**Characterization of IS*Lpl4*, a functional insertion sequence in
*Lactobacillus plantarum***

Keywords: Lactic acid bacteria; Transposase; Frameshift; Mobile element

Blanca de las Rivas, Ángela Marcobal, Alicia Gómez, Rosario Muñoz*

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

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

Abbreviations: IS, insertion sequence; IR, inverted repeat; DR, directed repeat; ORF, open
reading frame; PCR, polymerase chain reaction; RBS, ribosome binding site; CECT,
Spanish Type Culture Collection.

Abstract

A *Lactobacillus plantarum* strain, CECT 4645, was found to have insertions of a sequence (985 bp in length) at least in eight loci in its genome. The prototype copy (Lp1) of this insertion sequence (named IS*Lpl4*) has one open reading frame encoding a putative protein that is 292 amino acid in length with significant levels of similarity with IS982 family transposases. Perfect 16-bp inverted repeats were found at its termini. Upon transposition, generates 8-bp direct repeats of the target sequence, but no consensus sequences could be identified at either insertion site. The IS*Lpl4* pattern changed over many generations on the CECT 4645 strain. This finding strongly supports our hypothesis that IS*Lpl4* is a functional element in *L. plantarum*. Some of these elements may be cryptic, since point mutation or 1-nucleotide deletions were found in their transposase encoding genes. IS*Lpl4* copies have been detected in *Leuconostoc mesenteroides*, *Oenococcus oeni*, and *Lactobacillus sakei*. An IS*Lpl4* copy of *O. oeni* contained a +1 nucleotide insertion on its transposase encoding gene and, by using an experimental system, we were able to demonstrate that this specific sequence originates a +1 programmed translational frameshifting. Although the frameshitfing process reported here operates at a low rate, this description might represents the first case of a functional +1 frameshifting among IS.

1. Introduction

Insertion sequences (ISs) are small mobile genetic elements that are found in the genomes of numerous bacteria. They only contains genes involved in their transposition. ISs usually have inverted repeats (IRs) at their termini and duplicate a sequence consisting of several base pairs at the target site upon transposition (for reviews, see Mahillon and Chandler; Nagy and Chandler, 2004). Insertion of an IS can cause gene disruption or activation due to the creation or insertion of upstream promoters, and this contributes significantly to the plasticity of the host genome. Furthermore, a critical role of ISs in bacterial virulence is currently being recognized (Brynstad et al., 1997; Collins and Gutman, 1992; Hacker et al. 1997; Stroeder et al., 1995). IS elements are frequently used as markers for epidemiological purposes (Kivi et al., 2002; Stanley et al., 1993).

Lactobacillus plantarum is a flexible and versatile species that is encountered in a variety of environmental habitats, including some dairy, meat, and many vegetable or plant fermentations. Moreover, *L. plantarum* is frequently encountered as a natural inhabitant of the human gastrointestinal tract. The analysis of the complete genome sequence of *L. plantarum* WCFS1 strain (Kleerebezem et al., 2003) has revealed two classes of transposase-encoding regions that are likely to represent mobile genetic elements; and recently, Nicoloff and Bringel (Nicoloff and Bringel, 2003) described IS*Lpl1*, the first functional IS described in *L. plantarum*.

In the course of a investigation to identify genes possessing esterase activity, we isolated a segment of *L. plantarum* CECT 4645 DNA that exhibited all of the hallmarks of a bacterial IS. Since this IS was not present in the genome of the *L. plantarum* WCFS1 strain (Kleerebezem et al., 2003), here we report the sequence and characterization of IS*Lpl4*, a functional IS found in *L. plantarum* CECT 4645.

2. Materials and Methods

2.1. Bacterial strains, plasmids and growth conditions

Pure cultures of 33 lactic acid bacterial (LAB) strains were provided by the Spanish Type Culture Collection (CECT). A total of 51 LAB strains isolated from must grape or wine have been described in previous publications (Moreno-Arribas et al., 2003; Vaquero et al., 2004). *Escherichia coli* DH5 α (Sambrook et al., 1989) was also used. Plasmid pUC19 (New England Biolabs) and pGEM-T (Promega) were used to clone PCR fragments.

LAB strains were routinely growth in MRS medium (Difco) at 30 °C without shaking. *E. coli* cells were incubated in Luria –Bertani medium (Sambrook et al., 1989) at 37 °C with shaking. When required, ampicillin was added to the medium at 100 μ g per ml. Chromosomal DNA, plasmid purification, and transformation of *E. coli* were carried out as described (Sambrook et al., 1989).

The genetic stability of the IS element pattern was evaluated by subculturing a single-colony isolate of *L. plantarum* CECT 4645 strain on MRS broth during approximately 150 generations. Ten colony isolates were selected and chromosomal DNA was prepared as described previously, and hybridized with a IS*Lpl4*-specific DNA probe.

2.2. DNA purification and hybridization

Restriction endonucleases, T4 ligase, and the Klenow fragment of the DNA polymerase were obtained commercially and used according to the recommendations of the

suppliers. Gel electrophoresis of plasmids, restriction fragments, and PCR products was carried out in agarose gels as described previously (Sambrook et al., 1989). DNA was recovered from gel slices with the QIAquick gel extraction kit (Quiagen). An internal fragment of *ISLpl4* from *L. plantarum* CECT 4645 was PCR amplified and used as a probe. DNA was digoxigenin-labeled and chemiluminescently detected by using the DIG High Prime DNA labeling and detection Starter Kit (Roche) according to the manufacturer's instructions. DNA sequencing was carried out by using an Abi Prism 377TM DNA sequencer (Applied Biosystems). Sequence similarity searches were carried out using BLAST. Computer promoter predictions were carried out at http://www.fruitfly.org/seq_tools/promoter.html. Multiple alignment was done using CLUSTAL W at the EBI site (<http://www.ebi.ac.uk>).

2.3. Standard and inverse PCR amplification

PCR amplifications (25 µl) were performed with 2 U of AmpliTaq Gold DNA polymerase, 10 ng of template DNA, 200 µM of each dNTP, and 1 µM of each synthetic oligonucleotide primer in the buffer recommended by the manufacturer. The following primers, based on the nucleotide sequence of the prototype copy of the *ISLpl4* element of *L. plantarum* CECT 4645, the Lp1 copy (accession no. **AJ968657**) (see Sections 3.1 and 3.2), were used: 124 (876), 5'-TGAAACTAACTTAACGCGGAGTG; 141 (377/c), 5'-GCGCGACGGTTGAACCGTGAC; 147 (392), 5'-CGTGGTCAATGCGATCCGTACTG; and 168 (1 and 985/c), 5'-ACGCGTGGTGCTAGTT. The numbers in parentheses indicate the positions of the first nucleotide of the number in the sequence of Lp1, and c means that the corresponding sequence is on the complementary strand. The following primers were

used based on the *L. plantarum* WCFS1 strain complete sequence (accession [NC_004567](#)):
181 (315949/c), 5'-CAATAGCCGGTACATCTAACG; 182 (315699), 5'-
GGTTATCTTAATACTAGAGC; 183 (736741/c), 5'-TCAGAGTAAAGGTAGCCACG;
184 (736449), 5'-CTTAGCTGGTAATTATGTTAGC; 185 (2260012), 5'-
CGTTTTCTTGTAACATTACACC; and 186 (2260291/c), 5'-
TTTCGTCAAGTAACATTACACC.

To localize some of the *ISLpl4* integration sites in the hybridization positive strains, inverse PCR amplifications were performed. Chromosomal DNA was digested with *ClaI*, which cuts within the insertion element, and religated. The ligation mixture was used as template for inverse PCR. The PCR fragments were purified and sequenced. When the inverse-PCR amplified DNA fragment was a mix of two or more different fragments, it was cloned into pGEM-T vector (Promega).

2.4. Construction of β -galactosidase reporter plasmids

The β -galactosidase reporter plasmids pISO, pISO+1, and pISO-1 were constructed by inserting into the polylinker of pUC19 a 66-bp DNA fragment, from nucleotide 169 to 234 of the *ISLpl4* element of *O. oeni* RM83 (referred to as the Oo copy) identified in this study (accession no. [AJ781303](#); see Section 3.3). This DNA fragment includes the putative slippery sequence and the putative hairpin loop structure containing the stop codon (see Section 3.5 and Fig. 4). DNA fragments were inserted by using oligonucleotides ISO-up, 5'-TTTATAAAGCTTTGCCAGTCGCTTTAC; ISO+1-up, 5'-
TTTATAAAGCTTTGCCAGTCGCTTTAC; ISO-1-up, 5'-
TTTATAAAGCTTTGCCAGTCGCTTTAC; and ISO-down, 5'-

1 GTCATCGAATTCACGTGCGCCAC. The sequence underlined indicate introduced
2 restriction sites. β -Galactosidase activity was detected in cell extracts by using *o*-nitro-
3 phenyl- β -D-galactoside (Sigma) as substrate (Sambrook et al., 1989).

4 5 **3. Results**

6 7 *3.1. Identification of ISLpl4*

8
9 In the course of a research to identify *L. plantarum* CECT 4645 proteins showing
10 esterase activity, we constructed a *L. plantarum* genomic library in *E. coli*. When the insert
11 of one recombinant clone from the library possessing esterase activity was sequenced, a
12 DNA sequence that exhibited all the hallmarks of a bacterial IS was found close to the
13 putative esterase gene (unpublished results). This IS was located in an intergenic region in
14 *L. plantarum* CECT 4645 and was not present in the genome of the completely sequenced
15 *L. plantarum* WCFS1 strain (Kleerebezem et al., 2003). Except for the IS element, the
16 genetic organization of the DNA region was identical in both *L. plantarum* strains.
17 Therefore, we were able to identify the nucleotide position in WCFS1 where the IS was
18 inserted in CECT 4645. The 985-bp sequence (named ISLpl4) was inserted after position
19 equivalent to 736463 in the WCFS1 genome (Lp1 copy) (Fig. 1 middle).

20 21 *3.2. Structural analysis of ISLpl4*

22
23 As compared to the *L. plantarum* WCFS1 genome, the Lp1 copy of ISLpl4 was
24 flanked in strain CECT 4645 by a 8-bp (GCTAACAT) duplication (Table 1). The IS

1 element contains 16-bp perfect inverted repeat sequences at the ends (5'-
2 ACGCGTGGTGCTAGTT-3'). It consists of one unique open reading frame (ORF) with a
3 potential ATG start codon at position 97 to 99 and extending to a TAG terminator codon at
4 positions 973 to 975. The G+C content of the transposase gene (42.64%) is similar to the
5 average G+C content of the *L. plantarum* chromosome (44.5%) (Kleebeezem et al., 2003).
6 Computer predictions showed the nucleotide sequence (nt 35-63) that could function as a
7 promoter and also predicted the putative transcription start site at 6 nucleotides after the –
8 10 sequence. A typical ribosome binding site precede the ATG initiation codon. The
9 transposase gene putatively encodes a 292-amino acid protein with a high proportion of
10 basic amino acids and therefore a predicted pI value (10.05) characteristic of bacterial
11 transposases and consistent with the idea that it correspond to a DNA-binding protein
12 (Galas and Chandler, 1989).

13 Computer searches of the databases revealed that the transposase encoded by the
14 Lp1 copy of IS*Lpl4* (Tra_{Lp1}) was similar to some transposases from the IS982 family
15 (Mahillon and Chandler, 1998; Spellerberg et al. 2000). Figure 2A shows a multiple
16 alignment of Tra_{Lp1} (Pla in Fig. 2A) with related transposases. Tra_{Lp1} was 47-49% identical
17 to *Streptococcus* transposases belonging to the IS982 family. Tra_{Lp1} was 31% identical to
18 the transposase encoded by IS982 from *Lactococcus lactis* (Yu et al., 1995). In addition, the
19 lengths of these transposases were very similar, ranging from 287 to 296 amino acid
20 residues. All the features taken together strongly suggested that the 985-bp DNA fragment
21 correspond to a new IS element. In *Lactobacillus sakei* a 123 amino acid transposase
22 fragment (accession [AAU08010](#)) was found to be identical to the 123 amino acid residues
23 N-terminal fragment of Tra_{Lp1}.

3.3. Frequency and distribution of IS*Lpl4* among LAB

The presence of IS*Lpl4* related elements in other LAB was studied. We performed PCR analysis by using oligonucleotides 147 and 126 that amplifies an internal 524-bp DNA fragment, and also Southern hybridization experiments by using the same PCR fragments as probe. We studied the presence of IS*Lpl4* among 84 LAB strains. Only *L. plantarum* CECT 4645, *Leuconostoc mesenteroides* CECT 912 and *Oenococcus oeni* RM83 gave an 524-bp DNA amplicon on the PCR analysis. To determine the number of IS*Lpl4* copies, *Cla*I-digested DNAs from these strains were hybridized with the IS*Lpl4* probe. Since one cleavage site for *Cla*I is present in the IS element but not in the probe, the number of hybridization bands should reflect the copy number of IS*Lpl4*. As deduced from Figure 3A, the copy number in the IS*Lpl4* positive strains ranged from one in *L. mesenteroides* CECT 912 to, at least, six in *L. plantarum* CECT 4645.

Since the complete genome sequence for *L. plantarum* WCFS1 strain is available, and also the unfinished sequences for *L. mesenteroides* ATCC 8293 (accession **NZ AABH000000000**) and *O. oeni* PSU-1 (accession **NZ ABBJ000000000**) strains, we tried to localize the integration sites for some of the IS*Lpl4* copies in these genomes. Southern blot hybridization indicated that some of the IS copies were included in aprox. 3-kb *Cla*I fragments. Chromosomal DNAs from the IS*Lpl4* positive strains were then digested with this enzyme, religated and the ligation mixtures were used as templates for inverse PCR. To amplify these fragments we used primer 124, close to the 3' end of the IS*Lpl4*, and primer 141, located near the unique internal *Cla*I restriction site. The results are in accordance with those obtained in the Southern blot (data not shown). Since *L. mesenteroides* CECT 912 possesses only one IS*Lpl4* copy, we obtained only one apparent

2-kb band. Similarly, inverse PCR in *O. oeni* RM83 gave only one aprox. 2-kb DNA fragment corresponding to the smaller *Cla*I band observed in the Southern blot. Four apparent bands were obtained when using *L. plantarum* CECT 4645 DNA. All the PCR fragments were purified and sequenced. Since one of the *L. plantarum* bands produce a double sequence, we cloned it in the pGEM-T vector.

3.4. Localization and sequence of the IS*Lpl4* copies

We identified the localization of three *L. plantarum* IS*Lpl4* copies by comparing with the *L. plantarum* WCFS1 genome. The copies were inserted at positions equivalents to 315711, 736463 and 2260260 (Lp2, Lp1, and Lp3 IS*Lpl4* copies, respectively) (Fig. 1). In order to analyze how the ISs inserted into the genome, primers 181 to 186 were designed to amplify the chromosomal fragments containing these IS copies. The amplified fragments were sequenced. All the IS were properly inserted in the genome and the specific target site was determined on each copy (Table 1, Fig. 1).

Two additional IS*Lpl4* copies were localized in *L. plantarum* CECT 4645. Both copies were linked to the previously described IS*Lpl1* (Nicoloff and Bringel, 2003). The inverted repeat of one IS*Lpl4* copy (Lp4) was located contiguously at the end to the inverted repeat of a IS*Lpl1* copy. The other IS*Lpl4* copy is deleted at position 922 where again the inverted repeat of a IS*Lpl1* copy appeared (Table 1). Since *L. plantarum* WCFS1 did not possess any IS*Lpl1* copy, we could not find the right localization of these two IS*Lpl4* copies in the CECT 4645 genome.

When we compared the region adjacent to the IS*Lpl4* copies present in *L. mesenteroides* and *O. oeni*, with the available partial genome sequences, we could not find

any significant similarity. Therefore, we could not establish the genome localization of the IS*Lpl4* copies on these strains.

As mentioned above, when we searched in the databases for sequences similar to IS*Lpl4*, a 123 amino acid transposase fragment from *Lactobacillus sakei* identical to the putative transposase encoded by the Lp1 copy of IS*Lpl4* (accession [AAU08010](#)) appeared. This partial copy was inserted linked to a glucosyltransferase coding gene (Kralj et al., 2004) (Table 1).

The three *L. plantarum* CECT 4645 IS*Lpl4* copies that were localized by comparing the WCFS1 genome (Lp1, Lp2 and Lp3) were amplified using oligonucleotides based on the flanked regions and were fully sequenced. In order to sequence the additional IS*Lpl4* copies, chromosomal DNAs from positive IS strains were used in PCR experiments by using oligonucleotide 168, based on the IR sequence. All the IS*Lpl4* copies having both IR were amplified. The 985-bp PCR fragments obtained were cloned in pGEM-T vector. The insert of different recombinant plasmids were sequenced in order to obtain most of the IS*Lpl4* different copies (Fig. 2B).

Seven different IS*Lpl4* copies (Lp1 to Lp7) were obtained and sequenced from *L. plantarum* CECT 4645. Copies Lp1 to Lp3 were previously identified by inverse PCR and localized in the *L. plantarum* CECT 4645 chromosome (Table 1). Lp4 was previously identified by inverse PCR and linked to the IS*Lpl1* (Table 1). Lp5 to Lp7 were new copies of the IS*Lpl4* element not detected by the inverse PCR approach due to the size of the *Clal* DNA fragment where they were included (Fig. 3). An additional 8th copy was detected by inverse PCR and linked to another IS*Lpl1* copy, however the deletion of 75 nucleotides on its 3' end did not allow us to amplify and, therefore, sequence it (Table 1).

The IS*Lpl4* copies exhibited several nucleotide differences between them and with respect to the prototype Lp1 copy of *L. plantarum* CECT 4645 (Fig. 2B) and some of them also showed truncated reading frames, although at different positions. Comparison of the sequences revealed the occurrence of point mutation or 1-nt deletion or insertion. If we considered the Lp1 copy, predicted to encoded a 292-amino acid long transposase, as putative prototype sequence for the IS*Lpl4*, some of the additional copies contain different mutation that originated amino acid changes in their respective transposases, as the Gly-270-to-Arg change in the transposase encoded by the Lm copy. On the other hand, Lp2, Lp3, Lp5, and Oo copies coded for putative truncated transposases of 205, 235, 205, and 41-amino acids respectively (Table 1, Fig. 2B).

3.5. Putative +1 translational frameshift event required for transposition of the Oo copy

The observation given above suggested that the Lp2, Lp3, Lp5, and Oo copies contained different mutations rendering truncated, and therefore, inactive transposases. However, a comparison of the Oo sequence lead us to find the transposase interruption as consequence of overlapping proline codons, that was previously found to be cause +1 frameshifting (De Smit et al., 1994; Zimmer et al., 2003). Figure 4 shows that the potential frameshift in the Oo transposase gene occurs at a location close to the 5' end of the gene, at the mRNA sequence UUG'CCC'ACG (corresponding to the nucleotides 192 to 200 of the Lp1 copy). The ribosome apparently could slips from the CCC proline codon one nucleotide position into the 3'-direction (overlined), and continues from the overlapping proline triplet (CCA) in the +1 frame until it reaches the stop codon at position 975. Interestingly, the transposase RNA is capable of forming a putative 3' pseudoknot structure,

1 5'-GUGGCGC (224 to 230) and 5'-GCGCUAC (260-266), downstream of the slippage
2 site (Fig. 4A). These secondary structures were previously known to stimulate
3 frameshifting (Mejlhede et al., 2004; Zimmer et al., 2003). The putative +1 frameshifted
4 transposase is 292 amino acids long and only differs in the amino acid residue located at
5 position 33 (Cys to Leu) from the prototype transposase coded by the Lp1 copy of the
6 ISLp14 from *L. plantarum* CECT 4645.

7 To determine the role of this Oo sequence in frameshifting, three pUC19 derivatives
8 possessing LacZ fusions were constructed. They comprised DNA segment in its polylinker
9 from nucleotide 169 to 234, 26 bp before the putative slippery sequence to the end of the
10 stem loop structure containing the stop codon. These fragments were cloned by PCR using
11 different oligonucleotide primers only differing in one or two extra nucleotides, giving the
12 insert cloned in the three possible reading frames with respect to *lacZ*. The pISO plasmid
13 had the insert on frame with *lacZ*; in the pISO+1 plasmid a +1 ribosomal frameshift will be
14 needed to express the *lacZ* gene and thus, correct reading of *lacZ* takes place only after the
15 frameshift occurs, and finally, a -1 frameshift will be necessary in pISO-1 to have β -
16 galactosidase activity. The data showed that the pISO+1 plasmid presented a 98.5%
17 reduction in β -galactosidase activity with respect to pISO which had the fragment on frame
18 with *lacZ* (not shown). However, pISO+1 showed twofold β -galactosidase activity as
19 compared to pISO-1 (5.6 versus 2.6 Miller units), indicating that the +1 ribosomal
20 frameshift might play a functional role in the transposition process.

21 22 3.6. Genetic stability over time of ISLp4 in *L. plantarum* CECT 4645

To study the genetic stability of *ISLpl4* we choose *L. plantarum* CECT 4645 since it harbors the highest number of copies of the *ISLpl4* element. As described in Material and Methods, cells from a single colony were grown during 150 generations in liquid media. After several rounds of serial growth, ten clones were isolated on MRS agar plates. The *ISLpl4* pattern of these clones was compared to that of parental strain. Chromosomal DNA was digested with *Cla*I and southern blot hybridized with the *ISLpl4* specific probe (Fig. 3B). A different *ISLpl4* pattern was observed only in one clone, on which a band of aprox. 7 kb was lost. Since Lp1 was included in a 3 kb *Cla*I-DNA fragment, this 7 kb band might correspond to Lp6 and Lp7 copies which coded for non-truncated transposases. Since a different pattern was found after 150 generations of *L. plantarum* CECT 4645, it was concluded that *ISLpl4* was not stable over time. This observation suggested that at least a copy of the IS element is functional in the CECT 4645 strain and able to excise from the *L. plantarum* chromosome.

4. Discussion

The analysis of the nucleotide sequence of a multicopy DNA fragment identified in *L. plantarum* CECT 4645 showed that this fragment displays the characteristic features of an IS element (Mahillon and Chandler, 1998; Nagy and Chandler, 2004) that was named *ISLpl4*. These features include the following: (i) the IS element is bounded by two 16-bp terminal perfect inverted repeats, and (ii) there is a structure consisting of an ORF encoding a putative protein with a size of 292 amino acids. This protein is rich in basic amino acid residues (pI 10.05) and exhibits substantial similarity at the amino-acid level to putative transposases of the IS982 family: e.g., *Lactococcus lactis* subsp. *cremoris* SK11 (Yu et al.,

1995), *Lactococcus lactis* subsp. *lactis* biovar *diacetylactis* KR2 (Twomey et al., 1998), and in some strains belonging to the *Streptococcus* genera (Spellerberg et al., 2000). Moreover, similarity of the terminal IR sequences (Table 1) and the DDE motif lacking the conserved K/R residues (Fig. 2A), also suggest that the transposase coded by *ISLpl4* is probably a new member of the IS982 family (Mahillon & Chandler, 1998; Spellerberg et al., 2000).

PCR and hybridization test revealed that the genome of CECT 4645 contains at least eight copies of *ISLpl4*, but this finding not necessarily imply that all these copies were identical or even functional. Moreover, at least three of these copies (Lp2, Lp3, and Lp5) may be cryptic elements due to putative truncated transposases (Table 1, Fig. 2B) resulting from point mutation or 1-nt deletions within their transposase-coding genes. Evidence of *ISLpl4* activity in *L. plantarum* CECT 4645 includes the variation in the *ISLpl4* pattern seen after strain CECT 4645 was grown in MRS medium during 150 generations. Even if only one of the copies of *ISLpl4* detected in *L. plantarum* CECT 4645 were functional, such change in the IS bands would be expected.

The IS reported here was present in multiple copies in *L. plantarum* CECT 4645 strain but not in the other twelve *L. plantarum* strains studied. However, *ISLpl4* copies were found in *L. sakei* Kg15, *L. mesenteroides* CECT 912 and *O. oeni* RM83. *ISLpl4* may be transferred among different LAB genera, as suggested by the high degree of identity between them. Some of the *ISLpl4* copies in *L. plantarum* CECT 4645 were not localized in the WCFS1 complete genome sequence and they were linked to *ISLpl1*; we cannot exclude a plasmid localization of these copies since *ISLpl1*-related elements were associated to different plasmids (Kleerebezem et al., 2003; Nicoloff and Bringel, 2003). Moreover, plasmid localization of *ISLpl4* may facilitate gene exchange among LAB.

Transposition activity is generally maintained at a low level. Transposition frequency can be regulated indirectly through the control of transposase level or indirectly by modulation of the recombination reaction by alterations in transposase activity and the reactivity of the DNA substrates. The regulatory mechanisms used by transposable elements to control transpositions are numerous and act at virtually every level of gene expression (Mahillon and Chandler, 1998; Nagy and Chandler, 2004). Whereas the Lp1 copy in *L. plantarum* CECT 4645 encodes a complete transposase that seems to confer full functionality on the IS element, in the transposase encoded by the Oo copy a possible translational control of transposition by +1 frameshifting might be involved. It has been well established that several ISs encode transposases that are synthesized by a –1 frameshift, however, transposases translated by a +1 frameshift have not been reported (Baranov et al., 2002), although a +1 translational frameshift has been claimed to occur in the *Streptococcus pneumoniae* IS1381 transposase (Sánchez-Beato et al., 1997). In IS1381 an atypical motif, A'AAA'AAG, followed by a hairpin loop containing the stop codon was found. Nevertheless, the feature of overlapping proline codons, found in the transposase encoded by the Oo copy of the ISLpl4 element, was previously found to cause +1 frameshifting in an *E. coli* plasmid system (De Smit, 1994) and in the programmed +1 translational frameshifting of *Listeria monocytogenes* phage PSA (Zimmer et al., 2003). On these motifs, the proline-encoding CCC codon appears to be the point where the ribosome stalls and slips one nucleotide position forward into the 3' direction, facing the overlapping proline codon. Whereas CCC is a rare codon, representing only 14.7% of the possible proline codons in the *L. plantarum* genome (<http://www.kazusa.org.jp/codon/>), CCA features a codon usage of 41.4%. As proposed in the *L. monocytogenes* PSA phage, is it therefore reasonable to assume that the “hungry” codon negatively influences ribosomal

1 processivity, enhances stalling of the ribosome at the slippage site, and thereby favours
2 frameshifting.

3 The current model for ribosomal frameshifting implies that all factors decreasing the
4 speed of the elongating ribosome along the mRNA, such as RBS, hungry codons, and
5 secondary structures, which have to be unwound, result in increased frameshifting at
6 slippery sites. In the Oo copy of *ISLpl4* the sequence TTG'CCC'ACG is combined with an
7 mRNA secondary structure forming a pseudoknot (Fig. 4A). Further increase in the
8 efficiency of frameshifting is depending on the codon usage of individual host strains, when
9 aminoacyl-tRNA limitation causes the ribosome to halt on the mRNA at a hungry codon
10 and stimulates frameshifting.

11 We designed an experimental system for the convenient quantitative monitoring of
12 frameshifting using *lacZ* as a reporter gene. The basic plasmid carried the *lacZ* gene
13 coupled by a 66-bp fragment containing the *ISLpl4* frameshift site. The fusion gave about
14 1.5% of the β -galactosidase activity synthesized by a reference plasmid with an in-frame
15 fusion. This value is quite low as compared to other reported values, since 2-16% efficient
16 +1 frameshift have been reported (Fu and Parker, 1994; Zimmer et al., 2003), and also -1
17 frameshift efficiency values of 4.4% (Mejlhede et al., 2004), 10% (Condreay et al., 1989)
18 or 15% (Jacks et al., 1988). We can not exclude that sequences outside the bounds of the
19 cloned frameshift fragment play an important role in frameshifting. Moreover, as
20 mentioned above, the *E. coli* codon usage may be an additional reason of the low efficiency
21 of +1 frameshifting obtained in the experimental system designed. In *E. coli* K12 both
22 proline codons, CCC and CCA, are low represented on the possible proline codons, being
23 12.4 and 19.2% respectively. Since transposition activity is generally maintained at a low

level, this +1 frameshifting system of modulation may provide a sensitive mechanism for regulating rates of transposition.

Another general feature of IS elements is that, on insertion, most generate short directly repeated sequences (DR) on the target DNA flanking the IS. Attack of each DNA strand at the target site by one of the two IS ends in a staggered way during insertion provides an explanation for this observation. The length of the DR, between 2 and 14 bp, is characteristic for a given element and a given element will generally generate a duplication of fixed length (Nagy and Chandler, 2004). By comparing to the *L. plantarum* WCFS1 complete genome we were able to localize the insertion sites of three copies of *ISLpl4* in CECT 4645, and therefore, identified the short directly repeat sequences generated on the target DNA. In all these copies, a 8-bp different DR sequences were found. The *ISLpl4* copy linked to *ISLpl1* showed absence of DR (Table 1). Care should be taken in interpreting the absence of DRs in isolated cases. It has been described that in certain cases, ISs without DR can simply result from homologous inter- or intramolecular recombination between two IS elements, each with a different DR sequence. This would result in a hybrid element carrying one DR of each parent (Nagy and Chandler, 2004). This phenomenon of interdigitation of various intact or partial IS elements has been noted in the literature. Many of these observations are anecdotal and may reflect the scars of consecutive but isolated transposition events. On the other hand, several IS exhibit a true preference for insertion into other elements (Mahillon and Chandler, 1998).

The presence of a large number of IS elements might be expected to have strong influence on the stability of the genome. It is important to note that several of these *ISLpl4* copies carry mutations which render them inactive. This is an efficient way to keep transposition activity at a low level. Moreover, it has been described that transposases act

more efficiently on the element from which it is expressed than on other copies of the same element in a genome or plasmid (Mahillon and Chandler, 1998). In addition, the low efficiency of the +1 frameshift event obtained in the experimental system reported here might provide a sensitive mechanism for a tight regulation of the transposition rates in that specific *ISLpl4* copy.

Acknowledgments

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

The DNA sequences of the *ISLpl4* copies have been deposited in the GeneBank database under accession no. AJ968657 to AJ968664, and AJ781303. *ISLpl4* was named according to the IS database curators (<http://www-is.biotoul.fr>).

References

- Baranov, P.V., Gesteland, R.F., Atkins, J.F., 2002. Recoding: translational bifurcations in gene expression. *Gene* 286, 187-201.
- Brynstad, S., Synstaf, B., Granum, P.E., 1997. The *Clostridium perfringens* enterotoxin gene is on a transposable element in type A human food poisoning strains. *Microbiology* 143, 2109-2115.

1 Collins, C.M., Gutman, D.M., 1992. Insertional inactivation of an *Escherichia coli* urease
2 gene by IS3411. J. Bacteriol. 174, 883-888.

3 Condreay, J. P., Wright S.E., Molineux, L.J., 1989. Nucleotide sequence and
4 complementation studies of the gene 10 region of bacteriophage T3. J. Mol. Biol.
5 207, 555-561.

6 De Smit, M. H., van Duin, J., van Knippenberg, P.H., van Eijk, H.G., 1994. CCCUGA: a
7 new site of ribosomal frameshifting in *Escherichia coli*. Gene 143, 43-47.

8 Fu, C., Parker, J., 1994. A ribosomal frameshifting error during translation of the *arg1*
9 mRNA of *Escherichia coli*. Mol. Gen. Genet. 243, 434-441.

10 Galas, D.J., Chandler, M., 1989. Bacterial insertion sequences, p. 109-162. In D. E. Berg
11 and M. M. Howe (ed.), Mobile DNA. American Society for Microbiology,
12 Washington, D. C.

13 Hacker, J., Blum-Oehler, G., Mühldorfer, I., Tschäpe, H., 1997. Pathogenicity islands of
14 virulent bacteria: structure, function and impact on microbial evolution. Mol.
15 Microbiol. 23, 1089-1097.

16 Jacks, T., Madhani, H.D., Marsiarz, F.R., Varmus, H.E., 1988. Signals for ribosomal
17 frameshifting stimulates in the Rous sarcoma virus *gag-pol* region. Cell 55, 447-
18 458.

19 Kivi, M., Liu, X., Raychaudhuri, S., Altman, R.B., Small, P.M., 2002. Determining the
20 genomic locations of repetitive DNA sequences with a whole-genome microarray:
21 IS6110 in *Mycobacterium tuberculosis*. J. Clin. Microbiol. 40, 2192-2198.

22 Kleerebezem, M., Boekhorst, J., van Kranenburg, R, Molenaar, D, Kuipers, O.P., Leer, R.,
23 Tarchini, R., Peters, S.A., Sandbrink, H.M., Fiers, M.W.E.J., Stiekema, W,
24 Lankhorst, R.M.K., Bron, P.A., Hoffer, S.M., Groot, M.N., Kerkhoven, R., de

1 Vries, M., Ursing, B., de Vos, W.M., Siezen, R.J., 2003. Complete genome
 2 sequence of *Lactobacillus plantarum* WCFS1. Proc. Natl. Acad. Sci. USA. 100,
 3 1990-1995.

4 Kralj, S., van Geel-Schutten, G.H., Dondorff, M.M., Kirsanov, S., van der Maarel, M.J,
 5 Dijkhuizen, L., 2004. Glucan synthesis in the genus *Lactobacillus*: isolation and
 6 characterization of glucansucrase genes, enzymes and glucan products from six
 7 different strains. Microbiology 150, 3681-3690.

8 Mahillon, J., Chandler, M., 1998. Insertion sequences. Microbiol. Mol. Biol. Rev. 62, 725-
 9 774.

10 Mejlhede, N., Licznar, P., Prère, M.F., Wills, N.M., Gesteland, R.F., Atkins, J.F., Fayet, O.,
 11 2004. -1 Frameshifting at a CGA AAG hexanucleotide site is required for
 12 transposition of insertion sequence IS1222. J. Bacteriol. 186, 3274-3277.

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

16 Nagy, Z., Chandler, M., 2004. Regulation of transposition in bacteria. Res. Microbiol. 155,
 17 387-398.

18 Nicoloff, H., Bringel, F., 2003. IS*LplI* is a functional IS30-related insertion element in
 19 *Lactobacillus plantarum* that is also found in other lactic acid bacteria. Appl.
 20 Environ. Microbiol. 69, 6032-6040.

21 Sambrook, J., Fritsch, E.F, Maniatis, T (ed)., 1989. Molecular cloning: a laboratory manual.
 22 Cold Spring Harbor Laboratory, Cold Spring Harbor, N. Y.

- 1 Sánchez-Beato, A.R., García, E., López, R., García, J.L., 1997. Identification and
2 characterization of IS1381, a new insertion sequence in *Streptococcus pneumoniae*.
3 J. Bacteriol. 179, 2459-2463.
- 4 Spellerberg, B., Martin, S., Franken, C., Berner, R., Lütticken, R., 2000. Identification of a
5 novel insertion sequence element in *Streptococcus agalactiae*. Gene 241, 51-56.
- 6 Stanley, J., Baquar, N., Threlfall, E.J., 1993. Genotypes and phylogenetic relationships of
7 *Salmonella typhimurium* are defined by molecular fingerprinting of IS200 and 16S
8 *rrn* loci. J. Gen. Microbiol. 139, 1133-1140.
- 9 Stroecher, U. H., Jedani, K.E., Dredge, B.K., Morona, R., Brown, M.H., Karageorgos, L.E.,
10 Albert, M.J., Manning, P.A., 1995. Genetic rearrangements in the *rfb* regions of
11 *Vibrio cholerae* O1 and O139. Proc. Natl. Acad. Sci. USA. 92, 10374-10378.
- 12 Twomby, D.P., McKay, L., O'Sullivan, D.J., 1998. Molecular characterization of the
13 *Lactococcus lactis* LlaKR21 restriction-modification system and effect of an IS982
14 element positioned between the restriction and modification genes. J. Bacteriol. 180,
15 5844-5854.
- 16 Vaquero, I., Marcobal, A., Muñoz, R., 2004. Tannase activity by lactic acid bacteria
17 isolated from grape must and wine. Int. J. Food Microbiol. 96, 199-204.
- 18 Yu, W., Mierau, I., Mars, A., Johnson, E., Dunny, G., McKay, L.L., 1995. Novel insertion
19 sequence-like element IS982 in lactococci. Plasmid 33, 218-225.
- 20 Zimmer, M., Sattelberger, E., Inman, R.B., Calendar, R., Loessner, M.J., 2003. Genome
21 and proteome of *Listeria monocytogenes* phage PSA: an unusual case for
22 programmed +1 translational frameshifting in structural protein synthesis. Mol.
23 Microbiol. 50, 303-317.
- 24

Legends to Figures

Fig. 1. Overview of the putative genetic organization of the *L. plantarum* CECT 4645 genome where Lp1 (middle), Lp2 (top) and Lp3 (down) copies of the IS*Lpl4* element are inserted (see Section 3.4 and Table 1). Arrows represent complete ORFs. Black arrows represent the transposase encoding genes of the Lp1, Lp2, and Lp3 copies of the IS*Lpl4* element (*tra_{Lp1}*, *tra_{Lp2}*, and *tra_{Lp3}*). Some of the genes present in these regions are indicated: *tagG* (teichoic acid ABC transporter, permease), *tagH* (teichoic acid ABC transporter, ATP-binding protein), *amtB* (ammonium transport protein), *eriC* (chloride channel protein), *secG* (preprotein translocase, SecG subunit), *smpB* (SSRA RNA binding protein), *ptsI8CBA* (N-acetylglucosamine and glucose PTS, EIICBA), *panE1* (2-dehydropantoate 2-reductase), *homI* (homoserine dehydrogenase), *metY* (O-acetylhomoserine (thiol)-lyase), and *metA* (homoserine O-succinyl transferase). The numbers of the nucleotide positions corresponding to the *L. plantarum* WCSF1 complete genome sequence are indicated on the right and on the left. The insertion site position of the IS*Lpl4* copy is indicated above the black arrow.

Fig. 2. (A) Alignment of the deduced amino acid sequences of transposases of several IS*Lpl4*-related elements. The putative transposases aligned are from *Streptococcus criceti* (Scr) (accession no. **BAB64015**), *Streptococcus agalactiae* (Sag) (accession no. **AAF26677**), *Streptococcus mutans* (Smu) (accession no. **AAK48416**), *L. plantarum* CECT 4645 IS*Lpl4* prototype Lp1 copy, Tra_{Lp1} (Pla) and the IS982 transposase from *Lactococcus lactis* (Lla) (accession no. **L34754**). The multiple alignment was carried out with the Clustal W program. The acidic amino acid triad, DDE motif, present in all transposases and

involved in catalysis is indicated in bold. Identical amino acid residues in all sequences are indicated by asterisks, conserved or semiconserved substitutions are denoted by colons or dots, respectively. (B). Nucleotide sequence of individual copies of the IS*Lpl4* element. Lp1 to Lp7 represents the IS*Lpl4* copies present in *L. plantarum* CECT 4645. Lm copy corresponds to the sequence present in *L. mesenteroides* CECT 912, and Oo, the *O. oeni* RM83 copy that showed a putative +1 frameshift on its putative transposase. Vertical numbers indicated nucleotide position in the prototype Lp1 copy of the IS*Lpl4* element. Only sites that differ between sequences are shown. Dots and dashes indicated nucleotides identical to those of the prototype Lp1 copy and gaps, respectively.

Fig. 3. Presence of IS*Lpl4* in LAB. The Southern blottings were performed with *Cla*I-digested DNA from the indicated strains. (A) IS*Lpl4* distribution in *L. plantarum* CECT 4645 (lane 1), *O. oeni* RM83 (lane 2) and *L. mesenteroides* CECT 912 (lane 3). (B) Instability over time of IS*Lpl4* in a clone of CECT 4645 strain obtained after 150 generations of growth in liquid MRS medium (lane 1) as compared to the parental *L. plantarum* CECT 4646 strain (lane 2). The size of standards are indicated on the left.

Fig. 4. Programmed translational +1 frameshift in the IS*Lpl4* copy present in *O. oeni*. (A) The mRNA sequence immediately downstream of the slippery sequence has the potential to form a secondary structure and a 3' pseudoknot, which are thought to stimulate the frequency of frameshifting. (B) The +1 shift in decoding the transposase gene in *O. oeni* is shown, leading to the synthesis of an extended protein. The slippery sequence in mRNA is printed in bold letters. The stop codon is shown in italics. The potential stem-loop structure is overlined and the potential pseudoknot is indicated by brackets.

TABLE 1. Characteristics of IS*Lpl4* copies

Bacterial host	Genetic linkage	Adjacent sequence ^a	IR sequence (5' → 3')	Target duplication	nt sequence / Tra (aa)	Reference
<i>L. plantarum</i> CECT 4645	SSRA RNA binding protein	TTATATGCTAACAT GCTAACATAATTAC	ACGCGTGGTGCTAGTT (IRL) AACTAGCACCACGCGT (IRR)	GCTAACAT	Lp1 / 292	This study
<i>L. plantarum</i> CECT 4645	Teichoic acid ABC transporter	CTTTATGCTCTAGT GCTCTAGTATTAAG	ACGCGTGGTGCTAGTT (IRL) AACTAGCACCACGCGT (IRR)	GCTCTAGT	Lp2 / 205, Tra interrupted	This study
<i>L. plantarum</i> CECT 4645	Homoserine dehydrogenase	TACTATGTGTTAAC GTGTTAACTAAAAA	ACGCGTGGTGCTAGTT (IRL) AACTAGCACCACGCGT (IRR)	GTGTTAAC	Lp3 / 235, Tra interrupted	This study
<i>L. plantarum</i> CECT 4645	IS <i>Lpl1</i>	GTAGATTGTAAAAT	AACTAGCACCACGCGT (IRR)	None	Lp4 / 292	This study
<i>L. plantarum</i> CECT 4645	unknown	unknown	unknown	unknown	Lp5 / 205, Tra interrupted	This study
<i>L. plantarum</i> CECT 4645	unknown	unknown	unknown	unknown	Lp6 / 292	This study
<i>L. plantarum</i> CECT 4645	unknown	unknown	unknown	unknown	Lp7 / 292	This study
<i>L. plantarum</i> CECT 4645	IS <i>Lpl1</i>	unknown	unknown	unknown	75 nucleotides 3' deleted (Lp8?)	This study
<i>L. mesenteroides</i> CECT 912	Transcriptional regulator & IS	GATTCAATGCTATT	AACTAGCACCACGCGT (IRR)	unknown	Lm / 292	This study
<i>Oenococcus oeni</i> RM83	unknown	AAAATAAAGCTAT	AACTAGCACCACGCGT (IRR)	unknown	Oo / putative +1 frameshift	This study
<i>Lactobacillus</i> <i>sakei</i> Kg15	Glucosyl transferase	ATTAATAAAACACAT	ACGCGTGGTGCTAGTT (IRL)	unknown	Uncompleted sequence	Kralj et al., 2004

^a The proposed target duplications are indicated in bold.

IR, inverted repeat; nt, nucleotide; aa, amino acid; IS, insertion sequence; Tra, transposase

Figure 1

Figure 2

A

Scr	MSHLQYTAKSHHLQWNLKQLSKICHQLYRDYCPYSFKHRHNISLSKVS	48
Sag	MSHLQYTAKSHHLQWNLKQLSKICHQLYRDYCPESFKHRHNISLSKVS	48
Smu	MLLVIKRTGVTQMNHLHYTAKHHHLQYSMTQLKVICSYYYRHYAPKELTQRRNVHYCKVN	60
Pla	MLNHPKFKQIRHELQGNHYHYFYQLCQSLYQRYCPRCVTQRRNVAVHKID	49
Lla	MTYNSTLPKVFVYLLTTIETLYQTRVPLEVQNRKNVHLATSG	42
	** : *: * . :***: . .	
Scr	DQSLLVLLILQAELGIKSQRHFYRRLCHLFP-CGQLLERSRFNRRARQLVWLQVIRQAMN	107
Sag	DQSLLVLLILQAELGIKSQRHFYRICHLFP-CGCLLERSRFNRRARQLIWLQIIRQAMN	107
Smu	DDTILVLLVLQAELGIKSQRRFYQLCQVFL-VTTGLERTRFYRRARYLIPLLKRIHQGMT	119
Pla	DCRLLALLCLQTTLGIQSQRRFWRLMTAFMPQKITISRSRFNRRALQLLPVVNAIRTGLT	109
Lla	CLVIACYLWGVLFHSETLKAKHQLAQLSFP--NFLEYSRFVRRCNGLLPSIQVIRQALV	99
	: * : . : . * : . :*** ** * : : * . :	
Scr	TQISP-DTIAIIDSFPLPLCQPVNRCTRIFNDLADIGYNASKHLWFGYFKVHMLVTLTG	166
Sag	SHMSH-ETIVIIDSLPLLLCQPIRHHVRVIFEGDLADIGYNASKHLWFGYFKVHMLVTLTG	166
Smu	SQFHQ-DDIVIMDSFPIPVCCPARHYKVCIFRDQATISYNPSKKMWFYGFKAHMLVTLTG	178
Pla	REAAQPGQIAIIDSLPNPLCEPVRKFRARIFAGQANIGYNATKQMSFYGFKTHMLVTTDG	169
Lla	FKEVEGMSVSIIDSFPIPLCQPIRNFRRSKGLGDYANVGYNATKGQYFYGCKCHALVSESG	159
	. : :***: * * : : : . * : ** . : * * * * * : .	
Scr	YILNYVVTASVHDIRAVDDLENCRQPYILADLGYLSRELKDHLVQKGYHLWTPLRQNM	226
Sag	YILNYVVTASVHDIRAVDDLENCRQPYILADLGYLSRDLKDHLVQKGYHLWTPLRQNM	226
Smu	FIVNYVVTASVHDIRQVAEDLLENTFPPIVLADLGYLSQVLKQHLTQRGYHFWTPLRRNM	238
Pla	YILNYVVTASVHDTKVAVSLIDDCPCPIVLADVGYYVGKRLGAEFKQLGYTLWTPYRSNM	229
Lla	YVIDYTITPASMADSSMTTEEVLSQFGTPTVLGDMGYLGQSLHDRLELELEGIDLMTPVKKNM	219
	:***: . :***: * . . : : * : * :***: . : * . : * : * * *	
Scr	AGAKQHNHWKLMAMRRTIETRFSELCAFDVEHTLTRGLAGLQLRLEQIILTHNLRYFEI	286
Sag	EGSKQHNHWKLMTRRTIETRFSELCSLFDIEHTLTRGLTGLQLSIEQMILAYNLRYFEI	286
Smu	ANAKKHNNHWKLKAERTIETRFVLCSEFDIERPLARSLKGLEIWLEQAIFAYNLRRFN-	297
Pla	KGAKEHNKRALKALRRTIESRFSTLVSDFGIETNLTRSAFGFQLKIELAILVYNLGFFNF	289
Lla	K-QKNILFPNFSKRRKVIERVFSFLTNLGAERCKSL-SPQGFQLKLEMILLAYSLLLKSA	277
	*:: : ** ** * * : : . * . :***: * : : : . * .	
Scr	N-----	287
Sag	N-----	287
Smu	-----	297
Pla	VTN-----	292
Lla	KSLEPETLRYSIGYQVMAK	296

Figure 2 (cont.)

B

	<i>1111 1256677 78889</i>
	<i>0019 9660949 91190</i>
	<i>0253 4703700 14706</i>
Lp1	TGTT-GGCCGGG-CACCC
Lp2	T.-.....
Lp3	TT....G.GAT.
Lp4	..G.....G.T.
Lp5	T.-..G.GAT.
Lp6	C.....G.T.
Lp7	G.T.
Lm	A...G..A
Oo	.A..T...T.....G...

Figure 3

Figure 4