AEM Accepted Manuscript Posted Online 16 October 2015
Appl. Environ. Microbiol. doi:10.1128/AEM.02723-15

Genomic diversity of phages infecting probiotic strains of *Lactobacillus paracasei*

Diego J. Mercanti [a]#, Geneviève M. Rousseau [b], María L. Capra [a], Andrea Quiberoni [a], Denise M. Tremblay [b], Simon J. Labrie [b], and Sylvain Moineau [b]#

[a] Instituto de Lactología Industrial (UNL-CONICET), Facultad de Ingeniería Química, Santa Fe, Argentina;

[b] Département de Biochimie, de Microbiologie et de Bioinformatique & PROTEO, Faculté des Sciences et de Génie, Félix d'Hérelle Reference Center for Bacterial Viruses & GREB, Faculté de Médecine Dentaire, Université Laval, Québec City, Québec, Canada, G1V 0A6

Running Head: Genomic characterization of *L. paracasei* phages

#Address correspondence to Diego J. Mercanti (diegojav78@gmail.com)
and Sylvain Moineau (sylvain.moineau@bcm.ulaval.ca)

Abstract

Strains of the *Lactobacillus casei* group have been extensively studied because some are used as probiotics in foods. Conversely, their phages have received much less attention. Here we analyze the complete genome sequence of five *L. paracasei* temperate phages, C_L1, C_L2, i*Lp*84, i*Lp*1308 and iA2. Only phage iA2 could not replicate in an indicator strain. Genome length varied from 34,155 bp (iA2) to 39,474 bp (C_L1). Phages iA2 and i*Lp*1308 (34,176 bp) possess the smallest genomes reported, so far, for phages of the *L. casei* group. The GC content of the five ranged from 44.8% to 45.6%. As observed with many other phages, their genome was organized as follows: genes coding for DNA packaging – morphogenesis – lysis – lysogeny – replication. Phages C_L1, C_L2 and i*Lp*1308 are highly related to each other. Phage i*Lp*84 was also related to these three phages but the similarities were limited to gene products involved in DNA packaging and structural proteins. Genomic fragments of phages C_L1, C_L2, i*Lp*1308 and i*Lp*84 were found in several genomes of *L. casei* strains. Prophage iA2 is unrelated to these four phages but almost all of its genome was found in at least four *L. casei* strains. Overall, these phages are distinct from previously characterized *Lactobacillus* phages. Our results highlight the diversity of *L. casei* phages and indicate frequent DNA exchanges between phages and their hosts.

INTRODUCTION

The *Lactobacillus casei* group includes the species *L. casei*, *L. paracasei* and *L. rhamnosus*, which are very closely related. They are grouped within the larger *casei* group because their species boundaries have not always been clear and, historically, their nomenclature has been sometimes controversial (1-3). Nonetheless, progress has been made in their classification (2, 4, 5).

Strains within this group are used in many fermented dairy products, where they contribute to flavor development (6). They can also be introduced into foods at a specific final concentration to provide a functional characteristic (probiotic). They are also used as starter cultures and propagate during the fermentation (7). In all of these cases, lactobacilli cells are targets for phage attacks, as virulent bacterial viruses are natural contaminants of food processing factories (8). Additionally, it is well known that lysogeny is widespread in the genus *Lactobacillus* (9, 10). It was even suggested that these *Lactobacillus* prophages may evolve to become virulent phages (11). Within the *L. casei* group, high frequencies of prophage induction by mitomycin C treatment were observed in 10 out of 11 commercial strains, supporting the potential risks of introducing new phages within manufacturing facilities (10).

Genome sequencing has become necessary for the classification of prokaryote viruses and for understanding their evolution (12). Complete genome sequences of several phages infecting lactic acid bacteria (LAB) are now available in public databases. Most of these infect *Lactococcus lactis* (13), *Streptococcus thermophilus* (14) as well as *Lactobacillus* sp. Over 40 *Lactobacillus* prophage and phage genome sequences are available, including phages infecting diverse *Lactobacillus* species, such as *L. gasseri* (15-

18), *L. johnsonii* (19-21), *L. plantarum* (22-27), *L. casei/paracasei/rhamnosus/zeae* (16, 28-35), *L. helveticus* (36), *L. delbrueckii* (37-42), *L. salivarius* (16), *L. sanfranciscensis* (43), *L. fermentum* (44, 45), *L. brevis* (46), and *L. jensenii* (47).

Morphology is the historical and still useful parameter for classifying phages. Over 40 phages of the *L. casei* group have been reported to date, mostly belonging to the *Siphoviridae* family (isometric capsid, long noncontractile tail) (48). The first *casei* phages were isolated from abnormal production of the fermented milk, Yakult; phages J-1 in 1965 (49) and PL-1 two years later (50). Their genome sequences were released only recently (32). A few other genome sequences of *L. casei* group phages are available, namely A2 (28, 33, 34), phiAT3 (29), Lc-Nu (31), Lb338-1 (35) as well as prophages Lca1 (16) and Lrm1 (30). Among these fully sequenced phages, Lb338-1 is the only myophage (contractile tail) and possesses a much larger genome with a particularly low GC-content (35).

Here, we analyzed five *L. paracasei* phages, C_L1, C_L2, iA2, i*Lp*84, and i*Lp*1308. Phages C_L1 and C_L2 were previously isolated from a spontaneously lysed culture of the probiotic strain *L. paracasei* A used in commercial milk beverages (51). Phage iA2 was previously induced by mitomycin C from the same probiotic strain, *L. paracasei* A (51). Phage i*Lp*84 was isolated from a mitomycin C-induced strain, *L. paracasei* 84, and propagated on the indicator strain, *L. paracasei* INL3 (10). Similarly, phage i*Lp*1308 was isolated from mitomycin C treatment of *L. paracasei* CNRZ 1308 and replicated in *L. paracasei* A14 (10). Phage i*Lp*1308 has the same host range as phages C_L1 and C_L2, while the host range of phage i*Lp*84 is restricted to fewer strains. We sequenced the genomes of these five phages to shed light on their genetic relationships as well as to increase our understanding of phages infecting probiotic *L. paracasei* strains.

MATERIALS AND METHODS

Bacterial strains, phages, and culture conditions. *L. paracasei* strains A, A14 and INL3 were grown at 37°C in MRS broth (Difco). Identification of the strains used in this study was previously confirmed (10) by species-specific PCR (5). Prophage iA2 was induced from *L. paracasei* strain A using mitomycin C (Sigma-Aldrich) at a final concentration of 0.6 µg/ml, as described previously (51). Phage iA2 could not be propagated on any indicator strain. Phages C_L1 and C_L2, originally isolated from *L. paracasei* A, were propagated on that strain. Phages i*Lp*84 and i*Lp*1308 were propagated on the indicator strains *L. paracasei* INL3 and A14, respectively (10). Phages were then purified by three rounds of plaque purification using their respective indicator strain. All five phages were confirmed to belong to the *Siphoviridae* family by electron microscopy, as shown elsewhere (51, 52). To amplify the phages, host bacteria were grown to an optical density (600 nm) of 0.2 in MRS broth supplemented with 10 mM $CaCl_2$ (MRS-Ca) and infected with phages at approximately 10^4 PFU/ml, followed by incubation at 37°C until lysis. Phage lysates were filtered through a 0.45-µm filter (Sarstedt) and kept at 4°C until use. Phage enumeration was assessed through the double layer method (53), using MRS-Ca supplemented with 100 mM glycine (MRS-Ca-Gly) to increase plaque size (54). Plates were incubated at 37°C for 18 h. Phages and bacterial strains used in this study were stored at -80°C in MRS broth supplemented with 15% (v/v) glycerol, at both the INLAIN Collection (Santa Fe, Argentina) and the Félix d'Hérelle Reference Center for Bacterial Viruses of the Université Laval (Québec, Canada). The identification numbers at the Félix d'Hérelle Reference Center are HER510, HER511, HER512, HER513 (phages C_L1, C_L2, i*Lp*84, i*Lp*1308), and HER1510, HER1512, HER1513 (*L. paracasei* strains A, INL3, A14).

105 **DNA isolation and sequencing.** Phage genomic DNA was isolated using a Qiagen lambda
106 maxi kit (Qiagen). To determine the genome extremities, phage DNA was cleaved with
107 BglII (Roche Diagnostics) according to the manufacturer recommendations. Then, DNA
108 fragments were resolved in a 0.8% agarose gel in 40 mM Tris-acetate–1 mM EDTA (TAE)
109 buffer, stained with EZ-Vision Three (Amresco), and visualized under UV light. An aliquot
110 of each digested phage DNA was heated at 75°C for 10 min prior to gel electrophoresis.
111 The presence of an extra phage genomic fragment in the heated samples indicated the
112 presence of *cos*-type genome extremities.

113 Sequencing libraries were prepared with the Nextera XT DNA Library Preparation
114 Kit (Illumina) according to the manufacturer's instructions. The libraries were sequenced
115 using a MiSeq Reagent Kit v2 (Illumina – 500 cycles) on a MiSeq system. De novo
116 assembly was performed with Ray assembler versions 2.1.1-devel and 2.2.0-devel using a
117 *kmer* size of 31 (55). A single contig was obtained for all phages with mean coverage
118 depths of 2536, 2396, 2828, 938, and 2923 for phages C_L1, C_L2, iA2, i*Lp*84, and i*Lp*1308,
119 respectively. Genome extremities were amplified using converging primers, and the PCR
120 products were sequenced by the Sanger method using an ABI 3730xl apparatus at the
121 sequencing and genotyping platform of the Centre Hospitalier de l'Université Laval. These
122 latter phage sequences were assembled with Staden software (version 2.0.0b9) (56).
123
124 **Bioinformatics analysis.** Complete phage genomes were analyzed with BioEdit (57). Open
125 reading frames (ORFs) were predicted with the command line version of GeneMarkS
126 (version 4.29) using the setting for phage genomes (58). The identified ORFs were
127 confirmed with ORF Finder (http://www.ncbi.nlm.nih.gov/gorf/gorf.html). ORFs were
128 considered candidates for evaluation when they encoded 25 or more amino acids (aa), and

129 possessed both a conserved Shine-Dalgarno sequence (5'-AAGGAGGT-3') and a start
130 codon (AUG, UUG or GUG). BLASTp was used to compare translated ORF products with
131 known proteins. Hits were considered valid when the E-value was lower than $1x10^{-3}$.
132 Conserved domains in protein sequences were identified with the NCBI CD-search
133 interface to search the Conserved Domain Database
134 (http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi). Physicochemical parameters of
135 predicted proteins were calculated with the ProtParam Tool from the ExPASy proteomics
136 server (http://web.expasy.org/protparam/). The bioinformatic tools tRNAscan-SE 1.21
137 (http://lowelab.ucsc.edu/tRNAscan-SE/) and ARAGORN (59) were used for tRNA gene
138 detection. Phage codon usage was calculated using the Countcodon program
139 (http://www.kazusa.or.jp/codon/) and compared with the codon usage of *L. paracasei*,
140 obtained from the same database. For the alignment figures, phage and prophage protein
141 sequences were compared using BLASTp 2.2.28+ (60). The percent identity between
142 proteins was calculated by dividing the number of identical residues by the size of the
143 largest protein. The genome maps were generated using the GenoPlotR package (61).
144
145 **Terminase phylogeny.** The phylogenetic tree of the terminase was generated with a dataset
146 including sequences representing the main phage groups (62). The sequences were aligned
147 using MAFFT with the automatic settings. The alignment was converted to the PHYLIP
148 format with compatible name using in-house Python script. The most probable amino-acid
149 substitution model was determined using ProtTest 3.2 (63). The best model was then
150 implemented in Phyml 3.0 to calculate the best tree (64). Branch support value was
151 established using the Shimodaira-Hasegawa-like procedure (65). The leaves of the tree

were renamed using the Newick utility package (66). Finally, the tree was rendered using the web interface ITOL (67).

Nucleotide sequence accession numbers. The complete genome sequences of phages C_L1, C_L2, iA2, i*Lp*84, and i*Lp*1308 have been deposited in GenBank under accession numbers KR905066, KR905067, KR905068, KR905069, and KR905070, respectively.

RESULTS AND DISCUSSION

Phage genomes overview. The five phage genomes analyzed in this study were double-stranded DNA molecules and their general features are shown in Table 1. The genome size of phages C_L1, C_L2, and i*Lp*84 ranged from 38,751 to 39,474 bp and were similar to most *L. casei* group phages already described (Table 1). The genomes of phages iA2 (34,155 bp) and i*Lp*1308 (34,176 bp) were about 2 kb shorter than the genome of phage Lc-Nu (36,466 bp), the smallest genome of *L. casei* phage reported before this study (Table 1). The GC-content (44.8-45.6%) of the five phages was similar to the other *L. casei* group phages (Table 1). The previously characterized *L. paracasei* myophage ϕLb338-1 has a much lower GC content (37.4%) but it differs in structure from the *L. casei* phages and has a much larger genome (35). The genome of prophage iA2 has cohesive extremities with 3' overhangs that are 10-nt long (CGGCATGCAA). Of note, when this sequence is ligated, it generates an *orf* coding for a HNH endonuclease. Hence, we elected to end the map of prophage iA2 after this *orf* and we located the *cos*-site at the positions 34135-34144.

174 **Genome organization.** Table 2 provides the analysis of the predicted ORFs for phage C_L1.
175 Homologous ORFs predicted for phages C_L2, i*Lp*84, and i*Lp*1308 and their respective
176 amino acid identities with the deduced proteins of phage C_L1 are indicated as well. Phage
177 C_L1 was chosen as the reference because it has the highest identity with the other three
178 phages (C_L2, i*Lp*84, and i*Lp*1308; see Fig. 1, discussed below). Because prophage iA2 does
179 not have significant similarity to these four phages, Table 3 presents the analysis of its 50
180 predicted ORFs, including 24 with a probable function. The organization of these genomes
181 is modular and resembles that reported for other *L. casei* phages. Their genomes have
182 clusters of genes involved in DNA packaging, morphogenesis, lysis, lysogeny, and
183 replication (Figure 1; Tables 2 and 3). Still, as illustrated in Figure 1, the five phages
184 described in this study are clearly distinct from the other phages of the *L. casei* group
185 characterized to date.
186
187 **Phylogeny and comparative analysis of the phages.** We looked at the relationships of
188 these five phages to other phages using one of the most conserved genes: the large
189 terminase subunit. This gene was used previously in attempts to classify phages (70).
190 Casjens *et al.* (62) also proposed to use the terminase phylogeny for classification of phages
191 into different groups according to their encapsidation mechanism. Figure 2 presents the
192 phylogenic analysis of a dataset of terminase sequences including the five phages of this
193 study and others present in public databases. Based on terminase phylogeny, phages C_L1,
194 C_L2, i*Lp*84, and i*Lp*1308 are highly related, and possess a P22-like headful mechanism.
195 Among *Lactobacillus* phages, only phages Lj965 (infecting *L. johnsonii*) and LL-H (*L.*
196 *delbrueckii*) have this feature, and only Lj965 is related to the five *L. paracasei* phages. The

other *L. casei* group phages sequenced, so far, have a 3'-extended *cos* packaging system, similar to phage iA2. These findings highlight the diversity in this group of phages.

Among the other phages of the *L. casei* group presenting 3'-extended *cos* ends, the large terminase subunits of phages A2, J-1, PL-1, and prophage Lrm1 are closely related. As indicated previously, phages J-1 and PL-1 were isolated during the manufacture of Yakult in the mid-1960s and they are almost identical (49, 50). However, phage PL-1 infects a strain insensitive to J-1 (49, 50). Prophage iA2 shares a low level of similarity to these phages but it falls within the same cluster based on terminase phylogeny (Figure 2).

Genomic comparison of phages C_L1 and C_L2. As shown in Figure 1, phages C_L1 and C_L2 are highly related (>80% identity) and they were both amplified on the same strain, *L. paracasei* A. The differences were mainly found in genes coding for non-structural proteins, probably involved in replication, regulation, and lysogeny. Both phages have homing endonucleases of similar size (ORF61/ C_L1 and ORF3/ C_L2), but with low genetic identity. The presence of at least three genes encoding putative transposases in C_L1 (ORF33, ORF35, and ORF36) and one in C_L2 (ORF37) (Table 2) was somewhat surprising and may be responsible for phage diversity since transposases can promote genome rearrangements and genetic exchange. In C_L1, however, only ORF35 and ORF36 seem to be functional, based on their protein size. The former (ORF35) has 381 amino acid residues and transposases of similar length (and with >71% identity) as reported for diverse species of *Lactobacillus*, including *L. paracasei* phage Lb338-1 but also strains of *L. brevis*, *L. kisonensis*, *L. pentosus*, *L. malefermentans,* and *Sporolactobacillus laevolacticus*. The ORF36 transposase has 141 amino acid residues and also seems to be complete, though the identity to enzymes of similar size was lower and corresponded mainly to transposases of *L. versmoldensis*, *L. salivarius* and several species of *Staphylococcus*. Phage C_L2's ORF37,

which has a very high amino-acid identity to ORF33 of C_L1, has only 50 amino acid residues and this gene product is likely not functional. Based on BLAST analysis, similarities were found only to partial sequences of reported transposases of *Staphylococcus aureus* and *S. epidermidis*.

Genomic comparison of the phages C_L1 and i*Lp*1308. The genome of phage i*Lp*1308 is about 5 kb smaller than C_L1 and C_L2, but retains a high level of identity with them, especially in genes involved in DNA packaging, morphogenesis and cell lysis. Inverted repetitions (285 bases) flank *orf29* and *orf34* (4.3 kb apart) in the genome of i*Lp*1308 (Figure 3), but they are absent in the other four phages. The region between these repetitions in the genome of phage i*Lp*1308 contains six genes encoded on the opposite strand, probably as a result of a recombination event. Similarly, phage C_L1 has a 4.9 kb region (from *orf32* to *orf40*) that is missing between *orf32* and *orf33* in phage i*Lp*1308 (Figure 3). This 4.9 kb region contains genes encoding for antirepressor protein, tcdA-E operon negative regulator and transposases, as discussed. These DNA rearrangements may have occurred during past integration/excision events from bacterial genomes.

Analysis of phage i*Lp*84 genome. Phage i*Lp*84 has a genome size similar to C_L1 and C_L2 but the identity is limited to genes/proteins involved in DNA packaging and morphogenesis. The tail tape measure, tail fiber, and tail-host interaction proteins are the largest structural proteins among these phages. Some of them share high amino acid identities (>80%). The tail fiber and tail-host interaction protein of phage i*Lp*84 possess 72% to 78% amino acid identity with the corresponding proteins in phages C_L1, C_L2, and i*Lp*1308. In support of the similarities between the tail-host interaction proteins, these four

Applied and Environmental Microbiology

phages have the same host range (Mercanti et al 2011). Conversely, the tail tape measure protein of phage i*Lp*84 does not have significant identity with the other phages.

Analysis of phage iA2 genome. The genome of phage iA2 is completely different from the other phages infecting the *L. casei* group, suggesting a different origin. The endolysin (ORF22) is the only protein that has homology with proteins of phages C_L1, C_L2 and i*Lp*1308, whereas ORF38 (putative DNA repair protein) has homology with a protein of phage i*Lp*84. These genomic differences are in accordance with other distinct phenotypes. For example, phage iA2 is readily induced at high level after treatment with mitomycin C but we could not find an indicator strain for this phage. As discussed below, phage iA2 is highly similar to putative prophages extracted from the genomic sequences of strains of the *L. casei* group available in GenBank (Figure 4).

Comparison with prophages. The occurrence of prophage DNA within bacterial genomes is common and the presence of such sequences can be identified using bioinformatic tools such as PHAST. Bacterial genomes can harbor inducible prophages but also altered/remnants of prophages displaying insertions, deletions or rearrangements (16). The BLAST analyses of the five phage genomes characterized in this study revealed that they are more closely related to prophage sequences found in bacterial chromosomes (data not shown). These prophage sequences were found in the genome of *L. casei* strains but also in strains of the other two species of the *casei* group, *L. paracasei* and *L. rhamnosus*. This is likely due to the high degree of relatedness among species included in the *casei* group (5). Some *Lactobacillus* phages were previously assumed to be able to cross the species barrier.

For example, *L. rhamnosus* phage Lc-Nu is homologous to the *L. casei* temperate phages phiAT3 and A2 (31).

Phages C_L1 and C_L2 showed high identity with sequences found in *L. rhamnosus* strains ATCC 53103, GG, and LOCK900, and *L. casei* strains 12A and ATCC 334 (data not shown). However, alignments were always fragmented and covered, in the best cases, just 48 to 50% of phage C_L1 and C_L2 genomes. Homology between the genomes of C_L1/C_L2 and the genome of *L. casei* ATCC 334 was detected in two different regions, one being the prophage Lca1 (16). The genome of phage i*Lp*1308 aligned relatively well with sequences in the genomes of three *L. casei* strains: W56, BD-II and BL-23 (data not shown). Sequences identical to the phage i*Lp*84 genome were found in the genomes of the same *L. casei* strains (data not shown).

Interestingly, the genome of the prophage iA2 was also observed in the above *L. casei* strains (W56, BD-II, and BL-23) and also in *L. casei* LC2W (Figure 4). Unlike the other four phages studied, the identity to prophage iA2 sequences found in these four bacterial genomes was very high (from 99.90 to 99.98%) and included virtually the whole prophage genome, though with some rearrangements. Three other fully sequenced *Lactobacillus* strains contained very large fragments of iA2 prophage (*L. paracasei* N1115 and 8700:2, *L. casei* LOCK919) (Figure 4).

We extracted these prophage sequences from the genomes of strains of the *L. casei* group for a comparative analysis with prophage iA2. The comparison is presented in Figure 4. As expected from the nucleotide identity levels, most of the deduced ORFs of prophages from *L. casei* strains W56, BD II, and BL-23 (group 1) share high amino acid identity (>99%) with the predicted proteins of prophage iA2. A second prophage, found in each of these four bacterial strains, was found to be very well conserved (prophage group 2), but

Applied and Environmental Microbiology

with noticeably lower identity to group 1. About half of the deduced ORFs of *L. casei* LOCK919 and *L. paracasei* N1115 prophages, and only 4 ORFs of *L. paracasei* 87002 prophage possess amino acid identity over 99% (Figure 4). Nonetheless, for most ORFs of these three phages, identity with prophages of group 1 is still high, indicating that they could belong to the same group (data not shown). All prophages of group 1 and prophages of *L. casei* LOCK919 and *L. paracasei* N1115 are integrated at the same site, in a region coding for a tRNA-Leu. Prophage of group 2 are integrated in a region coding for a tRNA-Arg.

Taken altogether, these comparisons highlighted the widespread occurrence of phage-related sequences in strains of the *casei* group. These prophage sequences significantly contribute to their diversity. It is tempting to speculate that perhaps prophage sequences could be used as markers for either strain identification or tracking.

Presence of tRNA in phage genomes. Although the importance of tRNAs in phage genomes is not fully understood, their presence is relatively common, particularly in large genomes (69). One tRNA was detected in the genomes of phages C_L2 and i*Lp*84. No tRNA was detected in the smaller genomes of phages i*Lp*1308 and iA2 (Table 4). Table 4 lists the tRNA present in the available phage genomes of the *L. casei* group. Phage Lb338-1 is the only one with more than one tRNA in its genome. Phage i*Lp*84 tRNA provides the amino acid isoleucine (Ile, ATA). This tRNA is similar to the one found in phage A2, but the A2 tRNA provides the amino acid leucine (TTA). The isoleucine codon ATA is used by *L. paracasei* at frequencies of 3.7%, while for phage i*Lp*84, it is 13.2%. This observation supports the hypothesis that phages encode tRNAs corresponding to codons less frequently used by the host bacteria to favor the expression of phage proteins (61).

316

317 **Conclusion.** *Lactobacillus* phages have been studied less than other phages of lactic acid
318 bacteria (48). The interest in this genus has been eclipsed by substantial studies on
319 *Lactococcus lactis* and *Streptococcus thermophilus*, the main species of LAB used as
320 starters in the dairy industry (71). However, strains of the *L. casei* group (*L. casei*, *L.*
321 *paracasei* and *L. rhamnosus*) are now widely used as probiotics in foods. It is recognized
322 that increased used of LAB will eventually lead to infection by virulent phages (71).
323 Therefore, *Lactobacillus* phages deserve attention, considering the potential risk of losses
324 associated with phage infections of probiotic bacteria, whose replacement with suitable
325 strains is difficult to achieve.

326 In previous studies we demonstrated the high occurrence of prophages in probiotic
327 strains added to dairy foods (10, 51). Here, we characterize five new phages of the
328 increasingly important *L. casei* group, including four which can replicate in indicator
329 strains. These phages are distinct from the previously characterized *Lactobacillus* phages
330 and can be divided into two groups (C_L1/C_L2/i*Lp*1308/i*Lp*84 and iA2). Our results also
331 point out what appears to be a high frequency of recombination events between phages and
332 their host prophages, leading to phage and host diversity. These findings also suggest that
333 new *Lactobacillus* phages likely remain to be discovered.

334 Our comparative genomic analyses also suggest ample distribution of prophages in
335 the genomes of strains across the *L. casei* group. This offers a reservoir of genes that could
336 be used by virulent phages to rapidly evolve (72, 73). The impact of these phage-host
337 interactions on the properties (including probiotics), phenotypes, and stability of
338 *Lactobacillus* strains is currently unknown. Future research should look to discover the
339 frequency of these interactions and their impact on strain activities. It would also be of

interest to study the different molecular mechanisms used by *Lactobacillus* to defend against these phages (74).

ACKNOWLEDGMENTS

We would like to thank Barbara-Ann Conway (Medical Writer & Editor) for editorial assistance. We are grateful to Pier-Luc Plante and Jacques Corbeil for preliminary genome assembly and access to the Illumina platform. This work was supported by the Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET; Project PIP N°112-200801-01206; Argentina). D.M. acknowledges funding from the CONICET and by the Agencia Nacional de Promoción Científica y Tecnológica (ANPCyT; Project PICT 2010-0138; Argentina). S.M. holds a Tier 1 Canada Research Chair in Bacteriophages.

REFERENCES

1. **Tindall BJ.** 2008. The type strain of *Lactobacillus casei* is ATCC 393, ATCC 334 cannot serve as the type because it represents a different taxon, the name *Lactobacillus paracasei* and its subspecies names are not rejected and the revival of the name *'Lactobacillus zeae'* contravenes Rules 51b (1) and (2) of the International Code of Nomenclature of Bacteria. Opinion 82. Int J Syst Evol Microbiol **58:**1764-1765.
2. **Dicks LM, Du Plessis EM, Dellaglio F, Lauer E.** 1996. Reclassification of *Lactobacillus casei* subsp. *casei* ATCC 393 and *Lactobacillus rhamnosus* ATCC 15820 as *Lactobacillus zeae* nom. rev., designation of ATCC 334 as the neotype of *L. casei* subsp. *casei*, and rejection of the name *Lactobacillus paracasei*. Int J Syst Bacteriol **46:**337-340.
3. **Dellaglio F, Felis GE, Torriani S.** 2002. The status of the species *Lactobacillus casei* (Orla-Jensen 1916) Hansen and Lessel 1971 and *Lactobacillus paracasei* Collins et al. 1989. Request for an opinion. Int J Syst Evol Microbiol **52:**285-287.
4. **Collins MD, Phillips BA, Zanoni P.** 1989. Deoxyribonucleic Acid Homology Studies of *Lactobacillus casei*, *Lactobacillus paracasei* sp. nov., subsp. *paracasei* and subsp. *tolerans*, and *Lactobacillus rhamnosus* sp. nov., comb. nov. Int J Syst Evol Microbiol **39:**105-108.
5. **Desai AR, Shah NP, Powell IB.** 2006. Discrimination of dairy industry isolates of the *Lactobacillus casei* group. J Dairy Sci **89:**3345-3351.

6. **Liu M, Nauta A, Francke C, Siezen RJ.** 2008. Comparative genomics of enzymes in flavor-forming pathways from amino acids in lactic acid bacteria. Appl Environ Microbiol **74:**4590-4600.

7. **Saarela M, Mogensen G, Fondén R, Mättö J, Mattila-Sandholm T.** 2000. Probiotic bacteria: safety, functional and technological properties. J Biotechnol **84:**197-215.

8. **Capra ML, Binetti AG, Mercanti DJ, Quiberoni A, Reinheimer JA.** 2009. Diversity among *Lactobacillus paracasei* phages isolated from a probiotic dairy product plant. J Appl Microbiol **107:**1350-1357.

9. **Suarez V, Zago M, Quiberoni A, Carminati D, Giraffa G, Reinheimer J.** 2008. Lysogeny in *Lactobacillus delbrueckii* strains and characterization of two new temperate prolate-headed bacteriophages. J Appl Microbiol **105:**1402-1411.

10. **Mercanti DJ, Carminati D, Reinheimer JA, Quiberoni A.** 2011. Widely distributed lysogeny in probiotic lactobacilli represents a potentially high risk for the fermentative dairy industry. Int J Food Microbiol **144:**503-510.

11. **Brussow H.** 2001. Phages of dairy bacteria. Annu Rev Microbiol **55:**283-303.

12. **Ackermann HW, Kropinski AM.** 2007. Curated list of prokaryote viruses with fully sequenced genomes. Res Microbiol **158:**555-566.

13. **Mahony J, van Sinderen D.** 2014. Current taxonomy of phages infecting lactic acid bacteria. Front Microbiol **5:**7.

14. **Quiberoni A, Moineau S, Rousseau G, Reinheimer J, Ackermann H-W.** 2010. *Streptococcus thermophilus* bacteriophages. Int Dairy J **20:**657-664.

15. **Altermann E, Klaenhammer TR.** 2011. Group-specific comparison of four lactobacilli isolated from human sources using differential blast analysis. Genes & nutrition **6:**319-340.

16. **Ventura M, Canchaya C, Bernini V, Altermann E, Barrangou R, McGrath S, Claesson MJ, Li Y, Leahy S, Walker CD, Zink R, Neviani E, Steele J, Broadbent J, Klaenhammer TR, Fitzgerald GF, O'Toole P W, van Sinderen D.** 2006. Comparative genomics and transcriptional analysis of prophages identified in the genomes of *Lactobacillus gasseri*, *Lactobacillus salivarius*, and *Lactobacillus casei*. Appl Environ Microbiol **72:**3130-3146.

17. **Altermann E, Klein JR, Henrich B.** 1999. Primary structure and features of the genome of the *Lactobacillus gasseri* temperate bacteriophage (phi)adh. Gene **236:**333-346.

18. **Baugher JL, Durmaz E, Klaenhammer TR.** 2014. Spontaneously induced prophages in *Lactobacillus gasseri* contribute to horizontal gene transfer. Appl Environ Microbiol **80:**3508-3517.

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

20. **Ventura M, Canchaya C, Pridmore RD, Brussow H.** 2004. The prophages of *Lactobacillus johnsonii* NCC 533: comparative genomics and transcription analysis. Virology **320:**229-242.

21. **Desiere F, Pridmore RD, Brussow H.** 2000. Comparative genomics of the late gene cluster from *Lactobacillus phages*. Virology **275:**294-305.

22. **Kodaira KI, 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.

23. **Ventura M, Canchaya C, Kleerebezem M, de Vos WM, Siezen RJ, Brussow H.** 2003. The prophage sequences of *Lactobacillus plantarum* strain WCFS1. Virology **316:**245-255.

24. **Briggiler Marco M, Garneau JE, Tremblay D, Quiberoni A, Moineau S.** 2012. Characterization of two virulent phages of *Lactobacillus plantarum*. Appl Environ Microbiol **78:**8719-8734.

25. **Lu Z, Altermann E, Breidt F, Predki P, Fleming HP, Klaenhammer TR.** 2005. Sequence analysis of the *Lactobacillus plantarum* bacteriophage PhiJL-1. Gene **348:**45-54.

26. **Chibani-Chennoufi S, Dillmann ML, Marvin-Guy L, Rami-Shojaei S, Brussow H.** 2004. *Lactobacillus plantarum* bacteriophage LP65: a new member of the SPO1-like genus of the family Myoviridae. J Bacteriol **186:**7069-7083.

27. **Yoon BH, Jang SH, Chang HI.** 2011. Sequence analysis of the *Lactobacillus* temperate phage Sha1. Arch Virol **156:**1681-1684.

28. **Garcia P, Ladero V, Suarez JE.** 2003. Analysis of the morphogenetic cluster and genome of the temperate *Lactobacillus casei* bacteriophage A2. Arch Virol **148:**1051-1070.

29. **Lo TC, Shih TC, Lin CF, Chen HW, Lin TH.** 2005. Complete genomic sequence of the temperate bacteriophage PhiAT3 isolated from *Lactobacillus casei* ATCC 393. Virology **339:**42-55.

30. **Durmaz E, Miller M, Azcarate-Peril M, Toon S, Klaenhammer T.** 2008. Genome sequence and characteristics of Lrm1, a prophage from industrial *Lactobacillus rhamnosus* strain M1. Appl Environ Microbiol **74:**4601-4609.

31. **Tuohimaa A, Riipinen KA, Brandt K, Alatossava T.** 2006. The genome of the virulent phage Lc-Nu of probiotic *Lactobacillus rhamnosus*, and comparative genomics with *Lactobacillus casei* phages. Arch Virol **151:**947-965.

32. **Dieterle ME, Jacobs-Sera D, Russell D, Hatfull G, Piuri M.** 2014. Complete genome sequences of *Lactobacillus* phages J-1 and PL-1. Genome Announc **2(1):**e00998-13.

33. **Herrero M, de los Reyes-Gavilán CG, Caso JL, Suárez JE.** 1994. Characterization of ø393-A2, a bacteriophage that infects *Lactobacillus casei*. Microbiology **140:**2585-2590.

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

35. **Alemayehu D, Ross RP, O'Sullivan O, Coffey A, Stanton C, Fitzgerald GF, McAuliffe O.** 2009. Genome of a virulent bacteriophage Lb338-1 that lyses the probiotic *Lactobacillus paracasei* cheese strain. Gene **448:**29-39.

36. **Zago M, Scaltriti E, Rossetti L, Guffanti A, Armiento A, Fornasari ME, Grolli S, Carminati D, Brini E, Pavan P, Felsani A, D'Urzo A, Moles A, Claude JB, Grandori R, Ramoni R, Giraffa G.** 2013. Characterization of the genome of the dairy *Lactobacillus helveticus* bacteriophage {Phi}AQ113. Appl Environ Microbiol **79:**4712-4718.

37. **Vasala A, Dupont L, Baumann M, Ritzenthaler P, Alatossava T.** 1993. Molecular comparison of the structural proteins encoding gene clusters of two related *Lactobacillus delbrueckii* bacteriophages. J Virol **67:**3061-3068.

38. **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.

39. **Mikkonen M, Raisanen L, Alatossava T.** 1996. The early gene region completes the nucleotide sequence of *Lactobacillus delbrueckii* subsp. *lactis* phage LL-H. Gene **175:**49-57.

40. **Riipinen KA, Forsman P, Alatossava T.** 2011. The genomes and comparative genomics of *Lactobacillus delbrueckii* phages. Arch Virol **156:**1217-1233.

41. **Casey E, Mahony J, O'Connell-Motherway M, Bottacini F, Cornelissen A, Neve H, Heller KJ, Noben JP, Dal Bello F, van Sinderen D.** 2014. Molecular characterization of three *Lactobacillus delbrueckii* subsp. *bulgaricus* phages. Appl Environ Microbiol **80:**5623-5635.

42. **Casey E, Mahony J, Neve H, Noben JP, Bello FD, van Sinderen D.** 2014. Genome and proteome analysis of bacteriophage Ldl1 reveals the existence of a novel phage group infecting *Lactobacillus delbrueckii* subsp. *lactis*. Appl Environ Microbiol **81:**1319-1326.

43. **Ehrmann MA, Angelov A, Picozzi C, Foschino R, Vogel RF.** 2013. The genome of the *Lactobacillus sanfranciscensis* temperate phage EV3. BMC Res Notes **6:**514.

44. **Yoon BH, Chang HI.** 2011. Complete genomic sequence of the *Lactobacillus* temperate phage LF1. Arch Virol **156:**1909-1912.

45. **Zhang X, Wang S, Guo T, Kong J.** 2011. Genome analysis of *Lactobacillus fermentum* temperate bacteriophage ϕPYB5. Int J Food Microbiol **144:**400-405.

46. **Jang SH, Yoon BH, Chang HI.** 2011. Complete nucleotide sequence of the temperate bacteriophage LBR48, a new member of the family Myoviridae. Arch Virol **156:**319-322.

47. **Martin R, Escobedo S, Suarez JE.** 2010. Induction, structural characterization, and genome sequence of Lv1, a prophage from a human vaginal *Lactobacillus jensenii* strain. Int Microbiol **13:**113-121.

48. **Villion M, Moineau S.** 2009. Bacteriophages of *Lactobacillus*. Front Biosci **14:**1661-1683.

49. **Hino M, Ikebe N.** 1965. Studies on the lactic acid bacteria employed for beverage production II. Isolation and some properties of a bacteriophage infected during the fermentation of lactic acid beverage. J Agric Chem Soc Jpn **39:**472-476.

50. **Watanabe K, Takesue S, Jin-Nai K, Yoshikawa T.** 1970. Bacteriophage active against the lactic acid beverage-producing bacterium *Lactobacillus casei*. Appl Microbiol **20:**409-415.

51. **Capra ML, Mercanti DJ, Reinheimer JA, Quiberoni AL.** 2010. Characterisation of three temperate phages released from the same *Lactobacillus paracasei* commercial strain. Int J Dairy Technol **63:**396-405.

52. **Mercanti DJ, Ackermann HW, Quiberoni A.** 2015. Characterization of two temperate *Lactobacillus paracasei* bacteriophages: morphology, kinetics and adsorption. Intervirology **58:**49-56.

53. **Svensson U, Christiansson A.** 1991. Methods for phage monitoring. Bulletin - FIL-IDF **263:**29-39.

512 54. **Lillehaug D.** 1997. An improved plaque assay for poor plaque-producing temperate
513 lactococcal bacteriophages. J Appl Microbiol **83:**85-90.
514 55. **Boisvert S, Laviolette F, Corbeil J.** 2010. Ray: simultaneous assembly of reads
515 from a mix of high-throughput sequencing technologies. J Comput Biol **17:**1519-
516 1533.
517 56. **Staden R.** 1996. The Staden sequence analysis package. Mol Biotechnol **5:**233-
518 241.
519 57. **Hall TA.** 1999. BIOEDIT: a user-friendly biological sequence alignment editor and
520 analysis program for windows 95/98/NT. Nucleic Acids Symp Ser **41:**95-98.
521 58. **Lukashin AV, Borodovsky M.** 1998. GeneMark.hmm: new solutions for gene
522 finding. Nucleic Acids Res **26:**1107-1115.
523 59. **Laslett D, Canback B.** 2004. ARAGORN, a program to detect tRNA genes and
524 tmRNA genes in nucleotide sequences. Nucleic Acids Res **32:**11-16.
525 60. **Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K,**
526 **Madden TL.** 2009. BLAST+: architecture and applications. BMC Bioinformatics
527 **10:**1471-2105.
528 61. **Guy L, Roat Kultima J, Andersson SGE.** 2010. genoPlotR: comparative gene and
529 genome visualization in R. Bioinformatics **26:**2334-2335.
530 62. **Casjens SR, Gilcrease EB, Winn-Stapley DA, Schicklmaier P, Schmieger H,**
531 **Pedulla ML, Ford ME, Houtz JM, Hatfull GF, Hendrix RW.** 2005. The
532 generalized transducing *Salmonella* bacteriophage ES18: Complete genome
533 sequence and DNA packaging strategy. J Bacteriol **187:**1091-1104.
534 63. **Darriba D, Taboada GL, Doallo R, Posada D.** 2011. ProtTest 3: fast selection of
535 best-fit models of protein evolution. Bioinformatics **27:**1164-1165.

64. **Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O.** 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol **59:**307-321.

65. **Shimodaira H.** 2002. An approximately unbiased test of phylogenetic tree selection. Syst Biol **51:**492-508.

66. **Junier T, Zdobnov EM.** 2010. The Newick utilities: high-throughput phylogenetic tree processing in the UNIX shell. Bioinformatics **26:**1669-1670.

67. **Letunic I, Bork P.** 2011. Interactive Tree Of Life v2: online annotation and display of phylogenetic trees made easy. Nucleic Acids Res **39:**W475-478.

68. **Sambrook JJ, Russell DDW.** 2001. Molecular cloning: a laboratory manual, 3rd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

69. **Samson JE, Moineau S.** 2010. Characterization of *Lactococcus lactis* phage 949 and comparison with other lactococcal phages. Appl Environ Microbiol **76:**6843-6852.

70. **Lopes A, Tavares P, Petit MA, Guerois R, Zinn-Justin S.** 2014. Automated classification of tailed bacteriophages according to their neck organization. BMC genomics **15:**1027.

71. **Garneau J, Moineau S.** 2011. Bacteriophages of lactic acid bacteria and their impact on milk fermentations. Microbial Cell Factories **10:**S20.

72. **Labrie S, Moineau S.** 2007. Abortive infection mechanisms and prophage sequences significantly influence the genetic makeup of emerging lytic lactococcal phages. J Bacteriol **189:**1482-1487.

558 73. **Capra ML, Quiberoni A, Ackermann HW, Moineau S, Reinheimer JA.** 2006.
559 Characterization of a new virulent phage (MLC-A) of *Lactobacillus paracasei*. J
560 Dairy Sci **89:**2414-2423.
561 74. **Labrie S, Samson J, Moineau S.** 2010. Bacteriophage resistance mechanisms. Nat
562 Rev Microbiol **8:**317-327.

563

564

565 **Table 1.** General features of phage genomes of the *L. casei* group

Phage	Genome size (kb)	% CG	Total ORFs (+ strand / - strand)	Reference
C_L1	39,474	44.8	62 (48 / 14)	This study
C_L2	38,751	44.9	60 (59 / 1)	This study
iA2	34,155	45.6	50 (43 / 7)	This study
i*Lp*84	39,399	45.1	60 (49 / 11)	This study
i*Lp*1308	34,176	45.1	50 (42 / 8)	This study
A2	43,411	44.9	61 (55 / 6)	(34)
J-1	40,931	44.8	63 (57 / 6)	(32)
Lca1	46,986	44.8	71 (56 / 15)	(16)
Lc-Nu	36,466	44.3	51 (50 / 1)	(31)
Lrm1	39,989	45.5	54 (51 / 3)	(30)
phiAT3	39,166	44.6	53 (50 / 3)	(29)
PL-1	38,880	44.9	59 (57 / 2)	(32)
Lb338 1	142,111	37.4	199 (128 / 71)	(35)

566

567 **Table 2.** Features of ORFs of phage C_L1 and homology with ORFs of phages C_L2, i*Lp*84 and i*Lp*1308

Strand	ORF	Start	Stop	Size (aa)	MM (kDa)	pI	SD sequence AAGGAGGT	Putative function	Representative ORF or gene (% aa identity) in indicated phage		
									C_L2	i*Lp*1308	i*Lp*84
+	1	31	522	163	18.2	5.7	CGGGAGGTgtggtgatat**GTG**	small subunit terminase	ORF1 (100)	ORF1 (25.6)	ORF1 (25.8)
+	2	506	1759	417	48.3	7.7	AAGGATGGtgatgtca**ATG**	terminase large subunit	ORF2 (80.7)	ORF2 (99.8)	ORF2 (98.3)
+	3	1719	3191	490	54.3	4.5	ACGGAGAAA**TG**	portal protein	ORF4 (99.8)	ORF3 (100)	ORF3 (97.6)
+	4	3157	4155	332	37.6	6.9	ACGGTGGTggcaatg**ATG**	Mu protein F like protein	ORF5 (100)	ORF4 (96.4)	ORF4 (92.5)
+	5	4165	4491	108	12.5	4.8	TAGGAGGAaacg**ATG**		ORF6 (100)		
+	6	4662	5300	212	23.8	5.0	AAGGAGTTcttagc**ATG**		ORF7 (100)	ORF5 (99.5)	ORF5 (97.6)
+	7	5313	5627	104	10.8	5.2	AAGGAGGTacttttt**ATG**		ORF8 (99.0)	ORF6 (99.0)	ORF6 (95.2)
+	8	5641	6660	339	37.6	4.9	AAGGAGGAttaacttA**TG**		ORF9 (100)	ORF7 (100)	ORF7 (95.7)
+	9	6747	6926	59	5.4	3.7	TATCCGCTgaccct**GTG**		ORF10 (100)	ORF8 (100)	
+	10	6996	7370	124	14.2	4.8	AAGGAGGCatgaa**ATG**		ORF11 (100)	ORF9 (100)	ORF10 (87.9)
+	11	7375	7677	100	11.4	4.8	AAAGAGGTgatcat**ATG**		ORF12 (100)	ORF10 (100)	ORF11 (90.0)
+	12	7674	8039	121	13.5	10.5	TTGTGGGTgagacgaa**ATG**		ORF13 (100)	ORF11 (100)	ORF12 (86.0)
+	13	8040	8444	134	15.4	4.7	AAGCAGCAaagtag**GTG**	major structural protein	ORF14 (99.3)	ORF12 (99.3)	ORF13 (97.0)
+	14	8456	9061	201	21.7	4.4	TAGGAGGCataac**ATG**	major tail protein	ORF15 (100)	ORF13 (100)	ORF14 (84.6)
+	15	9147	9479	110	11.9	5.7	AAGGATTTtaaatc**ATG**		ORF16 (100)	ORF14 (100)	ORF15 (75.7)
+	16	9584	9937	117	13.7	9.3	CGAGACATgaacgt**ATG**		ORF17 (100)	ORF15 (100)	ORF16 (88.9)
+	17	9930	13019	1029	108.5	9.0	AAGGAGGGagcat**ATG**	tape measure	ORF18 (99.8)	ORF16 (99.9)	ORF17 (15.4)
+	18	13022	15133	703	77.9	4.5	AAGGGCTattttagtt**TTG**	tail tip protein	ORF19 (99.1)	ORF17 (99.7)	ORF18 (73.1)
+	19	15130	17595	821	89.6	4.9	CAGGAGGCatggct**ATG**	host interaction protein	ORF20 (99.4)	ORF18 (99.3)	ORF19 (47.0)
+	20	17605	17928	107	11.7	4.2	TAGGAGGTtgtt**ATG**		ORF21 (99.1)	ORF19 (100)	ORF20 (95.3)
+	21	17921	18052	43	4.8	4.1	AAGGCGGCgaaatcaa**GTG**		ORF22 (100)	ORF20 (100)	
+	22	18083	18469	128	14.4	6.0	AAGGAAGTgatgaca**ATG**	holin	ORF23 (87.7)	ORF21 (78.9)	
+	23	18450	18659	69	8.0	9.8	AAGGAGTGacagcc**GTG**		ORF24 (92.8)	ORF22 (81.2)	
+	24	18656	19651	331	34.5	4.5	AAGGAGACaagcaagca**ATG**		ORF25 (99.4)	ORF23 (98.5)	
+	25	19656	20162	168	18.3	6.0	ATTGAGGGgtgattga**ATG**		ORF26 (94.0)	ORF24 (100)	
+	26	20155	20616	153	15.9	5.0	GAGGAGGTgaaatagt**ATG**	holin	ORF27 (99.3)	ORF25 (100)	
+	27	20618	21670	350	37.5	5.3	GAGGAGGTgaaatagt**ATG**	lysin	ORF28 (98.6)	ORF26 (98.6)	
-	28	22319	21834	161	17.5	6.1	AAGGAGAC taaaa**ATG**	ssDNA-binding protein		ORF27 (99.4)	
-	29	23008	22334	224	25.9	8.6	CAGGTGGCgaacgtga**ATG**			ORF28 (54.5)	
-	30	23760	23005	251	27.3	8.5	TTGGAGGGtaaatg**ATG**	single-strand annealing protein		ORF31 (100)	
-	31	24668	23760	302	33.7	5.6	AAGGAAGCaaaca**ATG**			ORF32 (99.7)	
-	32	24886	24665	73	8.1	5.8	AAGGGGGTaacgtg**ATG**		ORF38 (80.8)		
-	33	25043	24891	50	5.7	10.3	TTGGAGGCagtgga**TTG**		ORF37 (92.0)		
-	34	25128	25036	30	3.3	6.7	AAGAGGCggagaa**ATG**		ORF36 (100)		
-	35	26436	25294	380	44.3	9.9	AAAGAGGTgaaataag**GTG**	transposase			
+	36	26530	26952	140	16.3	10.1	ATTGAGGTaatcat**ATG**	transposase			
-	37	27227	26997	76	8.6	7.9	GAGGCGGTatagat**ATG**	anti-repressor	ORF34 (56.6)		ORF35 (12.0)
-	38	27484	27230	84	9.6	9.5	AAGGAGAAtcatat**ATG**		ORF33 (100)		
-	39	28128	27766	120	12.9	9.8	TTGGAGGGattcatc**ATG**		ORF32 (100)		
-	40	29212	28535	225	24.4	4.2	TCGGAGGGaaata**ATG**	tcdA-E operon negative regulator			
-	41	29505	29296	69	7.9	9.3	AATGGATgacggt**ATG**	anti-repressor			
-	42	29755	29528	75	8.5	9.8	AACGAGGTgattac**ATG**			ORF34 (100)	
+	43	29910	30020	36	4.3	9.3	CCGTAGAActgcg**ATG**				
+	44	30218	30445	75	8.6	9.0	AAAGAGGTaaacaa**CATG**			ORF36 (97.4)	

+	45	30438	30761	107	12.4	5.7	AAGGAGGTaacgatc**ATG**	replication protein		ORF37 (100)	
+	46	30761	31705	314	36.4	8.6	AAGAAGGGtg**ATG**	replication protein		ORF38 (99.4)	
+	47	31702	32964	420	47.4	5.6	AGGGAGGCaacgagC**TTG**	DNA helicase		ORF39 (100)	
+	48	32966	33310	114	13.6	4.9	AAGGTGGCgtattgaa**ATG**			ORF40 (99.1)	ORF47 (13.7)
+	49	33323	33610	95	10.9	5.4	GAGGAGGCaaaaac**ATG**		ORF48 (40.0)	ORF41 (98.9)	
+	50	33597	33851	84	9.4	9.6	CAGGAGGAatcttcaa**ATG**		ORF49 (98.8)	ORF42 (98.8)	ORF51 (88.1)
+	51	33848	34213	121	14.4	6.7	AAGGACGTgaagccagc**ATG**	endodeoxyribonuclease RusA	ORF50 (80.2)	ORF43 (99.2)	ORF52 (98.3)
+	52	34225	34428	67	7.8	6.8	AAGGAGAAaaatc**ATG**		ORF51 (98.5)	ORF44 (97.0)	ORF46 (86.6)
+	53	34547	34930	127	15.0	10.1	CGAGTCGTtagcc**ATG**		ORF52 (100)		
+	54	35239	35772	177	20.0	9.6	AAAGAAAAaataaca**ATG**	putative endodeoxyribonuclease	ORF53 (99.4)		
+	55	35769	35951	60	7.2	8.9	TTGGAGGGaaaaca**ATG**		ORF54 (96.7)		
+	56	35948	36025	25	2.9	10.3	AGGGAGGAgcggca**ATG**		ORF55 (100)	ORF45 (96.0)	ORF35 (17.5)
+	57	36038	36256	72	8.1	5.7	TAGGAGGTgaataat**TTG**	transcriptional regulator	ORF56 (95.8)	ORF46 (93.1)	
+	58	36330	36758	142	16,5	8.4	AAGGAGTGgacgca**ATG**	transcriptional regulator	ORF57 (99.3)	ORF47 (71.4)	ORF56 (72.1)
+	59	37121	37375	84	9.4	4.2	CAGGAGGTaaagaac**ATG**				
+	60	37399	38616	405	46.5	6.2	TTGGAGGCgagtag**ATG**		ORF59 (97.5)		ORF56 (37.3)
+	61	38603	39157	184	20.9	10.1	AAGGCTACggtgcactt**ATG**	HNH homing endonuclease			
+	62	39150	39470	106	12.6	7.7	ACGGGCGAgtgtcccc**ATG**	ribonucleoside-diphosphate reductase	ORF60 (80.7)	ORF50 (66.4)	ORF60 (65.4)

568
569

570 **Table 3.** Features of the ORFs of prophage iA2, predicted function of proteins and best matches with databases

Strand	ORF	Start	Stop	Size (aa)	MM (kDa)	pI	SD sequence AAGGAGGT	Putative function	Organism (extent[a]; % amino acid identity)	E-value	Size (aa)	
											Size (aa)	GenBank accession no.
+	1	58	444	128	14.5	5.2	AAGGGGATgacgaaa**TTG**		Lpp229_06031 (*L. paracasei*) (128/128; 100%)	1E-88	131	EPC46831.1
+	2	447	2177	576	66.2	5.7	AGCGAGGCgaatagca**ATG**	terminase	terminase (*L casei* BL23) (576/576; 100%)	0	576	CAQ66068.1
+	3	2196	3431	411	45.8	5.0	AAGGAGGTgattat**TTG**	Portal protein	phage portal protein (*L casei* BL23) (410/411; 99%)	0	411	CAQ66069.1
+	4	3403	4116	237	25.5	4.6	ACGGAGGGaggtg**ATG**	ATP-dep. Clp protease	Clp protease (*L. casei* BL23) (235/235; 100%)	3E-168	235	CAQ66070.1
+	5	4121	5350	409	43.7	5.7	AAAGGGGAataacctc**ATG**	major capsid protein	phage capsid protein (*L. casei* BL23) (409/409; 100%)	0	409	CAQ66071.1
+	6	5424	5672	82	8.3	4.3	AAGGAGCTagaac**ATG**		LCABL_09860 (*L. casei* BL23) (82/82; 100%)	2E-48	82	CAQ66072.1
+	7	5686	6012	108	12.2	4.3	TAGGCGGTgatcaaag**ATG**		LCABL_09870 (*L. casei* BL23) (108/108; 100%)	1E-71	108	CAQ66073.1
+	8	5951	6289	112	13.2	8.0	TTGGAGGTgagaacag**ATG**	head-tail adaptor	LCABL_09880 (*L. casei* BL23) (112/112; 100%)	3E-74	112	CAQ66074.1
+	9	6273	6602	109	12.6	4.6	ATGGAGTTgagtttggact**ATG**		LCABL_09890 (*L. casei* BL23) (109/109; 100%)	3E-72	109	CAQ66075.1
+	10	6592	6975	127	14.1	9.2	ATCGAGGTgattgcgg**GTG**	tail component	LCABL_09900 (*L. casei* BL23) (127/127; 100%)	1E-88	127	CAQ66076.1
+	11	6987	7634	215	23.6	5.4	TAGGAGGAatcgct**ATG**	major tail protein	tail protein (*L. casei* BL23) (215/215; 100%)	4E-157	215	CAQ66077.1
+	12	7711	8076	121	13.7	6.7	TAGGAGGAcaaaggaat**ATG**		LCABL_09920 (*L. casei* BL23) (121/121; 100%)	4E-83	121	CAQ66078.1
+	13	8136	8318	60	7.2	9.0	CTGGGCCTtttat**TTG**		AF91_09510 (*L. paracasei* N1115) (58/60; 98%)	1E-33	60	AHJ33419.1
+	14	8338	11310	990	105.5	9.4	AAGGAGGAaaata**ATG**	tape-measure protein	tape measure (*L. casei* BL23) (990/990; 100%)	0	990	CAQ66080.1
+	15	11317	12012	231	25.3	5.4	AAGGAGGCtttgatgacgg**ATG**	tail protein	Gp13 protein (*L. casei* BL23) (231/231; 100%)	3E-166	231	CAQ66081.1
+	16	12009	16415	1468	158.8	4.9	TAGGTGGTgttatgt**ATG**	tail protein, endopeptidase	PblB (*L. casei* LC2W) (1468/1468; 100%)	0	1468	AEA53317.1
+	17	16444	16869	141	15.8	4.7	TGGGAGGAaaaac**ATG**		LCABL_09970 (*L. casei* BL23) (141/141; 100%)	4E-95	141	CAQ66083.1
+	18	16872	17141	89	9.7	4.5	AAGGAGAGaagtaatc**ATG**		orf59 (*L. casei* BL23) (89/89; 100%)	8E-55	89	CAQ66084.1
+	19	17187	17480	97	10.8	5.6	AAGAAGTGagaaa**GTG**		LCABL_09990 (*L. casei* BL23) (96/97; 99%)	4E-63	97	CAQ66085.1
+	20	17470	17904	144	14.8	4.8	AAGGAGAAaataaccc**ATG**	holin	holin (*L. casei* BL23) (144/144; 100%)	1E-92	144	CAQ66086.1
+	21	17904	18092	62	6.9	5.0	CAGGAGGCcaagta**ATG**		LCABL_10010 (*L. casei* BL23) (62/62; 100%)	1E-34	62	CAQ66087.1
+	22	18079	19053	324	34.8	4.8	GAGGAGGCgtacacga**ATG**	lysin	amidase (*L. casei* BL23) (324/324; 100%)	0	324	CAQ66088.1
-	23	19435	20574	379	43.8	9.7	ATGGGGGCaagtgac**ATG**	integrase	integrase (*L. casei* BL23) (379/379; 100%)	0	379	CAQ66041.1
-	24	20692	21759	355	40.6	6.3	AGTGGGGTaagataa**TTG**	Abi family protein	CAAX protease (*L. casei* BL23) (355/355; 100%)	0	355	CAQ66042.1
-	25	21904	22335	143	15.8	5.1	AAGGAGGTaatagc**ATG**		LCABL_09570 (*L. casei* BL23) (143/143; 100%)	7E-98	143	CAQ66043.1
-	26	22405	22887	160	19.0	4.9	AAGGAGGAttagctg**ATG**		LCABL_09580 (*L. casei* BL23) (160/160; 100%)	7E-110	160	CAQ66044.1
-	27	22891	23184	97	11.6	8.6	AAGGACCCtcgact**ATG**		LCABL_09590 (*L. casei* BL23) (97/97; 100%)	5E-63	97	CAQ66045.1
-	28	23368	24228	286	33.0	5.2	AGGGAAAAgcaaa**ATG**		LCABL_09600 (*L. casei* BL23) (286/286; 100%)	0	286	CAQ66046.1
-	29	24215	24583	122	13.5	6.6	TTGGAGGGatttttt**ATG**	Cro/CI regulator	Cro/CI regulator (*L. casei* BL23) (122/122; 100%)	1E-81	122	CAQ66047.1
+	30	24839	25036	65	7.3	10.0	TAGGAGGTgaccactggt**ATG**		LC2W_0948 (*L. casei* LC2W) (65/65; 100%)	2E-38	65	AEA53282.1
+	31	25033	25755	240	26.7	9.8	AAGGAGGAatccaa**ATG**	putative antirepressor	LCABL_09620 (*L. casei* BL23) (240/240; 100%)	8E-175	240	CAQ66048.1
+	32	25763	25975	70	7.7	8.2	AAGGAGTGaagtacgt**TTG**		BN194_09360 (*L. casei* W56) (69/70; 99%)	4E-40	70	CCK21883.1
+	33	26084	26308	74	9.1	8.9	AACAGCGTgccgac**ATG**		LCABL_09630 (*L. casei* BL23) (74/74; 100%)	3E-46	74	CAQ66049.1
+	34	26308	26400	30	3.3	8.0	AAGGAATTggtgaata**ATG**		LCABL_09640 (*L. casei* BL23) (30/30; 100%)	1E-11	30	CAQ66050.1
+	35	26397	26609	70	8.1	4.2	GAGGCAGTggatcga**ATG**		LCABL_09650 (*L. casei* BL23) (70/70; 100%)	2E-42	70	CAQ66051.1
+	36	26619	27494	291	33.7	5.9	TAGGGGGTtgtt**ATG**	recombinase	protein yqaJ (*L. casei* BL23) (291/291; 100%)	0	291	CAQ66052.1
+	37	27497	27691	64	7.1	4.3	AAGAAGATtgaat**ATG**		LCABL_09670 (*L. casei* BL23) (64/64; 100%)	1E-37	64	CAQ66053.1
+	38	27691	28578	295	32.8	5.4	AAGGAGGAtgaagacta**ATG**	RecT DNA repair protein	RecT protein (*L. casei* BL23) (295/295; 100%)	0	295	CAQ66054.1
+	39	28587	28832	81	9.2	8.9	AAGTAGGTgagctg**ATG**	HTH transcript. regulator	orf8 (*L. casei* BL23) (81/81; 100%)	3E-51	81	CAQ66055.1
+	40	28837	29682	281	32.7	9.1	GCAGAGGTgatctg**ATG**	DNA damage-inducible	DnaD (*L. casei* BL23) (281/281; 100%)	0	281	CAQ66056.1
+	41	29720	30541	273	30.6	9.6	AACGAGGTgaaac**ATG**	DNA replication protein	Prophage pi3 protein46 (*L. casei* BL23)(273/273; 100%)	0	273	CAQ66057.1

+	42	30538	30837	99	11.6	4.9	AAGGAAGGcatgtc**ATG**		LCABL_09720 (*L. casei* BL23) (99/99; 100%)	6E-65	99	CAQ66058.1
+	43	30800	31084	94	10.6	4.8	AAGCGGGAaagttcg**ATG**		LC2W_0960 (*L. casei* LC2W) (94/94; 100%)	2E-61	94	AEA53294.1
+	44	31053	31400	115	14.0	10.0	AATGAGGTgactgaaa**ATG**		LCBD_0957 (*L. casei* BD-II) (115/115; 100%)	8E-79	115	AEA56455.1
+	45	31393	31971	192	22.1	10.4	GAGGAGCTggcgct**ATG**		LCABL_09750 (*L. casei* BL23) (192/192; 100%)	2E-138	192	CAQ66061.1
+	46	31988	32407	139	15.6	9.0	AGGGAGGAttgatg**ATG**	Hollidayjunction resolvase	resolvase (*L. casei* BL23) (139/139; 100%)	2E-95	140	CAQ66062.1
+	47	32412	32675	87	9.6	5.1	AAGGAGGTctaacgcc**ATG**		LCABL_09770 (*L. casei* BL23) (87/87; 100%)	2E-53	87	CAQ66063.1
+	48	32699	33022	107	13.1	9.8	GCGGAGGTtaagaa**ATG**		LCABL_09780 (*L. casei* BL23) (107/107; 100%)	3E-71	107	CAQ66064.1
+	49	33101	33550	149	17.9	5.1	AAGGAGTGgggccg**TTG**	transcriptional regulator	LCABL_09790 (*L. casei* BL23) (148/149; 99%)	6E-101	149	CAQ66065.1
+	50	33775	34155	126	14.7	8.7	TTGGAGGTggatatg**ATG**	HNH endonuclease	LCBD_0963 (*L. casei* BD-II) (126/126; 100%)	2E-89	127	AEA56461.1

[a]Number of identical amino acids/total number of amino acids.

574 **Table 4.** tRNA found in *L. casei* group phages using the bioinformatic tool ARAGORN.

Phage	Accession number	Genome size (bp)	Number of tRNA	tRNA begin	tRNA end	%GC	tRNA type	Anticodon
Lb338-1	FJ822135.1	142,111	5	21,370	21,442	47.9	Undetermined (Asn|Stop)	
				23,807	23,880	41.9	Arg	cct
				24,706	24,779	48.6	Arg	tct
				25,603	25,691	51.7	Ile	tat
				27,758	27,832	44.0	Thr	tgt
C_L2	KR905067	38,751	1	36,764	36,836	53.4	Undetermined (Asp|seC)	
i*Lp*84	KR905069	39,399	1	36,908	36,981	52.7	Ile	tat
A2	AJ251789.2	43,411	1	38,396	38,466	62.0	Leu	taa
Lc-Nu	AY131267.2	36,466	1	32,542	32,614	53.4	Undetermined (Asp|seC)	
Lrm1	EU246945.1	39,989	1	36,665	36,737	53.4	Met	cat

575
576

577 **List of Figures.**

578

579 **Figure 1.** Schematic representation of the genome organization of different *Lactobacillus*
580 phages. Each line corresponds to a different genome and starts from its physical end.
581 Predicted ORFs are represented by arrows. ORFs with the same color in different phages
582 possess amino acid identity higher than 80%. The white ORFs have less than 80% identity
583 with any another ORF. Asterisks indicate transposases in the genomes of phages C_L1 and
584 C_L2.

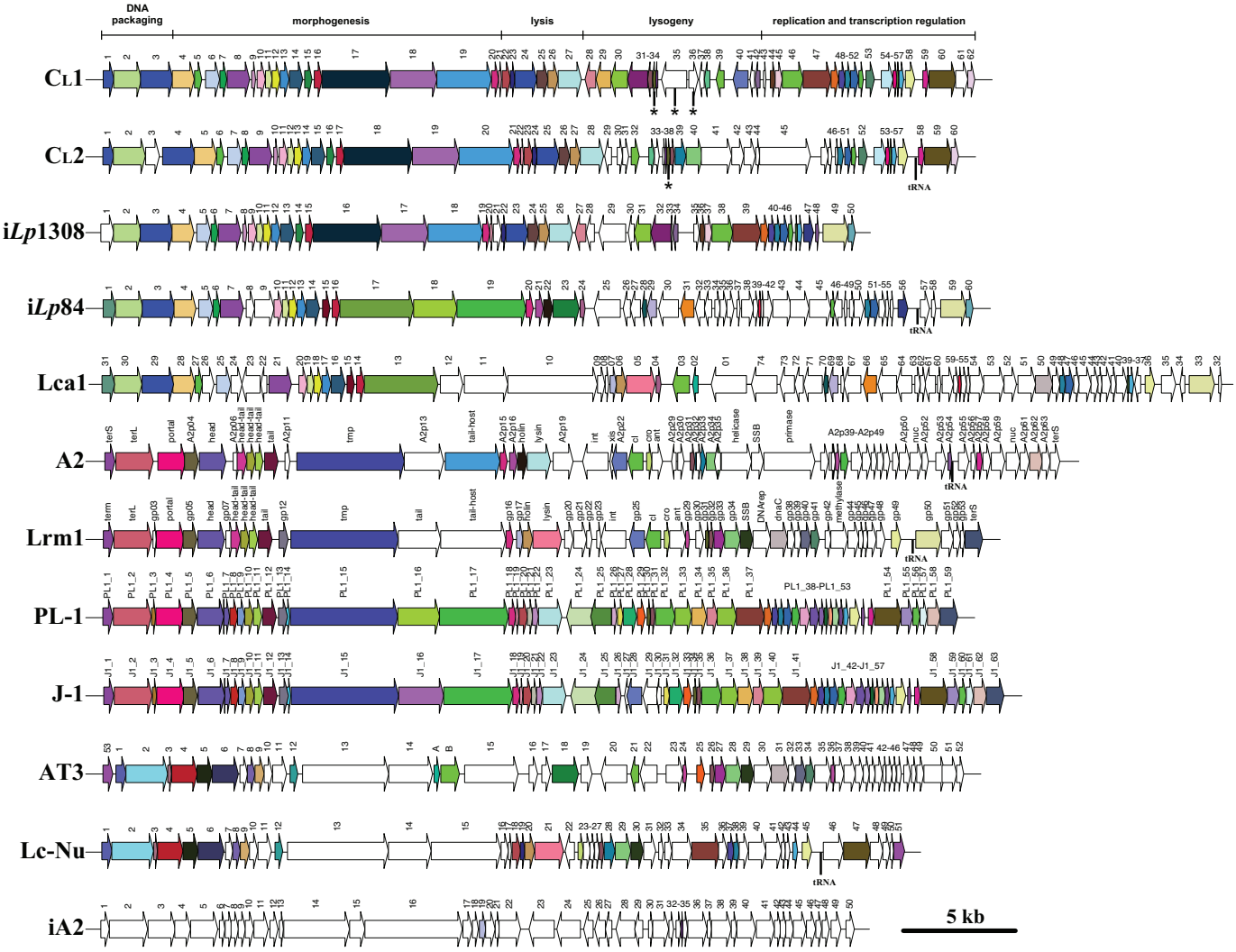
585

586 **Figure 2.** Maximum-likelihood tree for the taxonomy of several phages and prophages
587 according to amino acid sequences of the large terminase subunit. Sequences were obtained
588 in this study and from multiple phage genomes available in databases. Phage names are
589 indicated at the end of each terminal branch. Colors categorize phages with different
590 encapsidation systems: Blue, 3'-extended *cos* ends; orange, Mu-like headful; violet,
591 lambda-like 5'-extended *cos*; light blue, T7-like direct terminal repeats; light green, T4-like
592 headful; dark green, P2-like 5'-extended *cos* ends; pink, GTA headful; red, P22-like
593 headful.

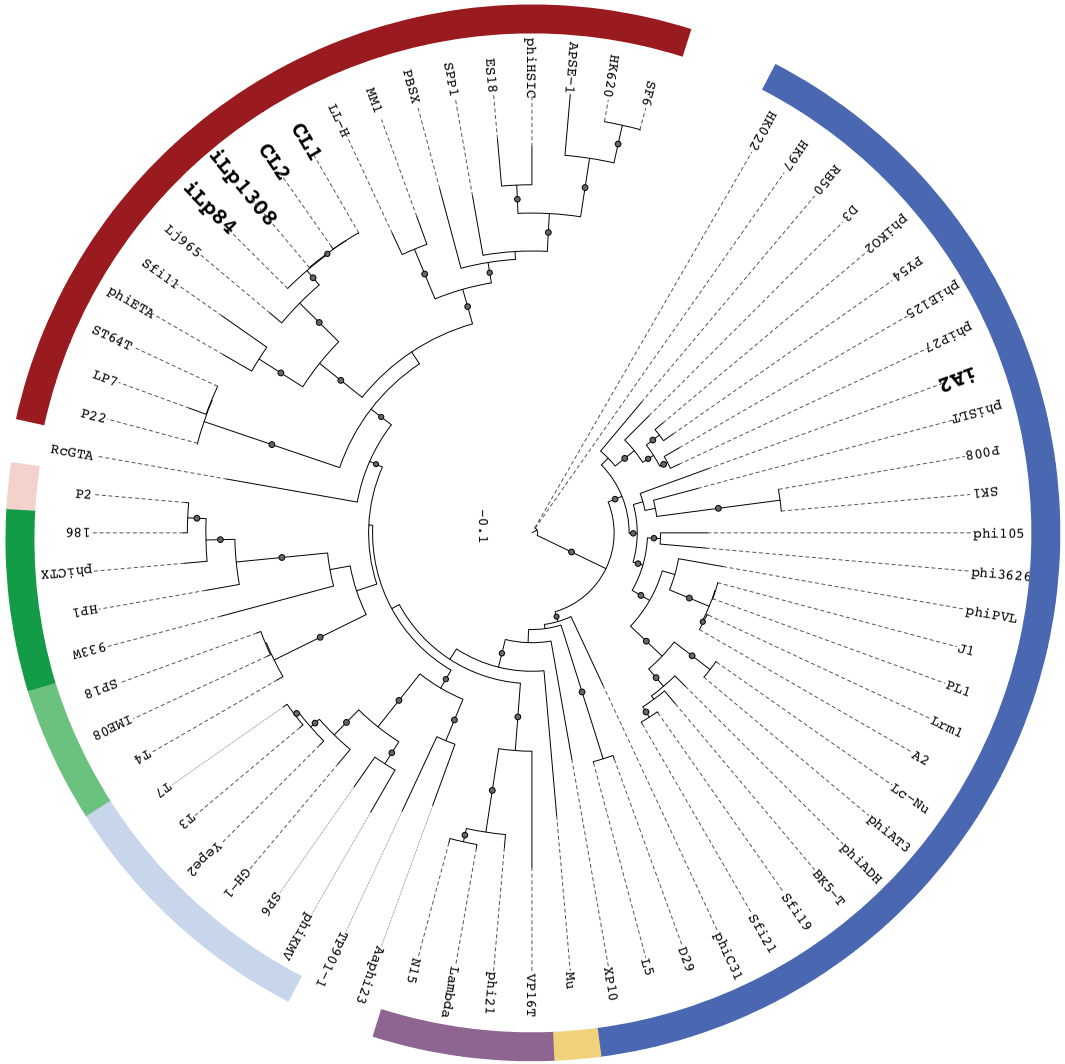
594

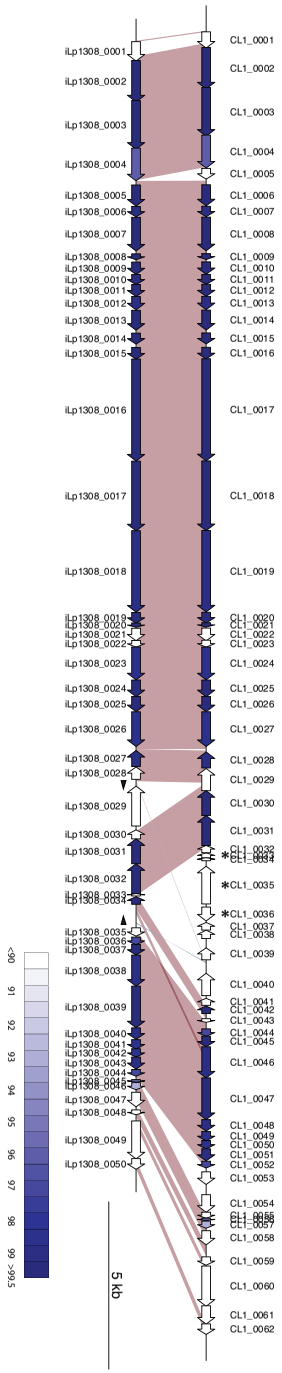
595 **Figure 3**. Comparison of the DNA and amino acid identity (blue-filled arrows) between the
596 phages C_L1 and i*Lp*1308. The level of amino-acid identity is shown with a white-blue
597 gradient where ORFs represented by white arrows share less than 90% identity. The ORFs
598 represented by blue arrows share 90% identity or more, with the shade of blue increasing
599 with the level of identity as indicated by the legend. DNA identity between both genomes is

600 also indicated for sequences running in the same (red shadows) or in opposite directions
601 (blue shadows). Asterisks indicate transposases in the genome of phage C_L1. Black
602 triangles indicate a 285-nt inverted repeats in the genome of phage i*Lp*1308.
603
604 **Figure 4.** Genome organization of prophage iA2 and prophages extracted from the
605 available genomic sequences of *L. casei* group strains. Genomes start from the attachment
606 sites. ORFs are indicated by arrows. ORFs with the same color in different prophages
607 possess amino acid identity higher than 99%. White ORFs are unique. Groups indicate
608 prophages that are highly conserved in different strains.



Applied and Environmental Microbiology


0.1
iLp1308
CL2
CL1
iLp84
SF6
HK620
APSE-1
phiHSIC
ES18
SPP1
PBSX
MM1
LL-H
Lj965
Sfi11
phiETA
ST64T
LP7
P22
RcGTA
P2
186
phiCTX
HP1
933W
SP18
IME08
T4
T7
T3
Yepe2
GH-1
SPO
phiKMV
TP901-1
Aaphi23
N15
Lambda
phi21
VP16T
Mu
XP10
L5
D29
phiC31
Sfi21
Sfi19
BK5-T
phiAD
phiAT3
Lc-Nu
A2
Lrm1
PL1
J1
phiPVL
phi3626
phi105
SK1
p008
phiSLT
iA2
phiP27
phiFL2S
PY54
phiKO2
D3
R850
HK97
HK022


iLp1308_0001
iLp1308_0002
iLp1308_0003
iLp1308_0004
iLp1308_0005
iLp1308_0006
iLp1308_0007
iLp1308_0008
iLp1308_0009
iLp1308_0010
iLp1308_0011
iLp1308_0012
iLp1308_0013
iLp1308_0014
iLp1308_0015
iLp1308_0016
iLp1308_0017
iLp1308_0018
iLp1308_0019
iLp1308_0020
iLp1308_0021
iLp1308_0022
iLp1308_0023
iLp1308_0024
iLp1308_0025
iLp1308_0026
iLp1308_0027
iLp1308_0028
iLp1308_0029
iLp1308_0030
iLp1308_0031
iLp1308_0032
iLp1308_0033
iLp1308_0034
iLp1308_0035
iLp1308_0036
iLp1308_0037
iLp1308_0038
iLp1308_0039
iLp1308_0040
iLp1308_0041
iLp1308_0042
iLp1308_0043
iLp1308_0044
iLp1308_0045
iLp1308_0046
iLp1308_0047
iLp1308_0048
iLp1308_0049
iLp1308_0050
CL1_0001
CL1_0002
CL1_0003
CL1_0004
CL1_0005
CL1_0006
CL1_0007
CL1_0008
CL1_0009
CL1_0010
CL1_0011
CL1_0012
CL1_0013
CL1_0014
CL1_0015
CL1_0016
CL1_0017
CL1_0018
CL1_0019
CL1_0020
CL1_0021
CL1_0022
CL1_0023
CL1_0024
CL1_0025
CL1_0026
CL1_0027
CL1_0028
CL1_0029
CL1_0030
CL1_0031
CL1_0032
CL1_0033
CL1_0034
*CL1_0035
*CL1_0036
CL1_0037
CL1_0038
CL1_0039
CL1_0040
CL1_0041
CL1_0042
CL1_0043
CL1_0044
CL1_0045
CL1_0046
CL1_0047
CL1_0048
CL1_0049
CL1_0050
CL1_0051
CL1_0052
CL1_0053
CL1_0054
CL1_0055
CL1_0056
CL1_0057
CL1_0058
CL1_0059
CL1_0060
CL1_0061
CL1_0062
<90
91
92
93
94
95
96
97
98
99
>99.5
5 kb