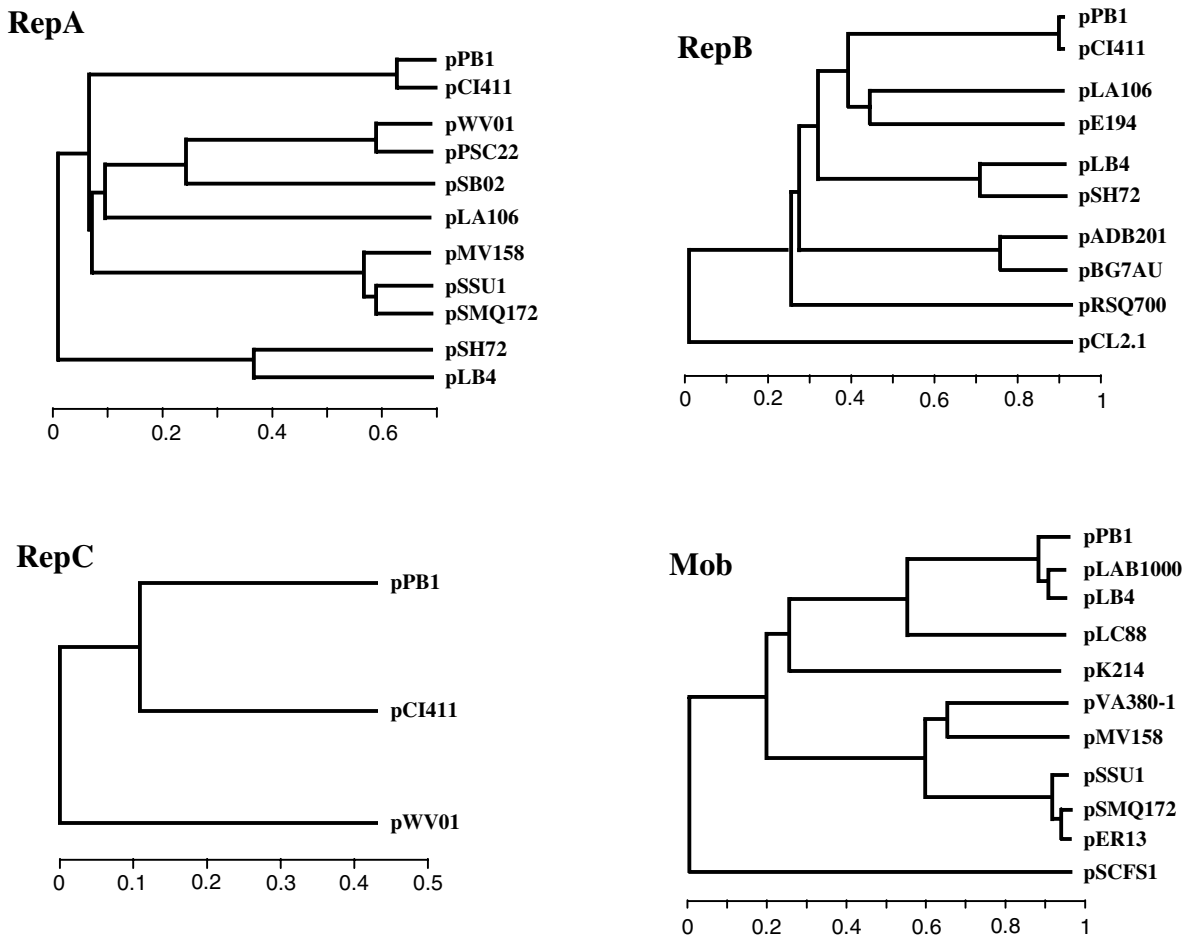
## Tables

Table 1
Characteristics of the ORFs of pBP1

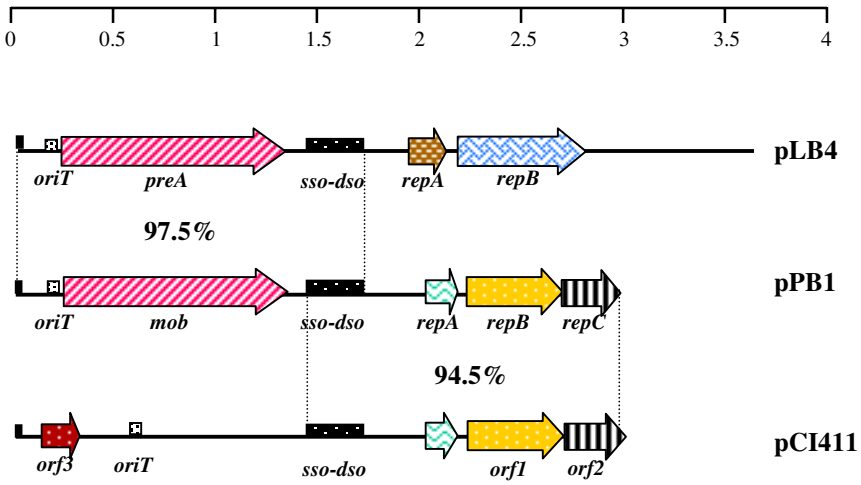
| DNA sequence characteristics | | | | Translation product characteristics | | | |
|---|---|---|---|---|---|---|---|
| ORF[a] | %GC | Position | D[b] | Size (aa) | Related protein (accession number) | Plasmid (organism) | Identity[c] (%) |
| *repA* | 29.0 | 447-611 | +71 | 54 | Replication protein RepA (Q04136) | pWV01(*Lactococcus lactis*) | 39 |
| | | | | | CopA protein (Q48820) | pPSC22 (*Lactobacillus plantarum*) | 37 |
| | | | | | RepA protein (O32744) | pLA106 (*Lactobacillus acidophilus*) | 31 |
| *repB* | 33.2 | 683-1309 | -4 | 208 | Replication initiation protein RepB (Q52222) | pCI411 (*Leuconostoc lactis*) | 97 |
| | | | | | ORF1 (Q48612) | pCL2.1 (*Lactococcus lactis*) | 61 |
| | | | | | RepB protein (O32745) | pLA106 (*Lactobacillus acidophilus*) | 47 |
| *repC* | 27.3 | 1306-1455 | +249 | 49 | Repressor protein RepA (Q52223) | pCI411 (*Leuconostoc lactis*) | 56 |
| | | | | | ORFD protein (Q48695) | pWV01 (*Lactococcus lactis*) | 49 |
| *mob* | 38.6 | 1704-2831 | +518 | 355 | Plasmid recombination protein (mobilization protein) PreA (P20046) | pLB4 (*Lactobacillus plantarum*) | 96 |
| | | | | | Mobilization protein PreA (P35856) | pLAB1000 (*Lactobacillus hilgardii*) | 95 |
| | | | | | Mobilization protein (Q48524) | pLC88 (*Lactobacillus casei*) | 56 |

[a] ORF are listed from the unique *Bgl*I site
[b] Distance between the start codon of the ORF and the stop codon of the upstream ORF. A negative value indicates an overlapping of two ORFs.
[c] Identities stretch over the entire length of the amino acid sequences of proteins.

**Figure**


RepA
pPB1
pCI411
pWV01
pPSC22
pSB02
pLA106
pMV158
pSSU1
pSMQ172
pSH72
pLB4
0
0.2
0.4
0.6
RepB
pPB1
pCI411
pLA106
pE194
pLB4
pSH72
pADB201
pBG7AU
pRSQ700
pCL2.1
0
0.2
0.4
0.6
0.8
1
RepC
pPB1
pCI411
pWV01
0
0.1
0.2
0.3
0.4
0.5
Mob
pPB1
pLAB1000
pLB4
pLC88
pK214
pVA380-1
pMV158
pSSU1
pSMQ172
pER13
pSCFS1
0
0.2
0.4
0.6
0.8
1

**Figure**


0
0.5
1
1.5
2
2.5
3
3.5
4
oriT
preA
sso-dso
repA
repB
pLB4
97.5%
oriT
mob
sso-dso
repA
repB
repC
pPB1
94.5%
orf3
oriT
sso-dso
orf1
orf2
pCI411

**Figure**

**A**

```
pLB4        TGTCCGTCAACGGTAAATCGACGTAGG-CGTTTTATAGCC-GCTGGGCTATTAGACGCCC
pPB1        TGTCCGTCAACGGTAAATCGACGTAGGGCGTTTTATAGCCCACTGGGCTATTAGACGCCC
pCI411      TGTCCGTCAACGGTAAATCGACGTAGAACG----------------GCTTTTAGCCGTTC
            *************************  **                *** **** **  *

pLB4        TAGGAGGCTTTAAGGAGTTGATAGACTAGC--GGATAAAACACTTTTGCACATGCAAAGA
pPB1        TAGGAGGCTTTAAGGAGTTGATAGACTAGC--GGATGAAACACTTTTGCACATGCAAAGA
PCI411      TGGGAGGCTTTAAGGAGTTGACGGACTCGCTAGGCCAAGACACTTTTGCGCATGCAAAGA
            * ******************  **** **  **   * ********** **********

pLB4        AAAGCACCCCTGCTTTTTTTGCCTGCCCCACGGCGAGTGCGGGGTGAGTTT-AGCGGGTG
pPB1        AAAGCACACCTGCTTTTTTTGCCTGCCCCACGGCGAGTGCGGGGTGAGTTTGAGCGGGTG
pCI411      AAAGCAACCCTGCTTTTTTTGCCTGCCTCACGGCGAGTGCGGGGTGAGTTTGAGCGGGGG
            ******  ****************** ********************** ****** *

pLB4        CTCCCG-TCATTTATGGGGTCAAGCTGACACAGCTTG---CGGGTTTGGGCAG--AGCCC
pPB1        CTCCCGCTCATTTATGGGGTCAAGCTGACACAGCTTG---CGGGTTTGGGCAG--AGCCC
pCI411      CTCCCTATC-TATAGGCGGTCAAGCTGGCGCACGTTGGACTGGGTTTGGGGCGTCAACCA
            *****  ** * ** * ********** * **  ***    *********     * **

pLB4        ATATTTTGGTTTGGTTTGAGTGGGATAAAAAAATTGGGCGAAAAACATGGGGG-TACTAC
pPB1        ATATTTTGGTTTGGTTTGAGTGGGATAAAAAAATTGGGCGAAAAACATGGGGGGTACTAC
pCI411      TGATTTTGATTTGGTTTGAGTGGGATAAAA--ATTGGGAGAAAAACATGGGGG-TACTAC
              ****** *********************  ****** ************** ******

pLB4        GACACCCCCCCATGTGTCCATTGTCCATTAAACAGAACACTTT   2061
pPB1        GACACCCCCCTATGTGGTCATTTGGTCATTTGGTCAAAATGGC    299
pCI411      AACACCCCCCTATGTGGTCATTTGGTCATTTGGTCAAAATGGC   1251
             ********* *****  ****      *      **
```

**B**

```
pLB4        TTTACCCCAATTGTCATGCGA--CTTTAAATAGAATTATT----GATTAATAAAAGCC   324
pPB1        GGAAGCTGATTTGCTGAG---------AAATAGAATTATT----GATTAATAAAAGCC  1463
pCI411      GGAAGTTGATTTACTGAGTTATACTTTACATGGAAGTCCTACCGGAATAATAAAATGA  2431
               *    * **     *          * ** *** *  *     * ********
```