ACADEMIC PRESS

Available online at www.sciencedirect.com

SCIENCE DIRECT®

VIROLOGY

Virology 316 (2003) 245–255

www.elsevier.com/locate/yviro

The prophage sequences of *Lactobacillus plantarum* strain WCFS1

Marco Ventura,[a,1,2] Carlos Canchaya,[a,1] Michiel Kleerebezem,[b] Willem M. de Vos,[b] Roland J. Siezen,[b,c] and Harald Brüssow[a,*]

[a] *Nestle Research Center, CH-1000 Lausanne 26, Vers-chez-les-Blanc, Switzerland*

[b] *Wagening Centre for Food Sciences, Wageningen, The Netherlands*

[c] *Center for Molecular and Biomolecular Informatics, University of Nijmegen, The Netherlands*

Received 16 April 2003; returned to author for revision 15 June 2003; accepted 3 August 2003

Abstract

The *Lactobacillus plantarum* commensal WCFS1 contains four prophage elements in its genome. Lp1 and Lp2 are two about 40-kb-long uninducible prophages that share closely related DNA packaging, head and tail genes defining a second lineage of *pac*-site *Siphoviridae* in *L. plantarum*, distinct from *L. plantarum* phage phig1e, but related to *Bacillus* phage SPP1 and *Lactococcus* phage TP901-1. Northern analysis revealed transcribed prophage genes exclusively near both attachment sites. Comparative genomics identified candidate lysogenic conversion genes (LCG) downstream of the lysis cassette and within the lysogeny module. Notable are genes with sequence similarities to putative LCG from *Streptococcus pyogenes* prophages and to a *Bacillus* plasmid. Both prophages harbored tRNA genes. R-Lp3 and R-Lp4 represent short prophage remnants; R-Lp3 abuts Lp2 and displays sequence links to *cos*-site *Siphoviridae.*

Introduction

One of the new insights from genome sequencing was the important role of bacteriophages for bacterial genome evolution (Bushman, 2002; Canchaya et al., 2003c). Prophages are not only quantitatively important genetic elements of the bacterial chromosome. Microarray analysis showed that prophages account in several bacterial species for a substantial amount of interstrain genetic variability (Baba et al., 2002; Bhattacharyya et al., 2002; Canchaya et al., 2003a; Chan et al., 2003; Murray et al., 2001; Porwollik et al., 2002; Smoot et al., 2002; Ventura et al., 2003). Notably, numerous virulence factors from bacterial pathogens are phage-encoded (Boyd and Brüssow, 2002; Wagner and Waldor, 2002). Polylysogeny, the presence of multiple prophage genomes in a bacterial host, offers the possibility of combining different virulence factors by a permutation principle. In fact, the emergence of highly virulent clones of *Streptococcus pyogenes* over recent decades was attributed to the sequential acquisition of three specific prophages with their distinct phage-encoded virulence factors (Banks et al., 2002; Beres et al., 2002).

The prophage–host interaction has been rationalized in simple evolutionary models (Canchaya et al., 2003a; Desiere et al., 2001; Lawrence et al., 2001; Lawrence and Roth, 1999) for both pathogenic and nonpathogenic bacteria (Brüssow and Hendrix, 2002). Prophages were postulated to contribute genes that increase the survival fitness of the lysogenic clone in its specific ecological niche. Interestingly, at the position where *S. pyogenes* prophages encode virulence factors, i.e., between the phage lysin gene and the right phage attachment site, prophages from many nonpathogenic low GC content Gram-positive bacteria also possessed genes that could not be attributed to any known phage function and lacked any database matches (Ventura et al., 2002b). Furthermore, transcription studies in *Streptococcus thermophilus* and *Lactococcus lactis* prophages demonstrated that these genes belong to the few prophage genes transcribed in the lysogenic state (Boyce et al., 1995; Ventura et al., 2002a). In the present article, we extended the prophage transcription analysis to lactobacilli. These

* Corresponding author.

E-mail address: harald.bruessow@rdls.nestle.com (H. Brussow).

[1] These authors contributed equally to the work.

[2] Present address: University College of Cork, Department of Microbiology, Cork, Ireland.

0042-6822/$ – see front matter © 2003 Elsevier Inc. All rights reserved.
doi:10.1016/j.virol.2003.08.019

bacteria are prominent commensals of the vagina and the alimentary tract. The recently sequenced *Lactobacillus plantarum* strain WCFS1 (Kleerebezem et al., 2003) was isolated from the oral cavity of a human subject.

Results

Prophage localization

Integrases are useful markers for mobile DNA elements (prophages, integrative plasmids, and pathogenicity islands) in bacterial genomes. The *L. plantarum* strain WCFS1 contains nine integrase genes in its 3.3-Mb large genome (Fig. 1A). Two integrase genes (*int 2* and *int 7*) identified the prophages Lp1 and Lp2 (Fig. 1A). *int 8* identified a 14-kb-long prophage remnant R-Lp3 and *int 9* was associated with a likely 10-kb-long prophage remnant R-Lp4 (Fig. 1A). No evidence for a prophage association was obtained for the other five integrases (*int 4* gene is flanked by capsular polysaccharide gene clusters; *int 6* is part of a transposase cluster).

Genome analysis of prophage Lp1

L. plantarum strain WCFS1 did not release phage spontaneously or after mitomycin C induction as assessed by extracellular phage isolation and phage DNA detection by PCR (data not shown). The likely extent of prophage Lp1 was from ORF 624 (phage integrase) to ORF 687 (Fig. 1C). This diagnosis was based on two observations. First, ORFs 624 and 687 were flanked downstream by a 14-bp repeat, suggestive of *attL* and *attR* sites (Fig. 1C). Second, primers placed in the adjacent ORFs 621 and 690 yielded a 2.8-kb-long PCR product with genomic DNA from *L. plantarum* strains lacking prophage Lp1 (Fig. 1B). The length of the PCR product corresponds exactly to the WCFS1 genome minus the Lp1 prophage, resulting in abutting *rplL* (ribosomal protein L12) and *mf2* (mitogenic factor 2) -like genes. Sequencing of this PCR product identified the presence of a single copy of the 14-bp repeat, suggesting an *attB* site (Fig. 1C). The 14-bp core sequence overlaps the very 3′ end of the *rplL* gene. Prophage integration complements exactly the 3′ end of the coding sequence from *rplL*, while the stop codon of *rplL* was changed from TAA to TAG (Fig. 1C).

Database matches allowed a subdivision of the Lp1 prophage genome into functional modules (Fig. 2). The lysogeny module covers ORF 624 to ORF 634. It comprised the integrase gene, the genetic switch region with divergently transcribed *cI*- and *cro*-like repressor genes (ORF 631–ORF 632, 43% aa identity with a genetic switch in *Enterococcus faecalis* and 38% identical with the repressor from *L. plantarum* phage phig1e), a likely superinfection exclusion gene *sie* (orf 628) related to *S. thermophilus* phage TPJ34 (40% aa identity, E value = 10^{-15}) and ORF 625 encoding a protein related to the mitogenic factor gene *mf4* from *S. pyogenes* prophage 315.3 (35% aa identity, E = 10^{-28}). Notably, in *S. pyogenes* prophage, the *mf4* gene is located between the phage lysin gene and the right phage attachment site *attR*. A total of four tRNA genes were located on the Lp1 genome. Interestingly, two flank the mf4-like gene (Fig. 2). A second repressor (ORF 637), with a H-T-H DNA-binding motif, is encoded to the right of the genetic switch region. Downstream of the lysogeny module, many ORFs encode proteins with homology to DNA replication proteins from various *Lactobacillus*, *Lactococcus*, and *Staphylococcus* prophages (Fig. 2). Database matches include RecT-, replisome organizer-, RusA-, and a helicase-like protein defining this genome segment as the likely DNA replication module.

Further database matches allowed a distinction of likely DNA packaging and head morphogenesis genes (the closest relative was *Bacillus subtilis* phage SPP1), possible head-to-tail joining genes (related to lactococcal phages TP901-1 and Tuc2009), and putative tail and tail fiber genes. With the exception of the tail fiber proteins, the structural proteins from prophage Lp1 shared no amino acid sequence identity with the corresponding structural proteins from *L. plantarum* phage phig1e (Kodaira et al., 1997; Kakikawa et al., 1996). Over the structural genes phig1e was closely related to *L. delbrueckii* phage LL-H (Mikkonen and Alatossava, 1994) (Fig. 2). This analysis identified at least two distinct lineages of structural genes in prophages from *L. plantarum:* Lp1 and phig1e. Both lineages belong to the recently proposed Sfi11 species of *pac*-site *Siphoviridae* in lactic acid bacteria (Brüssow, 2001; Brüssow and Desiere, 2001).

Downstream of the putative tail fiber genes, a likely lysis cassette was identified in Lp1. Unusual for prophages from LAB, the putative holin follows the lysin and precedes a further lysin gene fragment, suggesting a DNA rearrangement event.

Two genes at the prophage side of *attR* encode proteins with homology to different parts of a hypothetical protein from the *Bacillus anthracis* plasmid pX02 (gp 71, 28% aa identity, E = 10^{-38}). Interestingly, the two genes at the bacterial side of the putative *attR* site also encoded proteins with homology to phage proteins, i.e., to the mitogenic factor mf2 from *S. pyogenes* prophage 370.1 (32% aa identity, E = 10^{-18}), and to tail fiber (tf) proteins from phages of Gram-negative bacteria (prophages Rac, CP933X, Fels-1, T7-like Podovirus 28% aa identity, E = 10^{-10}) (Fig. 1C).

Genome analysis of prophage Lp2

Prophage Lp2 showed the common genome organization of Sfi11-like *Siphoviridae* (Figs. 2 and 3). However, the exact length of the prophage is unclear: two different direct repeats, R1 and R2 (17 and 15 bp, respectively, located at position 2′162′771–2′204′044 and 2′161′439–2′205′041), flank Lp2. Different primer combinations (primer pair 1–2 from ORF 2457 to ORF 2396, and primer pair 1–3 from ORF 2396 to ORF 2494) failed to amplify DNA from *L.*

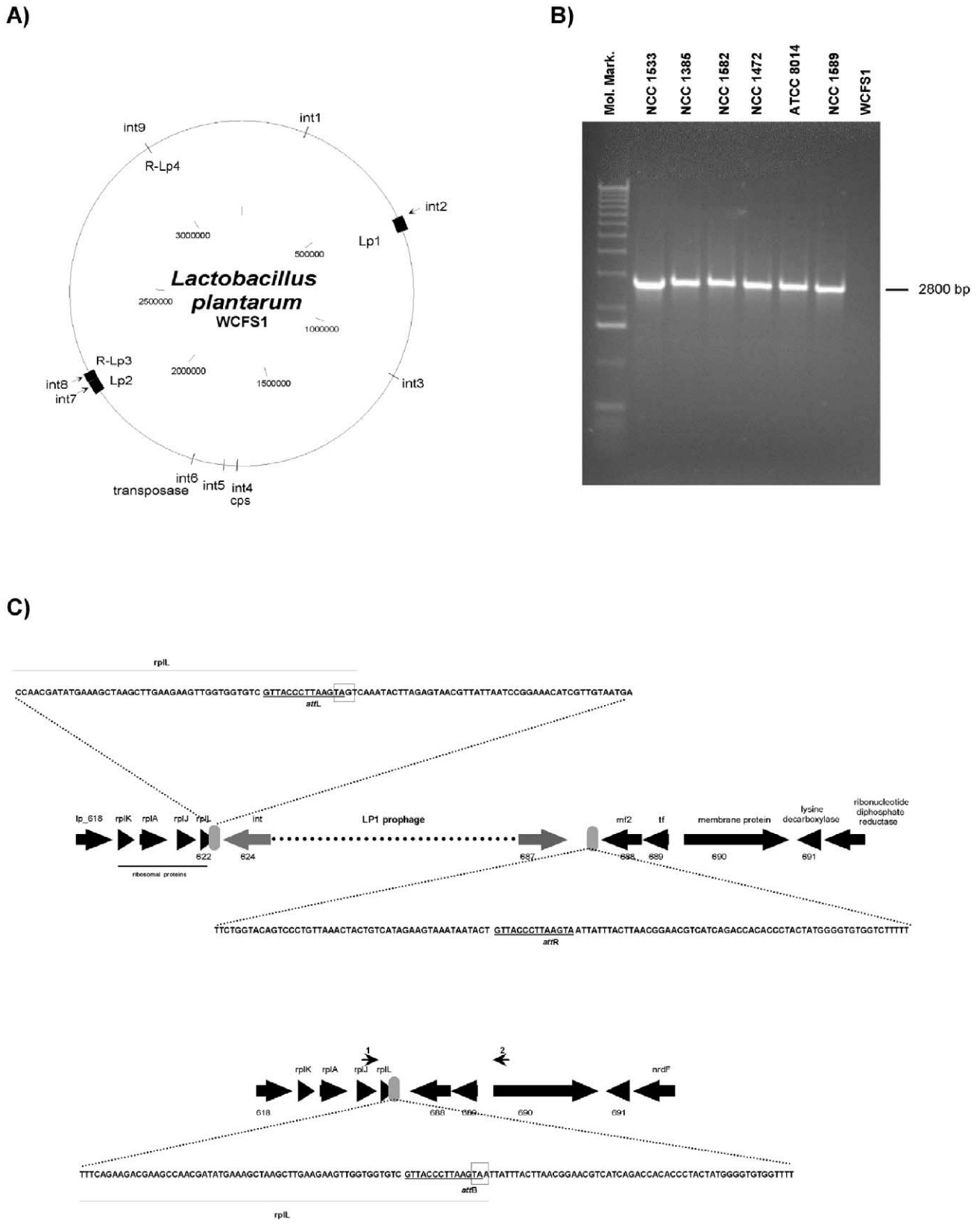

Fig. 1. Prophage integration. (A) The circular genome map of *L. plantarum* strain WCFS1 shows the location of the integrase genes, the prophages Lp1 and Lp2, and the prophage remnants R-Lp3 and R-Lp4. The origin of replication is at the top. The bp position of the prophages is for Lp1: 589′962 to 631′801; for Lp2: 2′161′732 to 2′203′817; for R-Lp3: 2′205′188 to 2′218′339; for R-Lp4: 2′997′299 to 3′007′277. (B) PCR amplification of genomic DNA of the indicated *L. plantarum* strains flanking the putative *attB* site into which Lp1 integrated. The location of the primers used in this PCR is indicated in C, bottom, with black arrows 1 and 2. (C) Integration site of Lp1. The bacterial genes are in black and are annotated with their ORF number and a shorthand notation of their function. The outmost prophage genes are shown in dark grey. The light grey vertical bars represent the likely attachment sites *att*L, *att*R, and *att*B. The nt sequence of these sites is indicated in the enlarged inserts. The conserved core sequence of each attachment site is underlined. The 3′ end of ORF 622 (encoding the ribosomal protein L12 *rplL*) is indicated by a line and the stop codon is boxed.

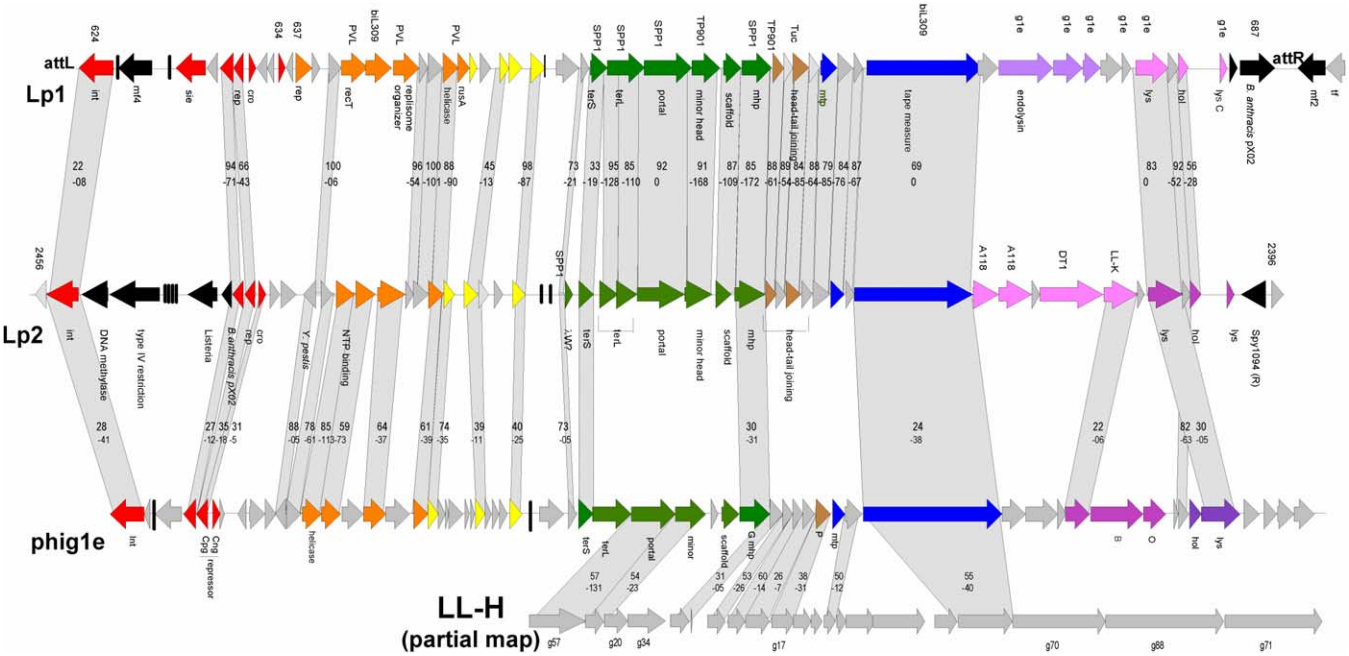

Fig. 2. Comparative genome maps of Lp1 and Lp2 prophages. Alignment of the genome maps of the *L. plantarum* prophages Lp1, Lp2, and phig1e with a partial genome map of *L. delbrueckii* prophage LL-H. Genes sharing similarity are linked by gray shading and the percentage of aa identity and the E value is given as exponents. Probable functions of encoded proteins identified by bioinformatic analysis are noted below the ORFs. The modular structure of the genomes is indicated by colors (red: lysogeny; orange: DNA replication; yellow: transcriptional regulation?; green: DNA packaging and head; brown: head-to-tail joining; blue: tail; mauve: tail fiber; violet: lysis modules; black: candidate lysogenic conversion genes; vertical black lines: tRNA genes).

plantarum strains (ATCC 8014, NCC1385, NCC1582, NCC1472, NCC1589) lacking Lp2.

Lp2 is closely related to Lp1 at the protein (Fig. 2) and at the DNA sequence level. With the exception of the putative tail fiber genes, both prophages shared essentially the entire structural gene cluster. Eighty-eight percent of DNA sequence identity was shared between the DNA packaging, head, head-to-tail, and tail genes from Lp1 and Lp2. In this region only few differences were detected between the two prophages: The putative small subunit terminase genes were only related at protein sequence level: a nonsense mutation disrupted the large subunit terminase gene in Lp2, and an internal segment of the putative tail tape measure gene lacked DNA sequence similarity. The tail fiber genes from Lp2 lacked similarity with Lp1, but showed homology with *Listeria* phage A118 and *Lactobacillus* phage LL-K (Fig. 2). The adjacent lysis cassette is similarly organized in Lp1 and Lp2. ORF2397, located downstream of the lysin in Lp2, is 32% aa identical to a gene of prophage remnant Spy1094 from *S. pyogenes* strain SF370, suggesting a lysogenic conversion gene (LCG). Further candidates for LCG were found between the integrase gene and the genetic switch region: a DNA methylase (41% aa identity with *Listeria*) and a protein with weak similarity to a type IV restriction-like enzyme, followed further upstream by four tRNA genes, and two genes encoding proteins with similarity to hypothetical proteins from *Listeria* (51% aa identity) and from *B. anthracis* plasmid pX02 (gp66, note that Lp1 ORF 687 shares similarity with the adjacent gp71 of plasmid pX02). Over the nonstructural genes, DNA sequence similarity between Lp1 and Lp2 was limited to single genes or small groups of genes in the lysis cassette and the DNA replication module (Fig. 2).

Genome analysis of prophage remnants R-Lp3 and R-Lp4

Directly downstream of the Lp2 integrase gene, another cluster of 14 phage-related genes was identified (ORF 2457–2480). The cluster belongs to an apparently truncated prophage remnant R-Lp3 (Fig. 4A). In R-Lp3 the integrase gene is preceded by a repressor-like gene and on the opposite strand by two phage DNA replication genes with sequence similarities to the helicase gene of lactococcal phage blL312. Further upstream, DNA packaging genes (encoding small and large terminase, portal protein) and head genes (encoding protease and major head protein) were identified (Fig. 4A). Their closest relatives were from *Lactobacillus* phage A2 and *Staphylococcus* phage PVL, two Sfi21-like *cos*-site *Siphoviridae* (Proux et al., 2002). R-Lp3 thus belongs to a third lineage of structural genes in *L. plantarum.* The prophage remnant R-Lp3 was apparently shaped by DNA deletion and rearrangement events. Not only are the lysogeny and DNA replication modules much shorter than in replication-competent dairy phages, there is also a "misplaced" head-to-tail gene between DNA replication and DNA packaging genes. A fusion gene (ORF 2463) encodes the N-terminus of a proteinase and the C-terminus of the major head protein, apparently caused by the loss of an

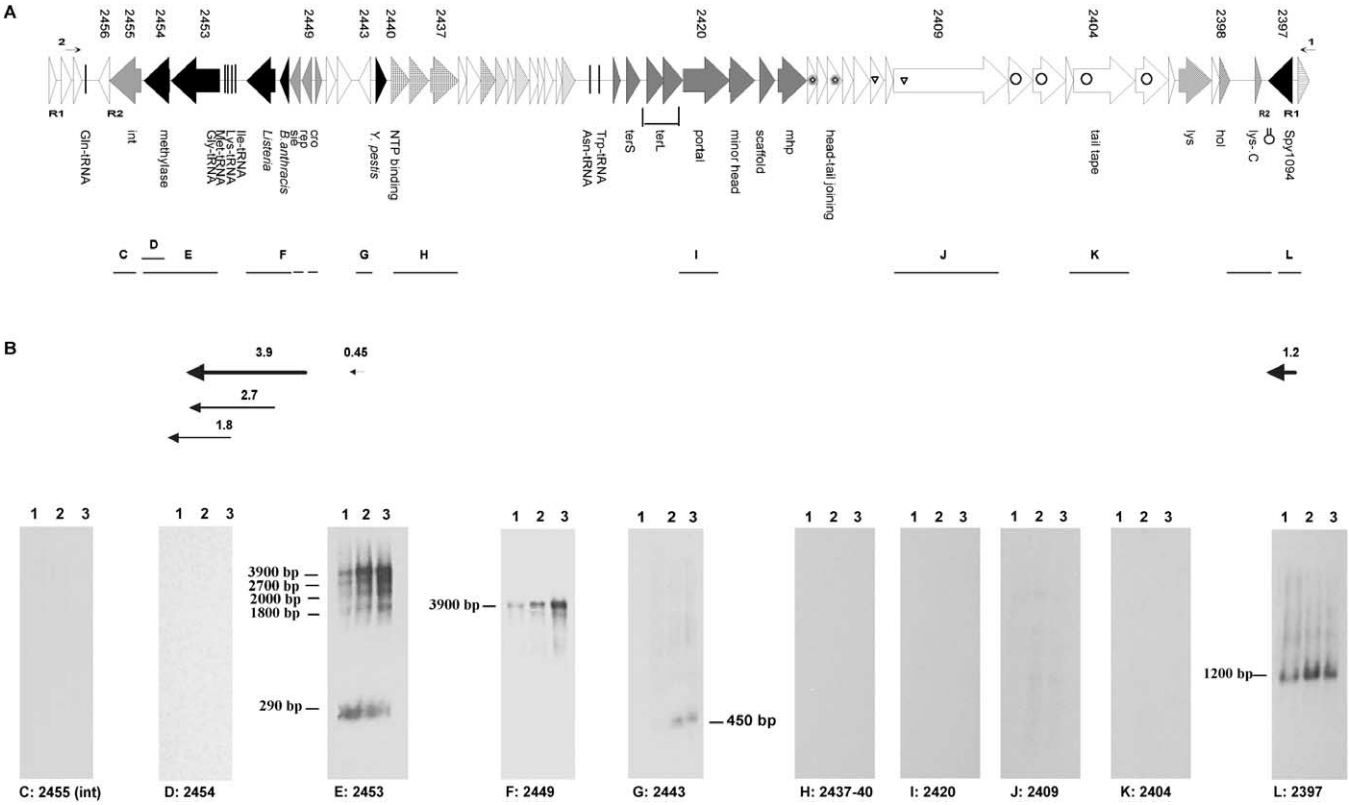

Fig. 3. Genome and transcription analysis of prophage Lp2. (A) Genome map of prophage Lp2. Genes belonging to the same module are marked with the same shading, see Fig. 2 for more details. Selected genes are annotated with their ORF number (top) and bioinformatic links (bottom). Arrows 1 and 2 are the positions of the primers for the unsuccessful *att*B PCR. R1 and R2 are repeats flanking Lp2. The hairpin symbolizes a rho-independent terminator with an energy >15 kcal/mol. (B) Transcription map. The position of the probes against which the Northern blots were hybridized is provided at the top line; the letters refer to the Northern blots shown below. The arrows indicate the likely position and orientation of the transcripts; the numbers give the lengths of the transcripts in kb; the thickness of the arrow indicates the autoradiographic intensity of the hybridization results. (C–L) Northern blot hybridization of total RNA from *L. plantarum* WCFS1 isolated at an OD of 0.4, 0.7, and 1.4, respectively (lanes 1–3). The blots were probed with radiolabeled PCR products covering the indicated regions of the genome (see top line of B) and the annotated genes. The molecular weights calculated for the hybridization signals are provided in bp.

internal DNA segment. Two likely head-to-tail joining genes and a tRNA gene followed the fusion gene (Fig. 4A).

A further prophage remnant R-Lp4, not flanked by DNA repeats, was identified between ORFs 3374 and 3392 (Fig. 4D). In this region an integrase and a possible genetic switch region remained from the lysogeny module and three genes showed homology to a phage DNA replication module. Interestingly, ORF 3378, which followed directly the DNA replication genes, encodes a protein with similarity to two hypothetical proteins encoded on the pathogenicity island of *E. faecalis* (Shankar et al., 2002).

Expression analysis of prophage genes

Only expressed prophage genes can confer a positive lysogenic conversion phenotype to its host. Therefore we tested the candidate LCG for transcription.

Lp1

The lysogeny module of Lp1 gave rise to three transcripts (Fig. 5B). A probe covering the repressor (ORF 631) and ORF 630 detected a 3.2-kb transcript (Fig. 5G). A 3.2-kb transcript was also revealed with a probe corresponding to the putative superinfection exclusion gene, ORF 628 (Fig. 5F). The *mf4*-like gene probe detected a prominent 1.2-kb transcript (Fig. 5E). A very weak 1.2-kb transcript was also revealed with an integrase-specific probe (Fig. 5D). In contrast, no transcripts were detected with probes covering prospective early lytic (Fig. 5H), DNA replication (Figs. 5I–J), transcriptional regulator (Fig. 5K), several structural (Figs. 5L–N and data not shown), or the phage holin gene (ORF 685, Fig. 5O). Probes corresponding to the two ORFs between the phage lysin gene fragment and *attR* yielded a 1.4-kb-long messenger (Fig. 5P), while the *mf2*-like gene located just outside of the confines from Lp1 was not transcribed (Fig. 5Q)

Lp2/R-Lp3

The bioinformatic analysis of Lp2 identified ORF 2397 as a candidate for a lysogenic conversion gene. Northern hybridization indeed revealed a 1.2-kb-long transcript when using a probe covering this ORF (Fig. 3L). A major 3.9-kb

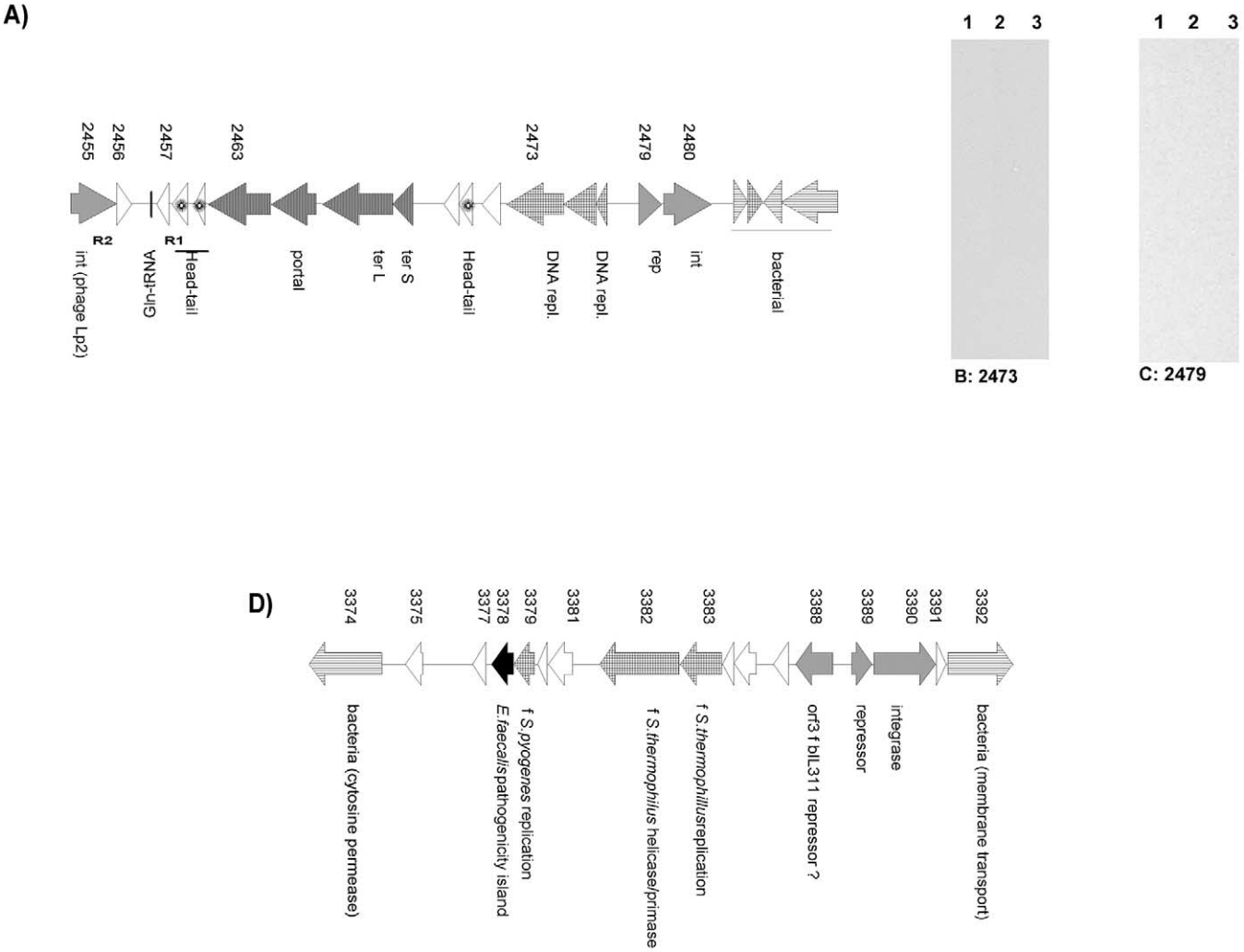

Fig. 4. Genome map of the prophage remnants R-Lp3 (A) and R-Lp4 (D). The ORFs are annotated with numbers and a shorthand notation according to their database matches. Genes are shaded according to their putative function (see Fig. 3 for code). (B–C) Northern blot hybridization of total RNA from *L. plantarum* WCFS1 isolated at an OD of 0.4, 0.7, and 1.4, respectively (lanes 1–3). The blots were probed with radiolabeled PCR products covering the indicated genes.

transcript and a series of smaller transcripts ranging in size from 2.7 to 0.3 kb were observed with probes covering the Lp2 repressor gene (ORF 2449) or genes downstream of the phage repressor (Figs. 3E–F). However, no transcripts were detected when using probes specific for the DNA methylase (ORF 2454) and the integrase gene (ORF 2455) from Lp2 (Figs. 3D and C). Lp2 contains a second constellation of divergently oriented ORFs near the putative genetic switch. An ORF 2443-specific probe revealed a small transcript of 450 bp (Fig. 3G). No transcripts were seen with probes covering the DNA replication (Fig. 3H) and several structural genes (Fig. 3I–K). No transcripts were observed when using probes from a DNA replication or the putative repressor gene of the prophage remnant R-Lp3 (Figs. 4B–C).

Distribution of Lp1- and Lp2-like phages

We selected five *L. plantarum* isolates from silage and various fermented food products of our collection and the *L. plantarum* ATCC8014 strain for analyzing the distribution of Lp1- and Lp2-like prophages. The six strains represented four different S*fi*I restriction patterns in pulsed-field gel electrophoresis; all were distinct from the pattern of the reference strain WCFS1 (Fig. 6A). Two Lp1-specific probes, located near the genetic switch region (ORF 633) and between the lysis cassette and the *attR* site (ORF 675), respectively, hybridized exclusively with a 190-kb fragment from WCSF1 DNA (Fig. 6C and data not shown), suggesting the absence of Lp1-like prophages in the test strains. An Lp2-specific probe covering the ORF 2454 hybridized exclusively with a 170-kb fragment from WCSF1 DNA (data not shown). In contrast, a probe from the lysis cassette of Lp2 (ORF 2398/2399) showed cross-hybridization with three test strains representing two different pulse field patterns (Fig. 6B), suggesting the presence of further prophages containing Lp2-like lysis cassettes. Indeed, prophages could be induced with mitomycin C from cells NCC1533

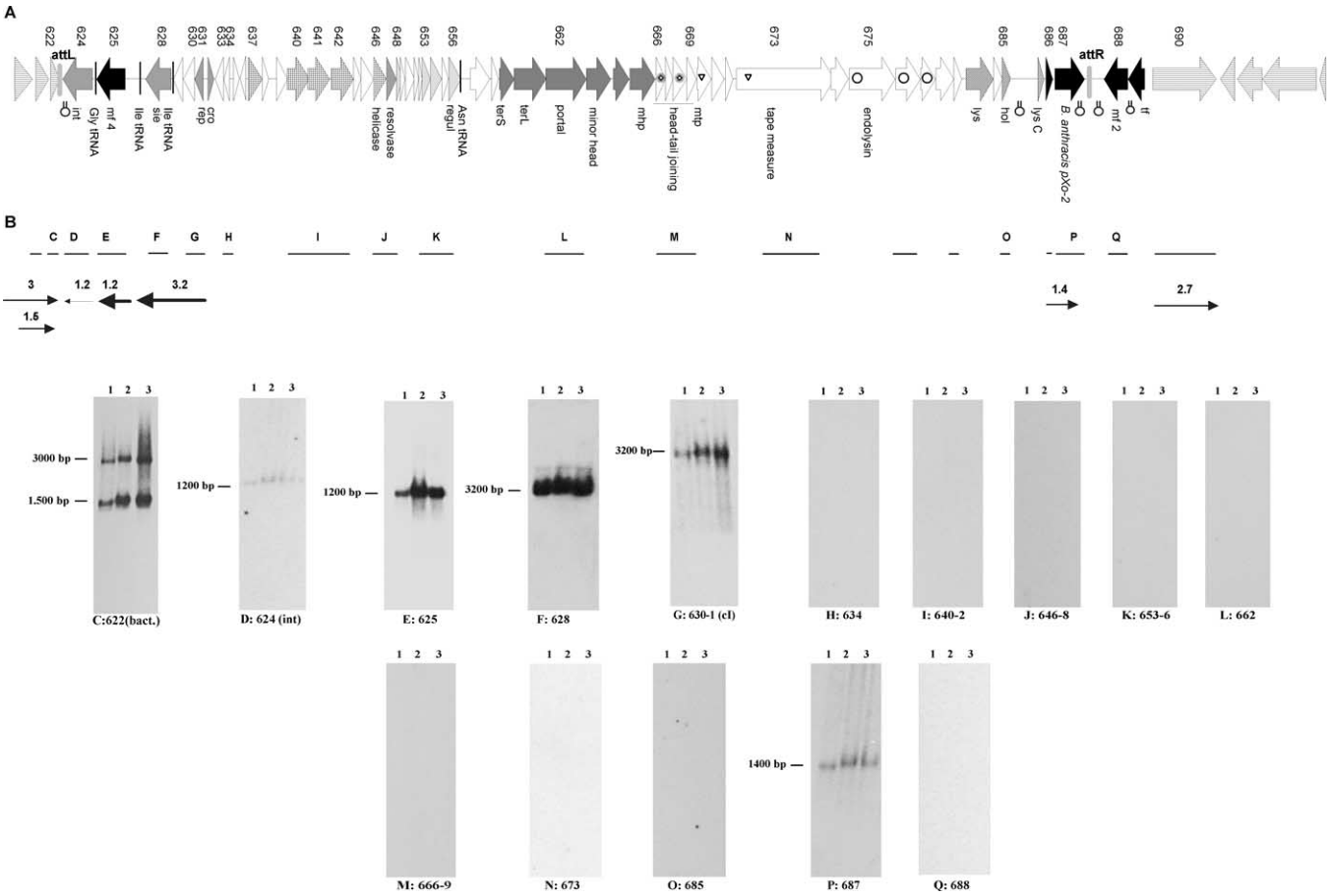

Fig. 5. Genome map and transcription analysis of prophage Lp1. (A) Genome map of the prophage Lp1. Genes are shaded according to their putative function (see Fig. 3 for code).The extent of prophage is defined by the *attL* and *attR* sites. Adjacent bacterial genes are shaded as in Fig. 4A (see Fig. 1C for their description). The hairpins symbolize rho-independent terminators with an energy >15 kcal/mol. (B) Partial transcription map of prophage Lp1. See Fig. 3 for explanations. (C–Q) Northern blot hybridization of total RNA from *L. plantarum* WCFS1 probed with radiolabeled PCR products covering the indicated genes and located on the map in the top line of B. The molecular weights calculated for the hybridization signals are provided in bp. A blot probed with the bacterial gene ORF 622 is given in C for a comparison of the hybridization intensity.

and NCC1472. By electron microscopy, *Siphoviridae* were visualized in the cell-free supernatant (data not shown).

Discussion

Prophages Lp1 and Lp2 were not released spontaneously from WCFS1 nor could they be induced by mitomycin C treatment. The reason for this noninduction is unclear in the case of Lp1, which showed a standard genome organization for an Sfi11-like siphovirus. Apparent DNA rearrangements occurred, however, in the lysis cassette. In contrast, the genome analysis of Lp2 suggested a defective prophage. First, Lp2 is not any longer flanked by *attL* and *attR* sequences characteristic for Campbell-like prophage integration events. In addition, directly adjacent to Lp2 a definitively truncated prophage remnant R-Lp3 was identified, suggesting that the Lp2/R-Lp3 genome region is the result of a series of chromosomal rearrangement events that resulted in the inactivation of the two involved prophages. Second, the large subunit of the terminase gene of Lp2 is interrupted by a stop codon. Similar inactivating point mutations have been observed in the genes coding for the portal protein or the replisome organizer in noninducible *S. pyogenes* prophages (Desiere et al., 2001).

The R-Lp3 and R-Lp4 are clearly defective prophages: genes from the lysogeny and the DNA replication modules are missing as are the entire tail, tail fiber, and lysis modules. In addition, a deletion in R-Lp3 led to the fusion of the 5′ half of the protease gene to the 3′ half of the major head gene. It is fascinating that very similar constellations of prophage genes were observed in prophage remnants from *S. pyogenes* (Canchaya et al., 2002) and *L. lactis* (Chopin et al., 2001).

Lp1 and Lp2 shared a high level of uninterrupted DNA sequence identity of 88% over about 10 kb. This observation is striking, but not without precedence. Two pairs of prophages in the *Xylella fastidiosa* strain 9a5c shared a high level of DNA sequence identity (Simpson et al., 2000) and a large set of lambda-like prophages from *Escherichia coli*

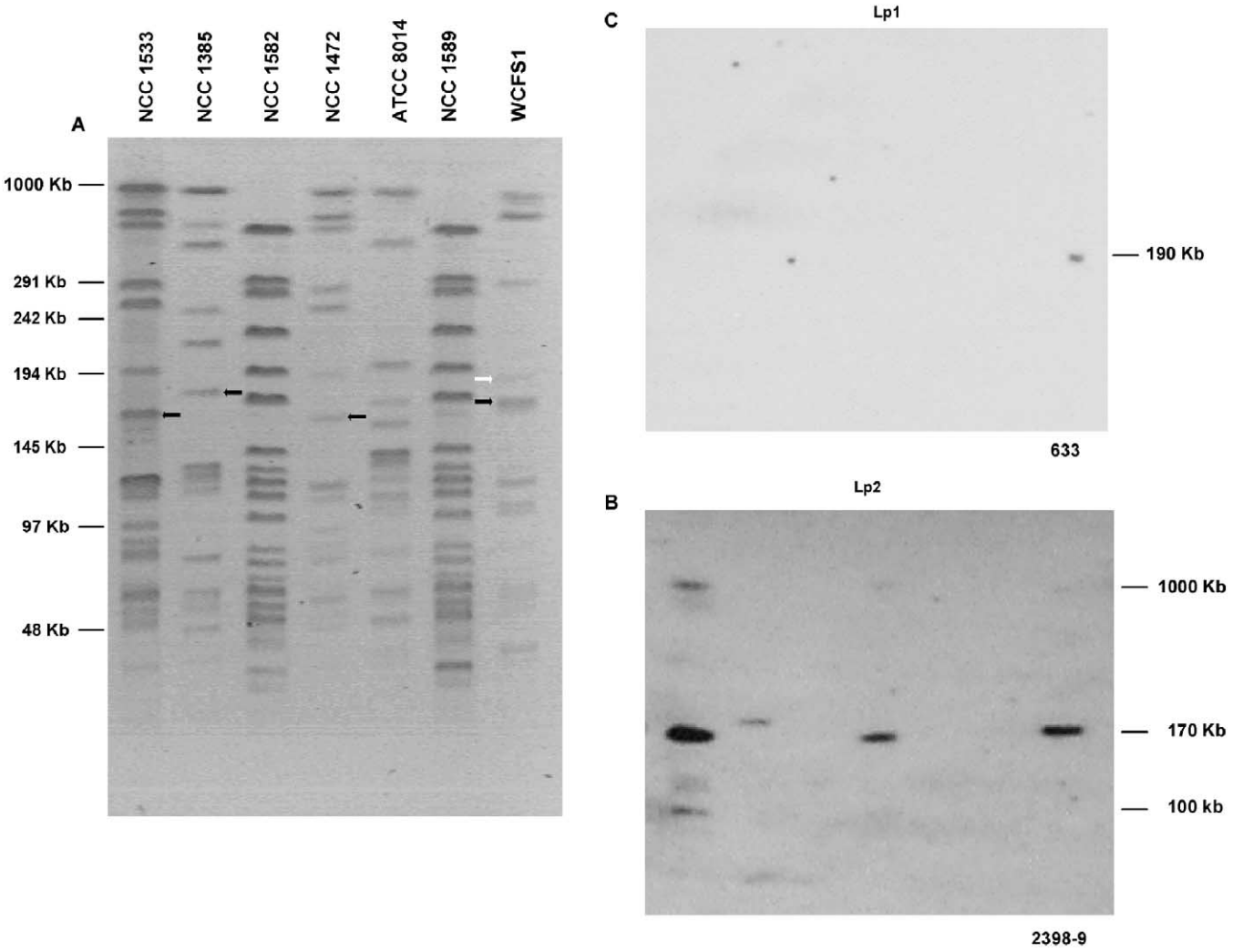

Fig. 6. Integration and distribution of Lp1 and Lp2 prophages in *L. plantarum* strains. (A) Pulsed field gel electrophoresis of S*fi*l-digested DNA from the indicated *L. plantarum* strains stained with ethidium bromide (negative of the photo). The molecular weight of the markers is given at the left. Bands hybridizing with the Lp2 and the Lp1 probes are marked with black or white arrows, respectively. (B–C) Corresponding Southern hybridization with the radiolabeled PCR probes from the ORF 2398–2399 of Lp2 (B) and ORF633 of Lp1 (C).

O157:H7 demonstrated extensive DNA sharing (Pema et al., 2001; Boyd and Brüssow, 2002; Ohnishi et al., 2001). Notably, one might predict instability of bacterial genomes containing two large DNA repeats as demonstrated by *X. fastidiosa* strain comparisons (Van Sluys et al., 2003).

S. pyogenes and *L. plantarum* occupy clearly distinct habitats. The first is an exclusive human pathogen living on the skin and in the oropharynx and nasal cavity, while the latter is commonly isolated from silage, sauerkraut, dairy products, cow dung, human intestinal tract, stools, and sewage. However, *L. plantarum* can also be isolated from human mouth; this was actually the isolation history of the sequenced WCFS1 strain (Kleerebezem et al. 2003). A recently formulated hypothesis proposes that prophages encode genes that are of selective benefit for the lysogen in their respective ecological setting (Canchaya et al., 2003a). In this context the sharing of distantly related, prospective LCG between prophages from *L. plantarum* and *S. pyogenes* is a fascinating observation. *S. pyogenes* prophages contribute enzymes (hyaluronidase, DNase, phospholipase), toxins, and superantigens (factors that hyperstimulate and disregulate the immune system) to the lysogen (Banks et al., 2002). Since commensal lactobacilli are discussed as probiotic organisms that stimulate the immune system in a nonspecific way (Ibnou-Zekri et al., 2003; Kaila et al., 1992), the observation of mitogenic factor-like genes within (*mf4*) and directly adjacent (*mf2*) of Lp1 is startling. However, the identity between the Lp1 genes and *mf2* and *mf4* is low (32–35% aa identity) and no definitive role has yet been attributed to these genes in *S. pyogenes* prophages. Some reports attribute a mitogenic activity on blood lymphocytes to the cloned and expressed protein (Iwasaki et al., 1993), while other authors report a DNase (streptodornase) activity (Podbielski et al., 1996; Sriskandan et al., 2000).

Previous genome analysis of prophages from low GC content Gram-positive bacteria identified candidate LCG in

the region between the lysin gene and the *attR* site. The current study identified prospective LCG within the lysogeny module. This location can be easily rationalized. The promoter of the *cI* repressor is constitutively active in the lysogen and genes located downstream of the *cI* repressor gene will be cotranscribed if not separated by a terminator. Further active promoters were previously detected between the *cI* and integrase genes (e.g., *sie* gene; Ventura et al., 2002a). Apparently, additional genes can be plugged into the lysogeny module without upsetting the lysogenic state as long as their promoters do not lead to transcription into the early lytic genes (rightward transcription) or activate integrase expression. Compared with the standard lysogeny module from prophages of lactic acid bacteria (Lucchini et al., 1999), Lp1 and Lp2 showed an impressive array of extra genes ["morons", i.e., more DNA in the terminology of the Pittsburgh phage group (Juhala et al., 2000)] in the lysogeny module. tRNA genes are a feature in both: Lp1 shows the *mf4*-like gene; Lp2 contains two genes that resemble weakly a restriction-modification system (a *Bacillus halodurans* prophage encodes a clear R-M system near its *attR* site, Canchaya et al. 2003b); Lp2 encodes a gene related to a gene from *B. anthracis* plasmid pX02. The function is unknown, but it is intriguing that Lp1 encodes near its *attR* site a gene that resembles another nearby gene from the same pX02 plasmid. However, none of these genes provides a clear lead to a lysogenic conversion phenotype. We clearly need more data from other prophages of *Lactobacillus* commensals to assess the role of prophages for the phenotype of this industrially and medically important class of lactic acid bacteria.

Materials and methods

Strain and culture conditions

Lactobacillus strains were grown anaerobically in MRS agar or broth (Difco, USA) for 24 h at 30°C. The strains used in this study were ATCC8014 and NCC1385, NCC1472, NCC1533, NCC1582, NCC1589 obtained from different culture collections [ATCC (American Type Culture Collection) and NCC (Nestlé Culture Collection)].

RNA isolation and Northern blot analysis

Total RNA was isolated by resuspending bacterial cell pellets in TRIzol (Gibco-BRL, UK), adding 106-μm glass beads (Sigma, USA), and shearing the slurry with a Mini-Beadbeater-8 cell disrupter (Biospec Products, USA) as described by Ventura et al., (2002b). Northern blot analysis of *Lactobacillus* prophage transcription was carried out on 15-μg aliquots of RNA isolated from 10 ml of *Lactobacillus* culture, collected at an optical density of 0.4, 0.7, and 1.4 (OD 600 nm). The RNA was separated in a 1.5% agarose-formaldehyde denaturing gel, transferred to a Zeta-Probe blotting membrane (Bio-Rad, UK) according to Sambrook (2001), and fixed by UV cross-linking using a Stratalinker 1800 (Stratagene, USA). Prehybridization and hybridization were carried out at 65°C in 0.5 M $NaHPO_4$ (pH 7.2), 1.0 mM EDTA, and 7.0% sodium dodecylsulfate (SDS). Following 18 h of hybridization, the membrane was rinsed twice for 30 min at 65°C in 0.1 M $NaHPO_4$ (pH 7.2), 1.0 mM EDTA, 1% SDS, twice for 30 min at 65°C in 0.1 mM $NaHPO_4$ (pH 7.2), 1.0 mM EDTA, 0.1% SDS and exposed to X-OMAT autoradiography film (Eastman Kodak, USA).

The probes for the Northern blot hybridization, located in Figs. 3B and 5B, were labeled with [α^{32}P] by using the random-primed DNA labeling system (Boehringer Mannheim GmbH) and purified with Nuc Trap probe purification columns (Stratagene).

DNA sequencing

The prophage sequences were retrieved from the genome sequence of *L. plantarum* WCFS1, accession code AL935263 (Kleerebezem et al., 2003). The *attB* site for Lp1 was PCR amplified from *L. plantarum* strains lacking a prophage at this position by using the primers 57-c (5′-CATGACGTCATACGTCAG-3′) and 3-b (5′-CTAAGGATTCATACGATGTTG-3′). The PCR products were sequenced on both strands using the fluorescent-labeled primer-sequencing kit (Amersham, Germany) following the supplier's instructions.

Pulsed-field gel electrophoresis (PFGE) and Southern blots

Agarose-embedded bacterial cells were prepared as described by Walker and Klaenhammer (1994). For digestion with restriction endonuclease, cells in agarose blocks were treated with 50 units of S*fi*I (Roche Molecular, UK) as described by the manufacturer. The digestion was stopped by washing the blocks for 20 min in TE buffer. PFGE was performed by a contour-clamped homogeneous electric field mode in a CHEF-DRII apparatus (Bio-Rad, USA). All DNA samples were separated in 1% agarose gels in 0.5× Tris borate EDTA (TBE) buffer, cooled to 14°C for 18 h at 6 V/cm, ramping pulse time from 1 to 20 s.

Southern blots of agarose gels were performed on Hybond N+ membranes (Amersham, UK) following the protocols of Sambrook (2001). The filters were hybridized with the radiolabeled PCR generated probe described in the text. Subsequent prehybridization, hybridization, and autoradiography were carried out according to Sambrook (2001).

Sequence analysis

Open reading frames have been predicted using FrameD and adjustments to the prediction with the ORF Predicter (NCBI) using ATG and GTG as possible start codons and a minimum size of 50 aa. Nucleotide and predicted amino

acid sequences were compared to those in the databases of GenBank; EMBL; PIR-Protein; SWISS-PROT; and PROPOSITE. Additional database searches have been conducted using BLAST (Altschul et al., 1997) and PSI-BLAST at the NCBI and FASTA (Lipman and Pearson, 1985). A search for tRNA genes was done using the tRNAscan-SE program (Lowe and Eddy, 1997). The motif search was performed by using Pfam.

Acknowledgments

We thank the Swiss National Science Foundation for the financial support of Marco Ventura and Carlos Canchaya (research Grant 5002-057832).

References

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

Baba, T., Takeuchi, F., Kuroda, M., Yuzawa, H., Aoki, K., Oguchi, A., Nagai, Y., Iwama, N., Asano, K., Naimi, T., Kuroda, H., Cui, L., Yamamoto, K., Hiramatsu, K., 2002. Genome and virulence determinants of high virulence community-acquired MRSA. Lancet 359, 1819–1827.

Banks, D.J., Beres, S.B., Musser, J.M., 2002. The fundamental contribution of phages to GAS evolution, genome diversification and strain emergence. Trends Microbiol. 10, 515–521.

Beres, S.B., Sylva, G.L., Barbian, K.D., Lei, B., Hoff, J.S., Mammarella, N.D., Liu, M.Y., Smoot, J.C., Porcella, S.F., Parkins, L.D., Campbell, D.S., Smith, T.M., McCormick, J.K., Leung, D.Y., Schlievert, P.M., Musser, J.M., 2002. Genome sequence of a serotype M3 strain of group A *Streptococcus*: phage-encoded toxins, the high-virulence phenotype, and clone emergence. Proc. Natl. Acad. Sci. USA 99, 10078–10083.

Bhattacharyya, A., Stilwagen, S., Ivanova, N., D'Souza, M., Bernal, A., Lykidis, A., Kapatral, V., Anderson, I., Larsen, N., Los, T., Reznik, G., Selkov Jr, E, Walunas, T.L., Feil, H., Feil, W.S., Purcell, A., Lassez, J.L., Hawkins, T.L., Haselkorn, R., Overbeek, R., Predki, P.F., Kyrpides, N.C., 2002. Whole-genome comparative analysis of three phytopathogenic *Xylella fastidiosa* strains. Proc. Natl. Acad. Sci. USA 99, 12403–12408.

Boyce, J.D., Davidson, B.E., Hillier, A.J., 1995. Identification of prophage genes expressed in lysogens of the *Lactococcus lactis* bacteriophage BK5-T. Appl. Environ. Microbiol. 61, 4099–4104.

Boyd, E.F., Brüssow, H., 2002. Common themes among bacteriophage-encoded virulence factors and diversity among the bacteriophages involved. Trends Microbiol. 10, 521–529.

Brüssow, H., 2001. Phages of dairy bacteria. Annu. Rev. Microbiol. 55, 283–303.

Brüssow, H., Desiere, F., 2001. Comparative phage genomics and the evolution of Siphoviridae: insights from dairy phages. Mol. Microbiol. 39, 213–222.

Brüssow, H., Hendrix, R.W., 2002. Phage genomics: small is beautiful. Cell 108, 13–16.

Bushman, F., 2002. Lateral DNA Transfer. CSH Laboratory Press, New York.

Canchaya, C., Desiere, F., McShan, W.M., Ferretti, J.J., Parkhill, J., Brüssow, H., 2002. Genome analysis of an inducible prophage and prophage remnants integrated in the *Streptococcus pyogenes* strain SF370. Virology 302, 245–258.

Canchaya, C., Fournous, G., Chibani-Chennoufi, S., Dillmann, M.-L., Brüssow, H. 2003b. Phage as agents of lateral gene transfer. Curr. Opin. Microbiol. 6, 417–424.

Canchaya, C., Proux, C., Fournous, G., Bruttin, A., Brüssow, H., 2003a. Prophage genomics. Microbiol. Mol. Biol. Rev. 67, 238–276.

Chan, K., Baker, S., Kim, C.C., Detweiler, C.S., Dougan, G., Falkow, S., 2003c. Genomic comparison of *Salmonella enterica* serovars and *Salmonella bongori* by use of an *S. enterica* serovar typhimurium DNA microarray. J. Bacteriol. 185, 553–563.

Chopin, A., Bolotin, A., Sorokin, A., Ehrlich, S.D., Chopin, M., 2001. Analysis of six prophages in *Lactococcus lactis* IL1403: different genetic structure of temperate and virulent phage populations. Nucleic Acids Res. 29, 644–651.

Davis, B.M., Waldor, M.K., 2003. Filamentous phages linked to virulence of *Vibrio cholerae*. Curr. Opin. Microbiol. 6, 35–42.

Desiere, F., McShan, W.M., van Sinderen, D., Ferretti, J.J., Brüssow, H., 2001. Comparative genomics reveals close genetic relationships between phages from dairy bacteria and pathogenic *Streptococci*: evolutionary implications for prophage-host interactions. Virology 288, 325–341.

Ibnou-Zekri, N., Blum, S., Schiffrin, E.J., von der Weid, T., 2003. Divergent patterns of colonization and immune response elicited from two intestinal *Lactobacillus* strains that display similar properties in vitro. Infect. Immun. 71, 428–436.

Iwasaki, M., Igarashi, H., Hinuma, Y., Yutsudo, T., 1993. Cloning, characterization and overexpression of a Streptococcus pyogenes gene encoding a new type of mitogenic factor. FEBS Lett. 331, 187–192.

Juhala, R.J., Ford, M.E., Duda, R.L., Youlton, A., Hatfull, G.F., Hendrix, R.W., 2000. Genomic sequences of bacteriophages HK97 and HK022: pervasive genetic mosaicism in the lambdoid bacteriophages. J. Mol. Biol. 299, 27–51.

Kaila, M., Isolauri, E., Soppi, E., Virtanen, E., Laine, S., Arvilommi, H., 1992. Enhancement of the circulating antibody secreting cell response in human diarrhea by a human *Lactobacillus* strain. Pediatr. Res. 32, 141–144.

Kakikawa, M., Oki, M., Tadokoro, H., Nakamura, S., Taketo, A., Kodaira, K., 1996. Cloning and nucleotide sequence of the major capsid proteins of *Lactobacillus* bacteriophage phi gle. Gene 175, 157–165.

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.

Kodaira, K.I., Oki, M., Kakikawa, M., Watanabe, N., Hirakawa, M., Yamada, K., Taketo, A., 1997. Genome structure of the *Lactobacillus* temperate phage phi g1e: the whole genome sequence and the putative promoter/repressor system. Gene 187, 45–53.

Lawrence, J.G., Hendrix, R.W., Casjens, S., 2001. Where are the pseudogenes in bacterial genomes? Trends Microbiol. 9, 535–540.

Lawrence, J.G., Roth, J.R., 1999. Genomic flux: genome evolution by gene loss and acquisition. in: Charlebois, R.L. (Ed.), Organization of the Prokaryotic Genome. ASM Press, Washington, DC, pp. 263–289.

Lipman, D.J., Pearson, W.R., 1985. Rapid and sensitive protein similarity searches. Science 227, 1435–1441.

Lowe, T.M., Eddy, S.R., 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. 25, 955–964.

Lucchini, S., Desiere, F., Brüssow, H., 1999. Similarly organized lysogeny modules in temperate *Siphoviridae* from low GC content Gram-positive bacteria. Virology 263, 427–435.

Mikkonen, M., Alatossava, T., 1994. Characterization of the genome region encoding structural proteins of *Lactobacillus delbrueckii* subsp. *lactis* bacteriophage LL-H. Gene 151, 53–59.

Murray, A.E., Lies, D., Li, G., Nealson, K., Zhou, J., Tiedje, J.M., 2001. DNA/DNA hybridization to microarrays reveals gene-specific differ-

ences between closely related microbial genomes. Proc. Natl. Acad. Sci. USA 98, 9853–9858.

Ohnishi, M., Kurokawa, K., Hayashi, T., 2001. Diversification of *Escherichia coli* genomes: are bacteriophages the major contributors? Trends Microbiol. 9, 481–485.

Perna, N.T., Plunkett III, G., Burland, V., Mau, B., Glasner, J.D., Rose, D.J., Mayhew, G.F., Evans, P.S., Gregor, J., Kirkpatrick, H.A., Posfai, G., Hackett, J., Klink, S., Boutin, A., Shao, Y., Miller, L., Grotbeck, E.J., Davis, N.W., Lim, A., Dimalanta, E.T., Potamousis, K.D., Apodaca, J., Anantharaman, T.S., Lin, J., Yen, G., Schwartz, D.C., Welch, R.A., Blattner, F.R., 2001. Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7. Nature 409, 529–533.

Podbielski, A., Zarges, I., Flosdorff, A., Weber-Heynemann, J., 1996. Molecular characterization of a major serotype M49 group A streptococcal DNase gene (sdaD). Infect. Immun. 64, 5349–5356.

Porwollik, S., Wong, R.M., McClelland, M., 2002. Evolutionary genomics of *Salmonella*: gene acquisitions revealed by microarray analysis. Proc. Natl. Acad. Sci. USA 99, 8956–8961.

Proux, C., van Sinderen, D., Suarez, J., Garcia, P., Ladero, V., Fitzgerald, G.F., Desiere, F., Brüssow, H., 2002. The dilemma of phage taxonomy illustrated by comparative genomics of Sfi21-like *Siphoviridae* in lactic acid bacteria. J. Bacteriol. 184, 6026–6036.

Sambrook, J.a.D.W.R., 2001. Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory, New York.

Shankar, N., Baghdayan, A.S., Gilmore, M.S., 2002. Modulation of virulence within a pathogenicity island in vancomycin-resistant *Enterococcus faecalis*. Nature 417, 746–750.

Simpson, A.J., Reinach, F.C., Arruda, P., Abreu, F.A., Acencio, M., et al., 2000. The genome sequence of the plant pathogen *Xylella fastidiosa*. Nature 406, 151–157.

Smoot, J.C., Barbian, K.D., Van Gompel, J.J., Smoot, L.M., Chaussee, M.S., Sylva, G.L., Sturdevant, D.E., Ricklefs, S.M., Porcella, S.F., Parkins, L.D., Beres, S.B., Campbell, D.S., Smith, T.M., Zhang, Q., Kapur, V., Daly, J.A., Veasy, L.G., Musser, J.M., 2002. Genome sequence and comparative microarray analysis of serotype M18 group A Streptococcus strains associated with acute rheumatic fever outbreaks. Proc. Natl. Acad. Sci. USA 99, 4668–4673.

Sriskandan, S., Unnikrishnan, M., Krausz, T., Cohen, J., 2000. Mitogenic factor (MF) is the major DNase of serotype M89 *Streptococcus pyogenes*. Microbiology 146 (Pt 11), 2785–2792.

Van Sluys, M.A., de Oliveira, M.C., Monteiro-Vitorello, C.B., Miyaki, et al. 2003. Comparative analyses of the complete genome sequences of Pierce's disease and citrus variegated chlorosis strains of *Xylella fastidiosa*. J. Bacteriol. 185, 1018–1026.

Ventura, M., Bruttin, A., Canchaya, C., Brüssow, H., 2002a. Transcription analysis of *Streptococcus thermophilus* phages in the lysogenic state. Virology 302, 21–32.

Ventura, M., Canchaya, C., Pridmore, D., Berger, B., Brüssow, H. 2003. Integration and distribution of *Lactobacillus johnsonii* prophages. J. Bacteriol. 185, 4603–4608.

Ventura, M., Foley, S., Bruttin, A., Chennoufi, S.C., Canchaya, C., Brüssow, H., 2002b. Transcription mapping as a tool in phage genomics: the case of the temperate *Streptococcus thermophilus* phage Sfi21. Virology 296, 62–76.

Wagner, P.L., Waldor, M.K., 2002. Bacteriophage control of bacterial virulence. Infect. Immunol. 70, 3985–3993.

Walker, D.C., Klaenhammer, T.R., 1994. Isolation of a novel IS3 group insertion element and construction of an integration vector for Lactobacillus spp. J. Bacteriol. 176, 5330–5340.