APPLIED AND ENVIRONMENTAL MICROBIOLOGY, Dec. 2005, p. 8371–8382 Vol. 71, No. 12
0099-2240/05/$08.00+0 doi:10.1128/AEM.71.12.8371–8382.2005


# Complete Sequences of Four Plasmids of *Lactococcus lactis* subsp. *cremoris* SK11 Reveal Extensive Adaptation to the Dairy Environment†

Roland J. Siezen,[1,2,3]* Bernadet Renckens,[1,2,3] Iris van Swam,[1,3] Sander Peters,[4] Richard van Kranenburg,[1,3]‡ Michiel Kleerebezem,[1,3] and Willem M. de Vos[1]

*Wageningen Centre for Food Sciences, P.O. Box 557, 6700AN Wageningen,[1] Centre for Molecular and Biomolecular Informatics, Radboud University, P.O. Box 9010, 6500GL Nijmegen,[2] NIZO food research, P.O. Box 20, 6710BA Ede,[3] and Greenomics, Plant Research International, P.O. Box 16, 6700AA Wageningen,[4] The Netherlands*

Received 28 June 2005/Accepted 4 August 2005

***Lactococcus lactis* strains are known to carry plasmids encoding industrially important traits. *L. lactis* subsp. *cremoris* SK11 is widely used by the dairy industry in cheese making. Its complete plasmid complement was sequenced and found to contain the plasmids pSK11A (10,372 bp), pSK11B (13,332 bp), pSK11L (47,165 bp), and pSK11P (75,814 bp). Six highly homologous *repB*-containing replicons were found, all belonging to the family of lactococcal theta-type replicons. Twenty-three complete insertion sequence elements segment the plasmids into numerous modules, many of which can be identified as functional units or containing functionally related genes. Plasmid-encoded functions previously known to reside on *L. lactis* SK11 plasmids were now mapped in detail, e.g., lactose utilization (*lacR-lacABCDFEGX*), the proteolytic system (*prtM-prtP*, *pepO*, *pepF*), and the oligopeptide permease system (*oppDFBCA*). Newly identified plasmid-encoded functions could facilitate the uptake of various cations, while the *pabA* and *pabB* genes could be essential for folate biosynthesis. A competitive advantage could be obtained by using the putative flavin adenine dinucleotide-dependent D-lactate dehydrogenase and oxalate:formate antiporter for enhanced ATP synthesis, while the activity of the predicted α-acetolactate decarboxylase may contribute to the formation of an additional electron sink. Various stress response proteins are plasmid encoded, which could enhance strain robustness. A substantial number of these "adaptation" genes have not been described before on *L. lactis* plasmids. Moreover, several genes were identified for the first time in *L. lactis*, possibly reflecting horizontal gene transfer.***

*Lactococcus lactis* is a fermentative lactic acid bacterium that is used extensively in food fermentations and has great biotechnological and economic importance. *L. lactis* is one of the main bacteria used in starter cultures by the dairy industry for the production of cheese, fermented milks, etc. (14, 15). Many *L. lactis* subsp. *lactis* and subsp. *cremoris* strains are known to carry plasmids encoding important traits such as lactose catabolism, citrate utilization, proteinase production, bacteriocin production and immunity, bacteriophage resistance, exopolysaccharide production, and antibiotic and heavy metal resistance (15, 48). These properties can contribute to the desired flavor and texture of the product and optimal growth on the milk components lactose and casein, as well as stability and survival. As plasmids are mobile elements, they can be lost and acquired. The latter process has been studied in *L. lactis*, and notably conjugal transfer has shown to be highly efficient (33). It is for this reason that *L. lactis* strains and their plasmids have been studied extensively in the past, in an effort to harness this extrachromosomal metabolic potential and exploit these characteristics to improve starters for the food industry.

* Corresponding author. Mailing address: Centre for Molecular and Biomolecular Informatics, Radboud University, P.O. Box 9010, 6500GL Nijmegen, The Netherlands. Phone: 31-318-659511. Fax: 31-318-650400. E-mail: Siezen@cmbi.ru.nl.

† Supplemental material for this article may be found at http://aem.asm.org/.

‡ Present address: Purac Biochem, P.O. Box 21, 4200AA Gorinchem, The Netherlands.

Most *L. lactis* strains have been found to harbor several plasmids, ranging in size from 2 to 80 kb (15). While several complete plasmid sequences are known (see Table S1 in the supplemental material), not many plasmids larger than 10 kb have been sequenced. The largest of these, the conjugative 60.2-kb plasmid pMRC01 isolated from *L. lactis* subsp. *lactis* DPC3147, has three modules, flanked by insertion sequences (IS), encoding bacteriocin production, bacteriophage resistance, and mobilization (23). The 42.8-kb mobilization plasmid pNZ4000 from *L. lactis* subsp. *cremoris* NIZO B40 has a module encoding exopolysaccharide biosynthesis but also encodes putative cobalt and magnesium transport systems (68).

*L. lactis* subsp. *cremoris* SK11 (previously described as *Streptococcus cremoris* SK11) is a phage-resistant strain used in cheese making. Several studies of this strain and its derivatives have shown that proteinase activity and phage resistance are plasmid-encoded functions that can be cured, while the plasmid-encoded lactose utilization ability is extremely stable (17, 18, 29, 38). Detailed studies demonstrated that SK11 cultures are heterogeneous with respect to plasmid content, and isolates were found to contain different combinations of at least eight different plasmids (17, 18). Only small parts of these plasmids have been sequenced in the past 2 decades. The largest plasmid, pSK111, which is 72 to 78 kb (17, 18, 60) long, has a region of about 10 kb, encoding the cell envelope proteinase PrtP and its maturase PrtM, that has been fully sequenced (71, 72). Flanking iso-IS*S1* elements were found, sug-

gesting that the proteinase regulon is part of a composite transposon, designated Tn*5277* (36). This plasmid, pSK111, is often referred to as the proteinase plasmid. This *prtP*/*prtM* system is required for utilization of the caseins in milk as a nitrogen source and for a fast milk coagulation phenotype (45). Curing experiments suggest that phage resistance is encoded on the next largest plasmid, pSK112, but no sequence information is available (17, 63). The most extensively studied plasmid is pSK113, better known as pSK11L or the lactose (Lac) plasmid, which is 47 to 48 kb long (17, 29, 38, 74, 75). This plasmid encodes a module with the lactose operon, flanked by IS*S1* elements, which is essential for its lactose-fermenting ability (38). In addition, a functional Opp oligopeptide transport system is encoded in a different module, flanked by IS*S1* and IS*982* elements (75). This Opp system, in combination with the PrtP/PrtM system, is required for a fast milk-coagulating phenotype (74). The lactose plasmid pSK11L is extremely stable in *L. lactis* subsp. *cremoris* strains (17, 29) but shows instability and a temperature-sensitive growth response when transferred to *L. lactis* subsp. *lactis* LM0230 (29, 38). In this case, stabilization has been observed at higher temperatures following integration of the Lac plasmid into the chromosome.

In this paper, we present the complete sequences of four plasmids of *L. lactis* subsp. *cremoris* NIZO B697, a derivative of strain SK11, obtained by shotgun sequencing of total plasmid DNA. Several of the known plasmid-encoded functions have now been characterized at the sequence level for the first time, while many new genes have been identified for which functions can be predicted.

## MATERIALS AND METHODS

**Bacterial strains, plasmid preparation, and manipulation.** *Lactococcus lactis* subsp. *cremoris* NIZO B697 (NIZO Culture Collection, Ede, The Netherlands), a single-colony isolate of *Lactococcus lactis* subsp. *cremoris* SK11 culture NCDO2004, with a $Prt^+$ $Lac^+$ phenotype (17), was selected. Isolation of plasmid DNA was performed as previously described (19). Standard recombinant DNA techniques and gel electrophoresis were performed as described by Sambrook et al. (55).

**Shotgun clone preparation.** Plasmid DNA was sheared by nebulization into fragments with an average size of 1,200 bp. Blunt repair of the ends was performed with *Pfu* DNA polymerase (Stratagene, Amsterdam, The Netherlands) according to the manufacturer's directions. DNA fragments were size fractionated by gel electrophoresis and cloned into the dephosphorylated EcoRV site of pBluescriptSK (Stratagene). After transformation into XL2-Blue competent cells (Stratagene), recombinant clones were randomly picked. For preparation of the phage l sequence library, the Lambda ZAP II RI Vector/Gigapack cloning kit (Stratagene) was used according to the manufacturer's protocol. DNA fragments were prepared by a Tsp509I partial digestion of genomic DNA such that an average restriction fragment size of 8 kb was obtained.

**Sequence determination.** Sequencing was performed using an ABI PRISM BigDye Terminator cycle sequencing ready-reaction kit with FS AmpliTaq DNA polymerase (PerkinElmer, Boston, Mass.) and analyzed on an ABI 3730XL DNA analyzer. A total of 1,536 sequence reads (*Escherichia coli* bank, 768 reads; λ bank, 768 reads), with an average length of 550 bp, provided over 800 kb which could be assembled (see below) into a unique sequence of about 140 kb distributed over 63 contigs with an average of 6.0-fold coverage.

Shotgun sequences were base called with the PHRED base caller and assembled using the Gap4 assembler of the Staden package (release 1.5.3). Using the PREGAP4 interface, GAP4-assembled sequences were parsed into the GAP4 assembly database (9). The GAP4 interface and its features were then used for editing and sequence finishing. Consensus calculations with a quality cutoff value of 40 were performed from within GAP4 using a probabilistic consensus algorithm based on expected error rates output by PHRED (26, 27). Sequencing the PCR products using custom-designed primers, we bridged the ends of contiguous fragments to fill the remaining sequence gaps. To verify the assembly, read pairs were analyzed by direction and size using a maximum spacing of 2.5 kb for normal shotgun clones and 8 kb for phage λ clones.

Gap closing was assisted by using contig information from the sequencing of total DNA, including plasmids, from *L. lactis* subsp. *cremoris* SK11 (42) by the Joint Genome Institute (JGI; http://genome.jgi-psf.org/mic_home.html) as templates. At the time of our analysis, the JGI draft sequence of about 2.3 Mb consisted of 83 scaffolds ranging in size from 2,643 to 198,531 bp. Major discrepancies with this JGI sequence information, resulting mostly from different assembly at IS elements, were checked by resequencing.

**Sequence annotation.** Annotation was first performed automatically using the Pedant-Pro (Biomax Informatics AG, Martinsreid, Germany) integrated annotation package, which includes BLASTP/BLASTN analysis (1), and homology analysis against the PFAM protein family database (3), the COG (cluster of orthologous groups) gene database (64), and the PROSITE (28) and BLOCKS (37) signature databases. Subsequently, all open reading frames (ORFs) were scrutinized manually using primarily the Artemis viewer (5), the ERGO bioinformatics suite (http://ergo.integratedgenomics.com/ERGO), and recent BLASTP analysis against the National Center for Biotechnology Information database (http://www.ncbi.nlm.nih.gov/BLAST) to identify putative frameshifts and point mutations and to refine the start codon assignments based on optimal ribosome-binding sites (RBS), full-length alignment of the encoded proteins with homologues, and the presence of identifiable motifs. Putative frameshift and point mutations were resequenced and corrected when appropriate, while ORFs containing verified mutations were designated as putatively inactive genes (pseudogenes). Intergenic regions were searched with TBLASTX for homology to fragments of known genes. IS elements were identified by comparison with the IS database (www-is.biotoul.fr) of the Centre National de la Recherche Scientifique. The codon adaptation index (CAI) was calculated using ribosomal proteins as a reference set (43, 58).

**Nucleotide sequence accession numbers.** The complete nucleotide sequences of plasmids pSK11A, pSK11B, pSK11L, and pSK11P have been submitted to the EMBL/GenBank database and are available under accession numbers DQ149242, DQ149243, DQ149244, and DQ149245, respectively.

## RESULTS AND DISCUSSION

**Plasmid DNA sequencing.** Plasmids are highly common in *L. lactis* strains, notably in industrially used strains (see Table S1 in the supplemental material). However, large variations in plasmid complement are known, and this has been described in detail for *L. lactis* subsp. *cremoris* SK11, for which isolates with up to seven different plasmids have been described (17, 18). The parental strain *L. lactis* subsp. *cremoris* SK11 was shown to have four plasmids of about 72, 48, 15, and 9.6 kb (29), and this was also observed in the derivative strain used in this study. Shotgun sequencing showed that nearly half of the sequence reads represented a single plasmid that assembled to a circular plasmid of 13,332 bp (~30-fold coverage), designated pSK11B (Fig. 1). Subsequent gap closing with the remaining contigs led to the identification of three additional complete circular plasmids, pSK11P of 75,184 bp (proteinase plasmid), pSK11L of 47,165 bp (lactose plasmid), and pSK11A of 10,700 bp, as depicted in Fig. 1 and 2. Their G+C content of 31 to 36% was similar to that of other *L. lactis* plasmids (see Table S1 in the supplemental material). The high overrepresentation of pSK11B reads suggests that this is a high-copy-number plasmid relative to the other three plasmids, which were represented about equally in the remaining sequence reads. These 4 plasmids are among the 10 largest *L. lactis* plasmids sequenced to date (see Table S1 in the supplemental material).

The plasmid sequences have been annotated and analyzed in detail (see Table S2 in the supplemental material). In addition to the salient specifics of each plasmid, we will discuss aspects of their replication and mobilization and the presence of IS elements and other possible mobile elements in detail below.

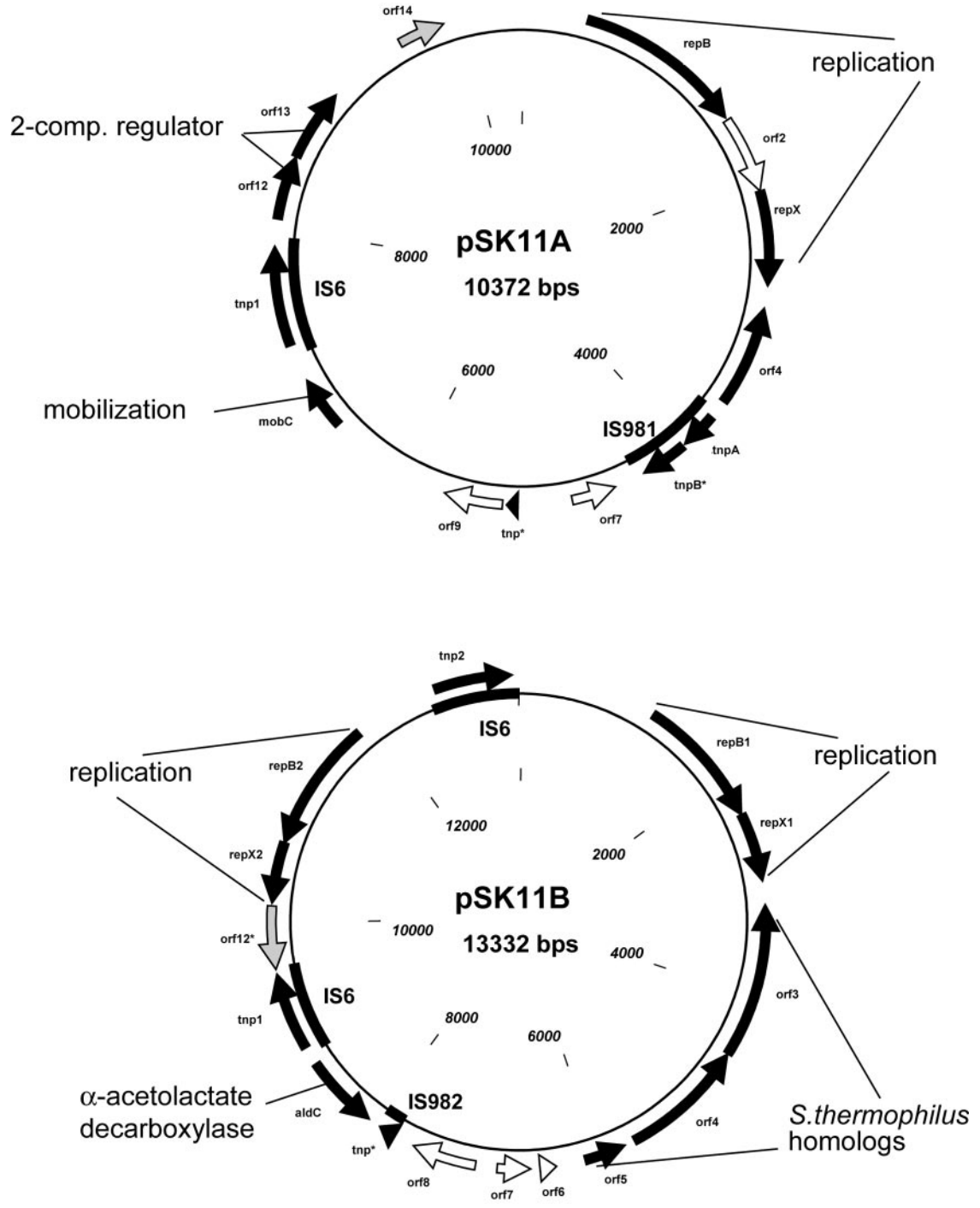
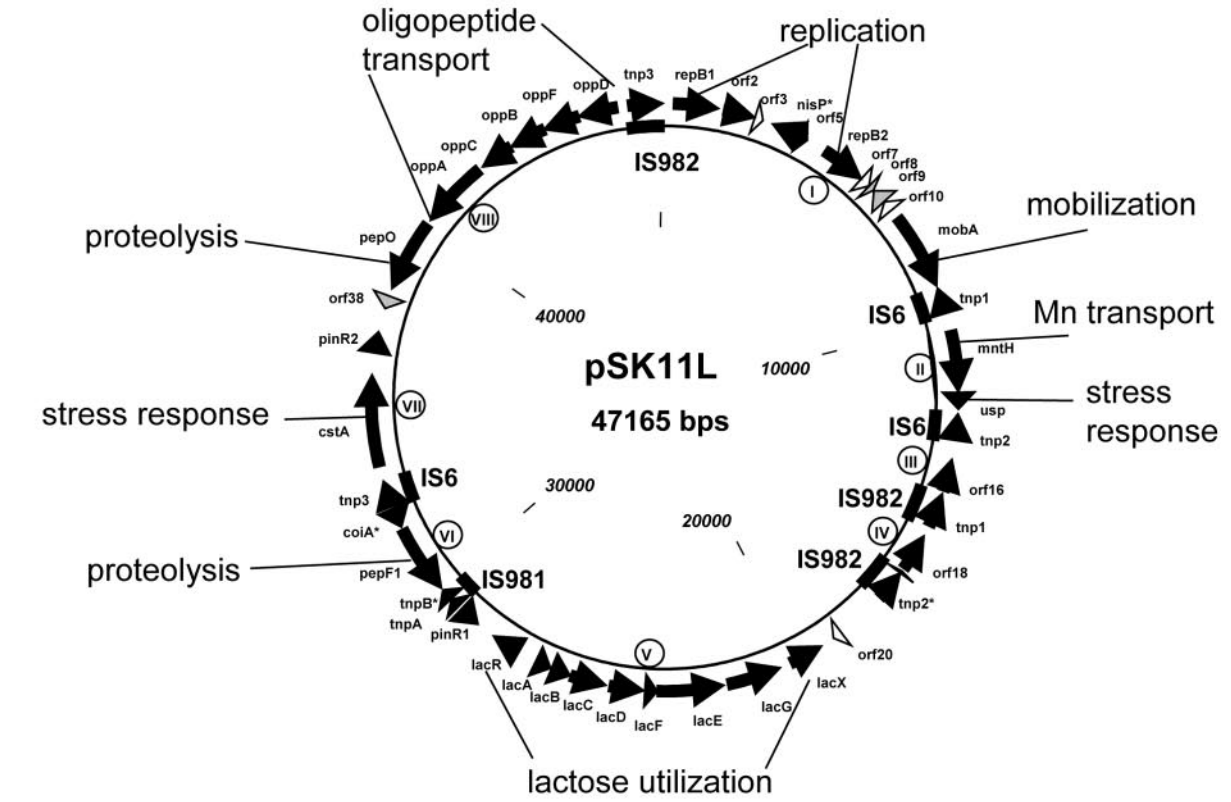
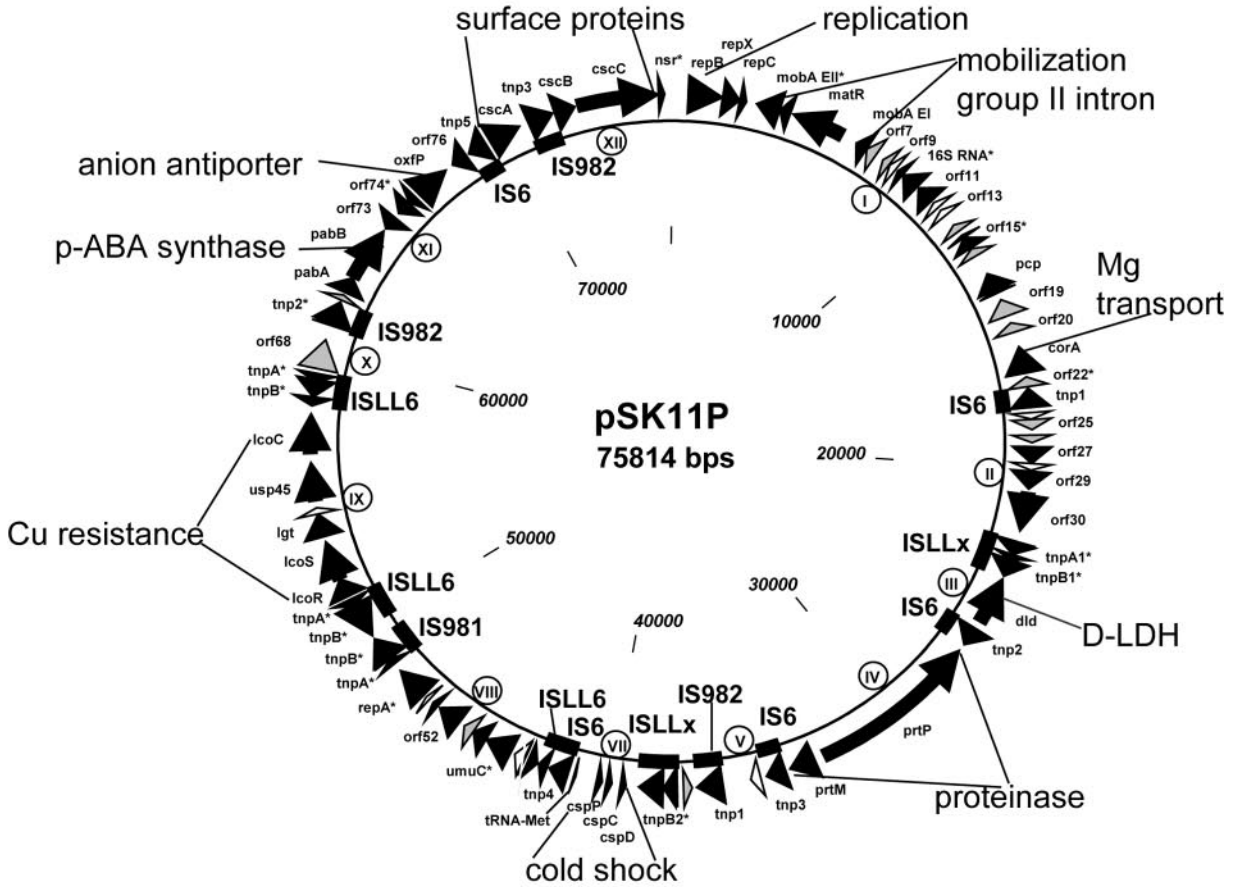



FIG. 1. Physical and genetic maps of plasmids pSK11A and pSK11B of *L. lactis* subsp. *cremoris* SK11. Arrows indicate positions, directions, and names of genes encoding proteins of predicted function (black), conserved hypothetical proteins (gray), and hypothetical proteins (white). An asterisk indicates a pseudogene. Inner circles show the nucleotide numbering and IS elements (black boxes). comp., component.

**Replication.** Generally, there are two types of plasmids in lactococci, known as the rolling-circle replicating and theta-replicating plasmids, the latter of which appears to be more common (16, 34, 41). Six replicons were identified on the four plasmids, all belonging to the family of lactococcal theta replicons, with two replicons on pSK11L and pSK11B each. These replicons are all characterized by a *repB* gene encoding the replication initiation protein RepB of 384 to 391 amino acid residues that has nicking-closing (topoisomerase I) activity.

The upstream noncoding regions of these *repB* genes contain several structural motifs characteristic of replication origins of the lactococcal theta-type replicons (Fig. 3) (34, 57), which are highly conserved among the theta-type replicons. These include an AT-rich region, the presumed origin of replication


oligopeptide transport
replication
mobilization
proteolysis
Mn transport
stress response
stress response
proteolysis
lactose utilization
pSK11L
47165 bps
IS982
IS6
IS982
IS6
IS982
IS6
IS981
oppD
oppF
oppB
oppC
oppA
tnp3
repB1
orf2
orf3
nisP*
orf5
repB2
orf7
orf8
orf9
orf10
mobA
tnp1
mntH
usp
tnp2
orf16
tnp1
orf18
tnp2*
orf20
lacX
lacG
lacE
lacF
lacD
lacC
lacB
lacA
lacR
pinR1
tnpA
tnpB*
pepF1
coiA*
tnp3
cstA
pinR2
orf38
pepO
40000
10000
20000
30000


surface proteins
replication
mobilization
group II intron
anion antiporter
p-ABA synthase
Mg transport
Cu resistance
D-LDH
proteinase
cold shock
pSK11P
75814 bps
IS982
ISLL6
IS6
ISLLx
IS6
ISLL6
IS981
ISLL6
IS6
ISLLx
IS982
IS6
repX
repB
repC
nsr*
cscC
cscB
tnp3
cscA
tnp5
orf76
oxfP
orf74*
orf73
pabB
pabA
tnp2*
orf68
tnpA*
tnpB*
mobA EII*
matR
mobA EI
orf7
orf9
16S RNA*
orf11
orf13
orf15*
pcp
orf19
orf20
corA
orf22*
tnp1
orf25
orf27
orf29
orf30
tnpA1*
tnpB1*
dld
tnp2
prtP
prtM
tnp3
tnp1
tnpB2*
cspD
cspC
cspP
tnp4
tRNA-Met
umuC*
orf52
repA*
tnpA*
tnpB*
tnpB*
tnpA*
IcoR
IcoS
lgt
usp45
IcoC
70000
10000
60000
20000
50000
30000
40000

where nicking and unwinding begins, followed by 22-bp direct repeats (DR; iterons) that are believed to interact directly with the RepB protein to initiate replication (32, 57). The number of DR is typically 3.5, as we found in four replicons, but variations are known to occur, as is also illustrated by the finding of 5.5 and 1.5 iteron-repeats upstream of *repB* of pSK11P and *repB2* of pSK11L, respectively (Fig. 3).

Although the 22-bp DR are highly homologous, they display sequence specificity in each of the six replicons found. Presumably, these iteron regions have evolved sufficiently to allow specific binding by the corresponding RepB proteins. Moreover, it has been suggested that, despite the high homology between their iterons, the interaction between the replication origin and the replication protein RepB is highly specific, in many cases rendering the plasmids compatible (34). Indeed, four functional and compatible theta-type replicons were identified on pNZ4000 (67, 68), and at least four of the six replicons on the four SK11 plasmids were found to be compatible with each other.

The iteron regions are followed by two characteristic short inverted repeats, IRa and IRb (Fig. 3). The IRa repeat of 6 to 11 bp overlaps with the last half iteron-repeat sequence and with the position of the *repB* promoter −35 box. The IRa is variable in the six replicons and is possibly involved in the autoregulation of transcription of the corresponding *repB* gene; binding of the RepB protein of pCI305 to a repeat similar to IRa has been observed previously (31). The IRb repeat of 13 to 15 bp is found between the extended promoter −10 site (consensus sequence TGNTATAAT) and the ribosome-binding site AAGGAG. The IRb repeat and the AT-rich region containing the 22-bp DR are probably required for recognition by host-encoded functions involved in replication (34, 41, 57) such as DnaA to unwind the AT-rich origin region. This IRb repeat is not present in the upstream region of *repB1* of pSK11L due to a deletion of about 47 bp (Fig. 3).

The downstream region of the *repB* gene is often conserved in theta-type replicons as *repB-orfX-hsdS*, for which translational coupling is presumed (11). The *orfX* gene is usually linked to the *repB* gene by a small overlap (one to two codons). OrfX has an N-terminal helix-turn-helix motif, characteristic for binding to DNA. The absence of o*rfX* on pUCL287 has no effect on its replication capability but did raise the copy number and the stability, suggesting that OrfX could act as a controlling element in plasmid replication (4). We propose to call this downstream gene *repX*, for replication-associated protein. This *repX* gene is found on plasmids of *L. lactis*, *Lactobacillus* spp. (*L. plantarum*, *L. sakei*, and *L. casei*), *Listeria innocua*, *Leuconostoc citreum*, and *Tetragenococcus* (*Pediococcus*) *halophilus.*

On the *L. lactis* subsp. *cremoris* SK11 plasmids, the *repX* gene was found downstream of *repB* in four of six cases, while in one case, a hypothetical gene appeared to be inserted between *repB* and *repX*. Downstream of *repX*, an *hsdS* gene of the restriction-modification system is often found (11), but this gene appears to be absent on the SK11 plasmids (for a possible exception, see pSK11B). On plasmid pSK11P, the organization of replication genes is *repB-repX-repC*, similar to that found on lactococcal plasmids pHP003, pND324, and pIL105, although the predicted RepB and RepX proteins are most homologous to those of pUCL22. On pSK11B, two different replicons with a set of *repB-repX* genes are found in opposite orientation from two origins of replication, separated by an IS*6* element. The *repB-repX* genes are clearly homologous to lactococcal plasmid proteins (82 to 96% amino acid identity). However, one of these sets is flanked on both sides by complete IS*1216E* elements (member of the IS*6* family) encoding transposases with 100% identity to those on *Enterococcus faecalis* plasmids. On pSK11A, the *repB* gene and adjacent gene are almost identical to those on pND306 of *L. lactis* M71.

**Mobilization.** Although conjugal transfer of the lactose and proteinase plasmids has been described, the SK11 plasmids do not appear to encode the complete set of canonical proteins that are required for mobilization and transfer, suggesting that none of the plasmids is self-transmissible. Plasmid pSK11A is predicted to encode only a mobilization protein, MobC (122 residues), which is 96% identical to that of pNZ4000 and belongs to the relaxase family (Pfam05713) of proteins that are involved in strand separation. Plasmid pSK11A also contains a more extensive 1.3-kb conserved region upstream of *repB* (bases 9,514 to 10,372 and 1 to 443) which is highly similar to the *repB* upstream region of many lactococcal plasmids, including pNZ4000 (67). This region can be divided into three segments. The first segment contains the putative origin of transfer (oriT) with conservation of the postulated *nic* site (hexamer CTTGCA just downstream of a conserved pair of inverted repeats). The next segment of 400 bases is also highly conserved between the lactococcal plasmids that have a *repB* gene downstream of an oriT sequence. This region has not been annotated with any particular function, but it contains at least three 8-bp perfect inverted repeats. The central spacer between the ends of these repeats is quite large (60 to 100 bp). Their distribution along this section is not uniform, with two regions overlapping each other. The final segment is a region with AT-rich regions and direct repeats (Fig. 3).

Plasmid pSK11L is predicted to encode a conjugal transfer protein (nicking enzyme) of the MobA-MobL family (Pfam03389), with the highest homology (32%) to a TraA protein encoded on *L. lactis* DPC3147 plasmid pMRC01. In contrast to pMRC01, the pSK11L *mobA* gene is not part of a large mobilization gene cluster.

Plasmid pSK11P contains a putative *mobA* gene that is interrupted by a group II intron, encoding the maturase MatR, a multifunctional protein that promotes group II intron splicing and mobility by acting on both RNA and DNA. The group II intron is found directly downstream of the replication genes *repB*, *repX*, and *repC*. A group II intron is also found in the

FIG. 2. Physical and genetic maps of plasmids pSK11L and pSK11P of *L. lactis* subsp. *cremoris* SK11. Arrows indicate positions, directions, and names of genes encoding proteins of predicted function (black), conserved hypothetical proteins (gray), and hypothetical proteins (white). Asterisks indicate pseudogenes. Encircled roman numerals refer to modules, which are described in the text. Inner circles show the nucleotide numbering and IS elements (black boxes).

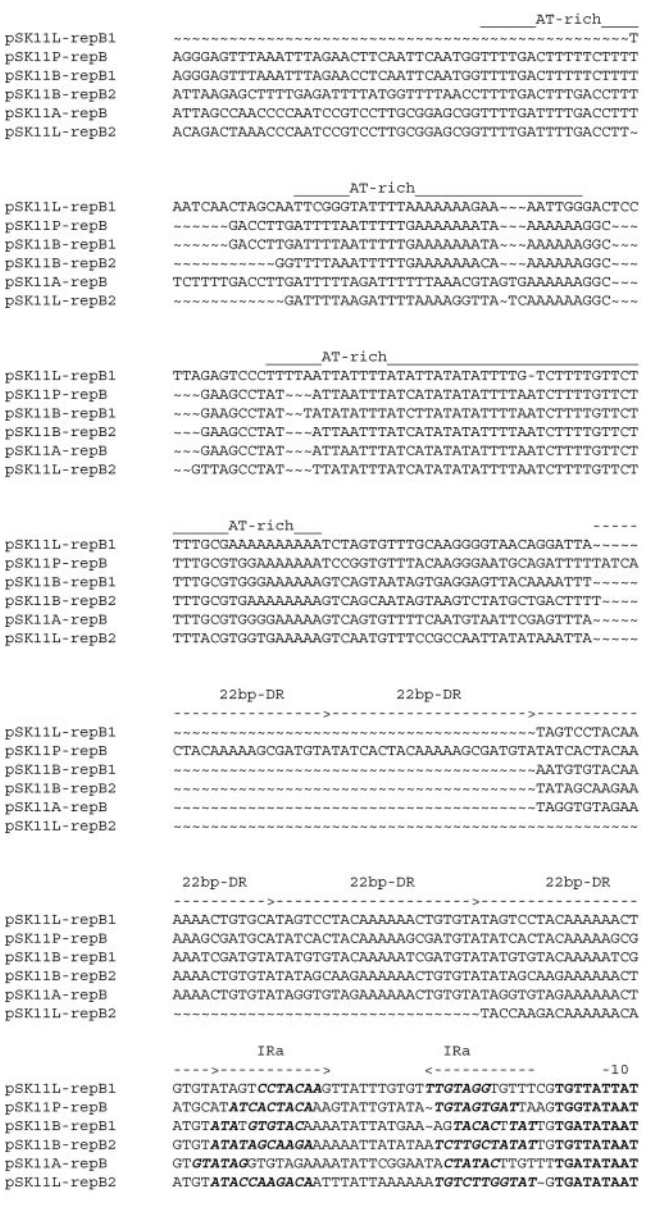



FIG. 3. Multiple sequence alignment of upstream region of *repB* genes, ending at the ATG start codon of *repB*. Indicated are the AT-rich region, the 22-bp direct repeats (22bp-DR), and, in boldface, the two inverted repeat regions (IRa, IRb), extended promoter −10 site (consensus TGNTATAAT), and RBS (nucleotide sequence AAGGAG).

chromosomally encoded sex factor of *L. lactis* 712 that controls conjugation and interrupts a mobilization gene, *mobA*, that was demonstrated to play an active role in conjugation (59). A group II intron also occurs on pAH82 of *L. lactis* adjacent to replication genes (53) and on pRS01 (49) and pCI2001 (AF179847) (K. Kearney and G. F. Fitzgerald, unpublished data). However, it is questionable whether MobA of pSK11P can be active after removal of the intron, since the *mobA* exon *EII* appears to contain a frameshift.

**IS elements.** IS elements have been shown to play an important role in lactococcal gene transfer, and they are often associated with industrially important traits (54). Various types of IS elements with their encoded transposases are known to occur in *L. lactis* both on plasmids and on the chromosome (8, 15). Numerous IS elements and fragments thereof were found on the four SK11 plasmids, belonging to three main IS families and their subfamilies, as summarized in Table 1. The pSK11P plasmid contains two copies of an IS element of 1,400 to 1,401 nucleotides (at positions 22434 to 23833 and positions 37651 to 39051), referred to as IS-*LLx*, which is not in the IS Database (www-is.biotoul.fr) but presumably belongs to the IS*3* family since the encoded transposases belong to the same COGs (Table 1). All 11 IS*6* elements and 4 of the 5 IS*982A* elements and their transposase genes are complete, and hence, they are potentially functional, in contrast to all other IS elements which appear to be nonfunctional since they are incomplete or encode truncated or frameshift-interrupted transposases (Tables 1 and 2; see Table S2 in the supplemental material).

The IS elements appear to divide the plasmids into modules or cassettes, which in some cases could be composite transposons (Fig. 1 and 2, modules indicated by roman numerals). Sequence homology analysis suggests that these modules are often from one origin, e.g., from other *L. lactis* plasmids or from other species. Many modules encode gene clusters for one function as described in more detail below.

**Plasmid-encoded functions and their origin.** Of the four plasmids with a total sequence of approximately 147 kb, we identified and annotated 156 putative genes or fragments thereof, as summarized in Table 2 and in Table S2 in the supplemental material. The majority (69%) of these genes displayed the highest homology to genes located on other *L. lactis* plasmids, but a significant amount (12%) displayed the highest homology to genes of other, mainly gram-positive bacteria, suggesting plasmid-mediated horizontal gene transfer between species. Putative functions could be assigned to 76% of the gene products, and the most interesting are described below.

Nearly 20% of the putative genes were pseudogenes, containing frameshifts, stop codons, or truncations; this occurred primarily in the transposases of IS elements (Table 2). Most of the conserved hypothetical proteins (15 of 19) had homologs only on plasmids (see Table S2 in the supplemental material). Nineteen ORFs encoded hypothetical proteins, generally smaller than 100 residues, with no homologs in databases. Some of these may have been overpredicted and may not have been protein-encoding genes. However, at least 29 of 37 (conserved) hypothetical genes were preceded by consensus ribosome-binding sites with high complementarity to the 3′ end of the *L. lactis* 16S rRNA that is expected for translated genes.

TABLE 1. IS elements

| Family | Name[b] | Size (bp)[c] | No. of IS elements[a] | | | | COG for transposase |
|---|---|---|---|---|---|---|---|
| | | | pSK11P | pSK11L | pSK11B | pSK11A | |
| IS*6* | IS*S1N* | 808 | 5 | | | | COG3316 |
| | IS*S1X* | 808 | | 1 | | 1 | COG3316 |
| | IS*1216E* | 809 | | | 2 | | COG3316 |
| | IS*1297* | 808 | | 2 | | | COG3316 |
| IS*3* | IS-*LL6* | 1,254 | 1 (2) | | | | For *tnpB*, COG2801; for *tnpA*, COG2963 |
| | IS*981* | 1,224 | (1) | (1) | | (1) | For *tnpB*, COG2801; for *tnpA*, COG2963 |
| | IS-*LLx* | 1,401 | 2 | | | | For *tnpB*, COG2801; for *tnpA*, COG2963 |
| IS*982* | IS*982A* | 1,003 | 2 | 3 | | | COG3039 |
| | IS*982B* | 1,003 | | | 1 | | COG3039 |
| | IS*982C* | 1,003 | 1 | | | | COG3039 |

[a] Values in parentheses indicate numbers of incomplete IS elements.
[b] For nomenclature, see www.is.biotoul.fr.
[c] Size of complete IS element.

**Plasmid pSK11A.** Plasmid pSK11A has mostly high homology to regions of other *L. lactis* plasmids (see Table S2 in the supplemental material). *orf4* could encode a serine/threonine protein phosphatase family protein, which is nearly 100% identical to that of pAH82 of *L. lactis* subsp. *lactis* bv. *diacetylactis* DPC220. The most conserved regions in this superfamily (Pfam00149.11) center around metal-chelating residues, which are all present in the Orf4 protein. This metal-dependent phosphoesterase could be involved in signal transduction and/or regulation. *orf12* is predicted to code for a response regulator of the LytR/AlgR family, with 30 to 35% sequence identity to regulators of several streptococci, staphylococci, and lactobacilli. *orf13*, predicted to encode a transmembrane protein, appears to be translationally coupled to *orf12*, suggesting that their gene products function together. Moreover, the *orf12-orf13* gene pair is conserved in various gram-positive bacteria. Since the Orf13 protein family (see Prodom database entry PD459887) contains conserved His and Asn residues, typical of histidine kinase sensor proteins, we speculated that the Orf13 protein acts as the cognate transmembrane sensor protein of response regulator Orf12, providing a two-component signal transduction system.

**Plasmid pSK11B.** This high-copy-number plasmid is quite complex and could be a cointegrate from plasmids of several different species (see Table S2 in the supplemental material).

TABLE 2. Summary of plasmid statistics, putative genes, and encoded functions

| Property | Value for: | | | |
|---|---|---|---|---|
| | pSK11A | pSK11B | pSK11L | pSK11P |
| Size (bp) | 10,372 | 13,332 | 47,165 | 75,184 |
| G+C content (%) | 31.0 | 34.3 | 34.8 | 35.5 |
| Total no. of: | | | | |
| ORFs | 14 | 15 | 45 | 82 |
| Pseudogenes | 2 | 2 | 4 | 23 |
| Transposases[a] | 4 (2) | 3 (1) | 7 (2) | 19 (12) |
| Replication proteins | 2 | 6 | 3 | 6 |
| Conserved hypotheticals[b] | 1 (1) | 1 (1) | 3 (3) | 14 (10) |
| Hypotheticals | 3 | 3 | 4 | 9 |

[a] Values in parentheses indicate numbers of pseudogenes.
[b] Values in parentheses indicate numbers of hypothetical genes conserved on plasmids.

Directly downstream of the second *repB-repX* replicon is a 3-kb region with 67% identity at the nucleotide level to *Streptococcus thermophilus* plasmid pSMQ173b (accession no. AY312235), encoding three proteins (Orf3, Orf4, and Orf5) with 52 to 72% identity to proteins encoded on this streptococcal plasmid but lacking known homologs in *L. lactis*.

The *aldC* gene encodes a protein of 221 amino acids with 92% identity to an α-acetolactate decarboxylase of the *L. lactis* IL-1403 chromosome. A second, chromosomally encoded AldC of strain SK11 has 98% identity to a second chromosomally encoded AldC of *L. lactis* IL1403. This enzyme converts α-acetolactate to acetoin and has potential significance in the dairy environment in the regulation of branched-chain amino acid biosynthesis (50). Alternatively, since acetoin can be further reduced to butanediol, the activity of α-acetolactate decarboxylase may contribute to the formation of an additional electron sink. The plasmid-encoded AldC of strain SK11 could provide a gene dosage enhancement of such a function, particularly since an IS element upstream of the *aldC* gene may boost transcription (10).

**Plasmid pSK11L (lactose plasmid).** The complete sequence determination of this plasmid as described here provides the molecular basis for several previously described phenotypic traits (29, 38, 63, 74). The calculated restriction map of our pSK11L (Fig. 2) is virtually identical to that of the 47.3-kb $Lac^+$ plasmid pSK11 determined by Horng et al. (38). Small fragments of pSK11L have been sequenced previously (38, 76), and the *repB1* gene and its upstream region are found to be 100% identical to the 1.5-kb fragment sequenced by Horng et al. (38).

Plasmid pSK11L is predicted to encode seven transposases of IS elements (Table 1) and two putative site-specific recombinases, PinR1 (*orf36*) and PinR2 (*orf43*), which have low homology (38% identity) to each other. It is known that site-specific recombinases are involved in the resolution of plasmid multimers (6). These elements appear to divide the plasmid into eight modules that in some cases have characteristics of composite transposons, and some are briefly outlined below (Fig. 2).

Module I is predicted to encode replication and mobilization proteins. *orf2*, the first gene downstream of *repB1*, encodes a

protein with low homology (34% identity) to a plasmid-partitioning ATPase of *L. lactis* plasmid pCI2000. This module also encodes the other replication initiator, RepB2, and MobA, a conjugal transfer protein (nicking enzyme).

Module II is flanked by two IS*6* elements and hence resembles a composite transposon. It may encode only two proteins, both with highest homology to the *Lactobacillus plantarum* WCFS1 chromosome (43). The *mntH* gene (*orf13*) is predicted to encode an $Mn^{2+}/Fe^{2+}$ transporter of the natural resistance-associated macrophage protein (NRAMP) family. In *L. plantarum*, the *mntH* genes are induced under $Mn^{2+}$ limitation (35). The advantage of having this plasmid-borne gene may be a gene dosage effect, since there is another highly homologous MntH protein encoded on the *L. lactis* SK11 chromosome. The *usp* gene (*orf14*) encodes the universal stress protein UspA, which is a cytoplasmic protein that enhances the rate of survival during prolonged exposure to various stress conditions (22).

Module V contains the *lacA*, *lacB*, *lacC*, *lacD*, *lacF*, *lacE*, *lacG*, and *lacX* genes of the lactose utilization operon and the divergently transcribed *lacR* gene encoding the lactose phosphotransferase system repressor (21, 69, 70). These plasmid-borne genes are responsible for the lactose-fermenting ability ($Lac^+$ phenotype) of strains carrying this plasmid. The encoded proteins of this module are 98 to 100% identical to those on the *L. lactis* subsp. *lactis* MG1820 chromosome (21).

Module VI carries a *pepF1*-like gene encoding the well-studied oligopeptidase F. Nearly identical proteins are encoded on the chromosome of *L. lactis* IL1403 and on the lactose plasmid pLP763 of *L. lactis* subsp. *cremoris* NCDO763 (51, 52). A gene encoding oligoendopeptidase PepF2, with about 80% amino acid sequence identity to PepF1, is carried on the chromosome of NCDO763 (52) but also on the chromosome of *L. lactis* subsp. *cremoris* SK11 (http://genome.ornl.gov/microbial/lcre/) The PepF proteins are part of a complex proteolytic system of *L. lactis* which is essential for its growth in milk (44). The advantage of having multiple genes for certain proteolytic functions could be in a more efficient utilization of peptides derived from milk proteins.

Module VII also resembles a composite transposon since it is flanked by an IS*6* element and a site-specific recombinase gene, *pinR2*. This module appears to contain a single *cstA*-like gene, which could be involved in peptide utilization during carbon starvation (56). The relevance of this plasmid-borne gene is not clear, since another CstA with 96% sequence identity is encoded on the SK11 chromosome, but it could be simply to enhance gene dosage.

Module VIII contains the *pepO* gene encoding neutral endopeptidase and the complete oligopeptide permease (Opp) system consisting of the *oppDFBCA* gene cluster. It has been demonstrated that this *opp* gene cluster is required, in addition to the *prtP*/*prtM* system (see pSK11P), for utilization of the milk caseins as a nitrogen source and for a fast milk coagulation phenotype, $Fmc^+$ (45, 65, 74, 75). A 7.4-kb fragment of the lactose plasmid pSK11L has previously been cloned and sequenced and found to contain this Opp system and an incomplete *pepO* gene (74). Growth studies of transformants showed that this *opp* cluster is functional and complemented the PrtP/PrtM system to achieve the $Fmc^+$ phenotype (74). The sequence of this fragment on pSK11L is essentially identical to the 7.4-kb fragment sequenced before and over 99% identical to the chromosomal Opp system from *L. lactis* SSL135 (65). The plasmid-borne *pepO* gene is now shown to be complete, but it is not unique, since another copy of *pepO* with very high homology (>90% identity) is found on the SK11 chromosome. However, the plasmid-encoded combination of PepO with the Opp system seems to be unique for this *L. lactis* strain and may be beneficial for the optimal uptake and utilization of oligopeptides generated by the extracellular PrtP serine proteinase.

**Plasmid pSK11P (proteinase plasmid).** The proteinase plasmid of *L. lactis* subsp. *cremoris* SK11 and some of its related phenotypic traits have also been described in several previous papers (17, 19, 20, 47, 60, 61). Plasmid pSK11P is found to contain at least 14 IS elements, belonging to three different families, which divide the plasmid into 12 different modules, ranging considerably in size (Fig. 2). Several IS elements are linked to each other, and many pseudogenes have been found, which presumably reflects the history of many recombination events within this plasmid.

Module I is quite large (~17 kb) and carries genes with a variety of functions, including genes encoding replication and mobilization proteins (see above), truncated integrase/recombinase genes (*orf15C*, *orf15N*, and *orf18*), and a 630-bp fragment of a 16S RNA with 99.7% identity to *L. lactis* subsp. *lactis* IL1403.

A predicted *pcp* gene encoding an N-terminal pyroglutamyl peptidase was present that was nearly identical to that of *L. lactis* MG1363. This component of the proteolytic system of *L. lactis* contributes to the N-terminal degradation of peptides. Furthermore, a *corA* gene encoding a CorA family transporter, which is expected to be the dominant $Mg^{2+}$ uptake system in bacteria, was found (62), as experimentally determined for *E. coli* and *Salmonella enterica* serovar Typhimurium. This CorA transporter of pSK11P has the highest homology by far to an *Oenococcus oeni* protein (91% identity), while homology to transporters of *Lactococcus lactis* and *Bacillus* is less than 30%. Another *corA* gene with low homology is carried on the chromosome of *L. lactis* SK11. Within species, paralogs are generally not closely related, suggesting that differential transport functions under variable conditions (39). The presence of multiple CorA-like transporters could be related to their optimal functioning under differential conditions, including specific growth-limiting conditions.

Module III has only a *dld* gene, predicted to encode a D-lactate dehydrogenase (D-LDH) (EC 1.1.1.28) of 559 residues with a putative flavin adenine dinucleotide (FAD)-binding site (25). This is a very unusual protein to find in *L. lactis* because the best homologs (about 50% sequence identity, same size) are found in gram-negative bacteria such as *E. coli* and *Shigella* and *Salmonella* spp. and there are no known homologs in gram-positive bacteria. This appears to be a clear case of horizontal gene transfer from a gram-negative organism. Gram-positive bacteria generally have a D-LDH of 320 to 350 residues belonging to an unrelated family, which is NADH dependent and is involved in D-lactate production during fermentation on sugars under strictly anaerobic conditions (12). Neither type of D-LDH has been found before in *L. lactis*. This NADH-independent D-LDH in *L. lactis* SK11 could function in D-lactate utilization under aerobic conditions and could

have been acquired and maintained in this strain during milk fermentation if D-lactate is produced by other lactic acid bacteria. Consumption of D-lactate could contribute to a pH reduction or could be converted to acetate with ATP production, both contributing to enhanced survival.

Module IV contains the divergently transcribed genes *prtP* and *prtM* encoding the proteinase PrtP and its maturation protein, PrtM/PrsA (foldase; EC 5.2.1.8). The sequence of this region of pSK11P has been described earlier (19, 71, 72). The *prtP-prtM* genes are bounded by IS*6* elements (subfamily IS*S1N*) encoding a transposase, previously also called the ORF-N1 protein (36). Between the *prtP* and *prtM* genes is located a highly AT-rich 0.3-kb region with overlapping, divergently oriented promoters which are peptide inducible (47).

Module VII contains a segment of 1.6 kb with 94% nucleotide identity to a region with the *cspC* and *cspD* genes, encoding cold shock proteins of 66 residues, located on the *L. lactis* MG1363 chromosome (73). This segment includes an upstream region of *cspC* in which another putative *csp* gene (*cspP*) is located. The encoded protein of 66 amino acids has some homology to CspC and CspD, so it is a putative cold shock protein. Both *cspC* and *cspD* have good RBS and −10/−35 promoter sequences, but *cspP* lacks a good promoter. This region is also present in both sequenced genomes of *S. thermophilus* (7).

Interestingly, a tRNA gene with the anticodon CAT (74 bp) is found upstream of the *csp* genes. tRNA-Met is important for initiation of translation, and a higher concentration of tRNA-Met may be needed for highly expressed genes. Codon analysis shows that the *cspD* gene has a very high CAI, the same as for highly expressed ribosomal protein genes; *cspC* has an average CAI, and *cspP* has a low CAI. One more *cspC* gene is found on the SK11 chromosome. Csp proteins could be of importance for the survival of starter cultures at low temperatures. The advantage of having additional plasmid-encoded cold-shock proteins could be an enhanced dosage effect, particularly if these genes can be highly expressed.

Module IX is another putative composite transposon bound by IS-*LL6* elements. It is predicted to contain a complete copper resistance gene cluster, *lcoRSABC*, that is nearly identical to corresponding genes on plasmid pND306 from *L. lactis* subsp. *lactis* LL58-1 (40, 46). This cluster contains a two-component regulatory system (LcoR and LcoS) that is involved in the copper-inducible transcription of the downstream operon *lcoABC*. LcoA is a typical lipoprotein signal peptidase/transferase which has a homolog on the SK11 chromosome. LcoC is a member of the multicopper oxidase family and a putative copper resistance protein; it has a typical lipoprotein signal peptide, so it is probably translocated across the cell membrane to the outside of the cell and covalently anchored by LcoA to membrane lipids. LcoB is a well-known and abundantly secreted protein of *L. lactis* of unknown function and was previously called Usp45 (66). It contains a C-terminal cysteine, histidine-dependent amidohydrolase/peptidase (CHAP) domain (PF05257) with conserved Cys and His residues and could be an amidase involved in peptidoglycan metabolism. Experimental studies have shown that all three genes of the *lcoABC* operon are required for copper resistance and that this operon confers copper resistance by reducing the intracellular accumulation of copper ions in *L. lactis* (40, 46).

Module X is predicted to encode only a single protein of 345 residues (Orf68) which has no known homologs in *L. lactis*. The best homology (highest at 33% identity) is to proteins of about 400 residues from *Methanococcus*, *Leuconostoc*, and *Listeria* spp. This family of proteins of unknown function shows homology only in their C-terminal domains of about 180 residues (ProDom Family PD464329). This domain has a highly conserved motif, GDGHGYDI, possibly containing conserved catalytic residues of an enzyme family of unknown function. Family members are not very homologous in sequences outside of this motif; such low homology is typical for esterase and lipase families (2, 13, 24, 30).

Module XI is predicted to encode the two subunits PabA and PabB of the para-aminobenzoate (PABA) synthase complex (EC 6.3.5.8), an enzyme in the pathway of chorismate to PABA. These two proteins are more than 95% identical to those on the *L. lactis* IL1403 chromosome but appear to be absent on the *L. lactis* SK11 chromosome. Knockout studies of the chromosomally encoded *pabA* and *pabB* genes of *L. lactis* MG1363 led to a 100-fold reduction of folate biosynthesis, which can be fully complemented with plasmid-borne *pabAB* (A. Wegkamp, personal communication). This suggests that the *pabA* and *pabB* genes of plasmid pSK11P provide *L. lactis* SK11 with the ability to synthesize folate, an essential cofactor in single-carbon metabolism and may contribute to plasmid stabilization under conditions of PABA requirement.

Module XI also encodes a putative oxalate-formate antiporter, a member of the major facilitator superfamily of permeases, which could be involved in a proton-motive metabolic cycle to drive ATP synthesis.

Module XII carries a *csc* gene cluster encoding cell surface proteins, which is also found in a subgroup of gram-positive bacteria (R. J. Siezen, J. Boekhorst, L. Muscariello, D. Molenaar, B. Renckens, and M. Kleerebezem, unpublished data). All encoded proteins have a regular signal peptide for secretion by the Sec-dependent pathway. The Csc proteins are postulated to form an extracellular, cell-bound complex that could be involved in the degradation of complex polysaccharides (R. J. Siezen et al., unpublished data). This gene cluster on pSK11P appears to be interrupted by an IS*982* element, possibly rendering it nonfunctional.

**Adaptation to dairy environment.** Table 3 provides a summary of genes and encoded proteins found on the *L. lactis* SK11 plasmids which reflect adaptations to the dairy environment, from which this strain was originally isolated. These adaptations are related to growth on milk components such as lactose and caseins but also to the facilitated uptake of cations such as $Mg^{2+}$, $Mn^{2+}$, and possibly $Fe^{2+}$. The plasmid-borne *pabA* and *pabB* genes are predicted to be essential for folate biosynthesis in this strain, which in turn is required as a cofactor for single-carbon transfer reactions. A competitive advantage could also be obtained by using the putative FAD-dependent D-lactate dehydrogenase and oxalate:formate antiporter for enhanced ATP synthesis, while the activity of the putative α-acetolactate decarboxylase may contribute to the formation of an additional electron sink. Strain robustness could certainly be enhanced by various encoded stress response proteins, such as cold shock proteins, universal stress protein, carbon starvation protein, and a copper resistance system. A substantial

TABLE 3. Predicted adaptation to dairy environment

| Plasmid and function | Gene(s)[a] | Encoded protein(s) | Predicted function/role(s) |
|---|---|---|---|
| Carbohydrate utilization | | | |
| pSK11L | *lacRABCDFEGX* | Lactose operon | Growth on lactose |
| pSK11P | *cscABC*[a] | Cell surface proteins | Degradation of complex polysaccharides |
| Peptide utilization | | | |
| pSK11P | *prtP-prtM* | Cell envelope proteinase, maturase | Growth on casein; protein breakdown to peptides |
| pSK11L | *oppDFBCA* | Oligopeptide transport | Growth on peptides; peptide uptake |
| pSK11L | *pepO* | Neutral endopeptidase | Growth on peptides; peptide breakdown |
| pSK11L | *pepF* | Oligoendopeptidase F | Growth on peptides; peptide breakdown |
| pSK11P | *pcp*[a] | Pyrrolidone-carboxylate peptidase | Growth on peptides; peptide breakdown |
| Cofactor and ion requirements | | | |
| pSK11P | *pabAB*[a] | *para*-Aminobenzoate synthase | Folate precursor biosynthesis |
| pSK11P | *corA* | $Mg^{2+}$ transporter | Enhancement of $Mg^{2+}$ uptake |
| pSK11L | *mntH*[a] | $Mn^{2+}/Fe^{2+}$ transporter, NRAMP family | Enhancement of $Mn^{2+}$ uptake |
| Energy metabolism | | | |
| pSK11P | *dld* | D-Lactate dehydrogenase, FAD dependent | D-Lactate to acetate and ATP conversion |
| pSK11P | *oxfA* | Oxalate:formate antiporter | ATP synthesis |
| pSK11B | *aldC*[a] | α-Acetolactate decarboxylase | Electron sink; α-acetolactate degradation |
| Stress response | | | |
| pSK11L | *cstA*[a] | Carbon starvation protein | Peptide utilization during carbon starvation |
| pSK11L | *usp*[a] | Universal stress protein | General stress response |
| pSK11P | *cspCDP*[a] | Cold shock proteins | Cold shock stress response |
| pSK11P | *lcoRSABC* | Copper resistance proteins, regulators | Copper resistance system |

[a] Genes never found before on plasmids of *L. lactis*.

number of these adaptation genes have not been described before on *L. lactis* plasmids (Table 3).

No indication was found for phage resistance genes on these four plasmids, suggesting that the ~60-kb plasmid pSK112, which confers phage resistance to *L. lactis* SK11 (17, 63), was not present in our isolate. However, the *L. lactis* subsp. *cremoris* strain NIZO B697 used here is known to carry an inducible prophage (F. Kingma, unpublished data), which could confer immunity to related phages.

**Horizontal gene transfer.** Table 4 lists examples of known-function genes not found before in *L. lactis* and the bacterial proteins to which these have the corresponding highest homology, in some cases with >90% identity. Some of these genes also have G+C contents that are significantly different from the average of about 35% for *L. lactis* plasmids. These are the most likely candidate genes to have been acquired by horizontal transfer, possibly through the exchange of plasmid DNA between enterococci, streptococci, and lactobacilli in a dairy environment.

**Concluding remarks.** By definition, plasmids are extrachromosomal elements that code for autonomous replication and, in many cases, stabilization. In addition, they also can carry mobilization functions that allow their promiscuous dissemination. The SK11 plasmid complement is no exception, and the functional redundancy on the various plasmids suggests that there is no synergy. This is confirmed by the observation that pSK11P can be cured without effect on the other plasmids, although attempts to create plasmid-free derivatives of SK11 have not been successful (18).

In addition to these expected plasmid-located functions,

TABLE 4. Novel genes not found before in *L. lactis*

| Plasmid | Gene(s) | Function(s) | Best homolog found in: | % Identity | G+C content (%) |
|---|---|---|---|---|---|
| pSK11A | *orf12-orf13* | Response regulator and sensor protein: signal transduction system (putative) | *Lactobacillus gasseri* | 31 | 23.2–27.9 |
| pSK11B | *orf3-orf4-orf5* | DNA segregation ATPase, replication initiation factor, hypothetical | *Streptococcus thermophilus* | 52–72 | 35.1–42.3 |
| pSK11B | *tnpIS6* | Transposase | *Enterococcus faecalis* | 100 | 37.6 |
| pSK11L | *mntH* | $Mn^{2+}/Fe^{2+}$ transporter, NRAMP family | *Lactobacillus plantarum* | 77 | 37.1 |
| pSK11L | *orf16* | Aminoglycoside 3-*N*-acetyltransferase | *Bacillus subtilis* | 39 | 31.4 |
| pSK11P | *dld* | D-Lactate dehydrogenase, FAD dependent | *Escherichia coli* | 52 | 38.9 |
| pSK11P | *orf76* | VanZ-family protein | *Streptococcus thermophilus* | 98 | 34.9 |
| pSK11P | *orf29* | Amidohydrolase | *Enterococcus faecium* | 75 | 40.6 |
| pSK11P | *corA* | Mg transporter, CorA family | *Oenococcus oeni* | 91[a] | 32.1 |

[a] Best homology to a *L. lactis* CorA protein was less than 30%.

three main sets of sequences have been found. First, a great number of IS elements are present on the SK11 plasmids that may add to the mobility of the plasmid-located genes via homologous or site-specific recombination in which their transposases may play a role. This also holds for the observed group II intron that has often been found to be located on mobile elements on *L. lactis*. Second, the plasmids appear to encode a variety of functional properties that are providing a competitive advantage on the plasmid-carrying cells. In addition to the well-known proteins involved in lactose and protein metabolism, these include proteins involved in cofactor biosynthesis, ion uptake, energy metabolism, and stress response. Finally, there seems to be a high density of genes on these plasmids for which we presently have no known function but that are also found on mobile elements in other bacteria.

The complete sequences of these four plasmids, totaling 150 kb of extrachromosomal DNA, provide new insight into the repertoire of plasmid-encoded functions in *L. lactis* and will not only provide leads for the targeted determination of gene functions by experimental methods but also be an important reference for comparative analysis with other lactococcal plasmids.

## ACKNOWLEDGMENTS

We thank Douwe Molenaar for analysis of tRNA-Met and CAI, Arno Wegkamp for analysis of *pabA* and *pabB* genes, and Robert Kerkhoven for the calculation of terminators.

## REFERENCES

1. **Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman.** 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. **25:**3389–3402.
2. **Anthonsen, H. W., A. Baptista, F. Drablos, P. Martel, S. B. Petersen, M. Sebastiao, and L. Vaz.** 1995. Lipases and esterases: a review of their sequences, structure and evolution. Biotechnol. Annu. Rev. **1:**315–371.
3. **Bateman, A., L. Coin, R. Durbin, R. D. Finn, V. Hollich, S. Griffiths-Jones, A. Khanna, M. Marshall, S. Moxon, E. L. Sonnhammer, D. J. Studholme, C. Yeats, and S. R. Eddy.** 2004. The Pfam protein families database. Nucleic Acids Res. **32:**D138–D141.
4. **Benachour, A., J. Frere, S. Flahaut, G. Novel, and Y. Auffray.** 1997. Molecular analysis of the replication region of the theta-replicating plasmid pUCL287 from Tetragenococcus (Pediococcus) halophilus ATCC33315. Mol. Gen. Genet. **255:**504–513.
5. **Berriman, M., and K. Rutherford.** 2003. Viewing and annotating sequence data with Artemis. Brief. Bioinform. **4:**124–132.
6. **Blake, J. A., N. Ganguly, and D. J. Sherratt.** 1997. DNA sequence of recombinase-binding sites can determine Xer site-specific recombination outcome. Mol. Microbiol. **23:**387–398.
7. **Bolotin, A., B. Quinquis, P. Renault, A. Sorokin, S. D. Ehrlich, S. Kulakauskas, A. Lapidus, E. Goltsman, M. Mazur, G. D. Pusch, M. Fonstein, R. Overbeek, N. Kyprides, B. Purnelle, D. Prozzi, K. Ngui, D. Masuy, F. Hancy, S. Burteau, M. Boutry, J. Delcour, A. Goffeau, and P. Hols.** 2004. Complete sequence and comparative genome analysis of the dairy bacterium Streptococcus thermophilus. Nat. Biotechnol. **22:**1554–1558.
8. **Bolotin, A., P. Wincker, S. Mauger, O. Jaillon, K. Malarme, J. Weissenbach, S. D. Ehrlich, and A. Sorokin.** 2001. The complete genome sequence of the lactic acid bacterium Lactococcus lactis ssp. lactis IL1403. Genome Res. **11:**731–753.
9. **Bonfield, J. K., K. Smith, and R. Staden.** 1995. A new DNA sequence assembly program. Nucleic Acids Res. **23:**4992–4999.
10. **Bongers, R. S., M. H. N. Hoefnagel, M. J. C. Starrenburg, M. A. J. Siemerink, J. G. A. Arends, J. Hugenholtz, and M. Kleerebezem.** 2003. IS*981*-mediated adaptive evolution recovers lactate production by *ldhB* transcription activation in a lactate dehydrogenase-deficient strain of *Lactococcus lactis*. J. Bacteriol. **185:**4499–4507.
11. **Boucher, I., E. Emond, M. Parrot, and S. Moineau.** 2001. DNA sequence analysis of three Lactococcus lactis plasmids encoding phage resistance mechanisms. J. Dairy Sci. **84:**1610–1620.
12. **Bunch, P. K., F. Mat-Jan, N. Lee, and D. P. Clark.** 1997. The ldhA gene encoding the fermentative lactate dehydrogenase of Escherichia coli. Microbiology **143:**187–195.
13. **Cygler, M., J. D. Schrag, J. L. Sussman, M. Harel, I. Silman, M. K. Gentry, and B. P. Doctor.** 1993. Relationship between sequence conservation and three-dimensional structure in a large family of esterases, lipases, and related proteins. Protein Sci. **2:**366–382.
14. **Daly, C.** 1983. The use of mesophilic cultures in the dairy industry. Antonie Leeuwenhoek **49:**297–312.
15. **Davidson, B. E., N. Kordias, M. Dobos, and A. J. Hillier.** 1996. Genomic organization of lactic acid bacteria. Antonie Leeuwenhoek **70:**161–183.
16. **del Solar, G., R. Giraldo, M. J. Ruiz-Echevarria, M. Espinosa, and R. Diaz-Orejas.** 1998. Replication and control of circular bacterial plasmids. Microbiol. Mol. Biol. Rev. **62:**434–464.
17. **de Vos, W. M., and F. L. Davies.** 1984. Plasmid DNA in lactic streptococci: bacteriophage resistance and proteinase plasmids in Streptococcus cremoris SK11, p. 201–206. *In* Third European Congress on Biotechnology, vol. III. Verlag Chemie, Munich, Germany.
18. **de Vos, W. M., H. M. Underwood, and F. L. Davies.** 1984. Plasmid encoded bacteriophage resistance in Streptococcus cremoris SK11. FEMS Microbiol. Lett. **23:**175–178.
19. **de Vos, W. M., P. Vos, H. de Haard, and I. Boerrigter.** 1989. Cloning and expression of the Lactococcus lactis subsp. cremoris SK11 gene encoding an extracellular serine proteinase. Gene **85:**169–176.
20. **de Vos, W. M., P. Vos, G. Simons, and S. David.** 1989. Gene organization and expression in mesophilic lactic acid bacteria. J. Dairy Sci. **72:**3398–3405.
21. **de Vos, W. M., I. Boerrigter, R. J. van Rooyen, B. Reiche, and W. Hengstenberg.** 1990. Characterization of the lactose-specific enzymes of the phosphotransferase system in Lactococcus lactis. J. Biol. Chem. **265:**22554–22560.
22. **Diez, A., N. Gustavsson, and T. Nystrom.** 2000. The universal stress protein A of Escherichia coli is required for resistance to DNA damaging agents and is regulated by a RecA/FtsK-dependent regulatory pathway. Mol. Microbiol. **36:**1494–1503.
23. **Dougherty, B. A., C. Hill, J. F. Weidman, D. R. Richardson, J. C. Venter, and R. P. Ross.** 1998. Sequence and analysis of the 60 kb conjugative, bacteriocin-producing plasmid pMRC01 from Lactococcus lactis DPC3147. Mol. Microbiol. **29:**1029–1038.
24. **Drablos, F., and S. B. Petersen.** 1997. Identification of conserved residues in family of esterase and lipase sequences. Methods Enzymol. **284:**28–61.
25. **Dym, O., E. A. Pratt, C. Ho, and D. Eisenberg.** 2000. The crystal structure of D-lactate dehydrogenase, a peripheral membrane respiratory enzyme. Proc. Natl. Acad. Sci. USA **97:**9413–9418.
26. **Ewing, B., and P. Green.** 1998. Base-calling of automated sequencer traces using phred. II. Error probabilities. Genome Res. **8:**186–194.
27. **Ewing, B., L. Hillier, M. C. Wendl, and P. Green.** 1998. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. Genome Res. **8:**175–185.
28. **Falquet, L., M. Pagni, P. Bucher, N. Hulo, C. J. Sigrist, K. Hofmann, and A. Bairoch.** 2002. The PROSITE database, its status in 2002. Nucleic Acids Res. **30:**235–238.
29. **Feirtag, J. M., J. P. Petzel, E. Pasalodos, K. A. Baldwin, and L. L. McKay.** 1991. Thermosensitive plasmid replication, temperature-sensitive host growth, and chromosomal plasmid integration conferred by *Lactococcus lactis* subsp. *cremoris* lactose plasmids in *Lactococcus lactis* subsp. *lactis*. Appl. Environ. Microbiol. **57:**539–548.
30. **Fojan, P., P. H. Jonson, M. T. Petersen, and S. B. Petersen.** 2000. What distinguishes an esterase from a lipase: a novel structural approach. Biochimie **82:**1033–1041.
31. **Foley, S., S. Bron, G. Venema, C. Daly, and G. F. Fitzgerald.** 1996. Molecular analysis of the replication origin of the Lactococcus lactis plasmid pCJ305. Plasmid **36:**125–141.
32. **Frere, J., M. Novel, and G. Novel.** 1993. Molecular analysis of the Lactococcus lactis subspecies lactis CNRZ270 bidirectional theta replicating lactose plasmid pUCL22. Mol. Microbiol. **10:**1113–1124.
33. **Gasson, M. J.** 1990. In vivo genetic systems in lactic acid bacteria. FEMS Microbiol. Rev. **7:**43–60.
34. **Gravesen, A., J. Josephsen, A. von Wright, and F. K. Vogensen.** 1995. Characterization of the replicon from the lactococcal theta-replicating plasmid pJW563. Plasmid **34:**105–118.
35. **Groot, M. N., E. Klaassens, W. M. de Vos, J. Delcour, P. Hols, and M. Kleerebezem.** 2005. Genome-based in silico detection of putative manganese transport systems in Lactobacillus plantarum and their genetic analysis. Microbiology **151:**1229–1238.
36. **Haandrikman, A. J., C. van Leeuwen, J. Kok, P. Vos, W. M. de Vos, and G. Venema.** 1990. Insertion elements on lactococcal proteinase plasmids. Appl. Environ. Microbiol. **56:**1890–1896.
37. **Henikoff, S., J. G. Henikoff, and S. Pietrokovski.** 1999. Blocks+: a non-redundant database of protein alignment blocks derived from multiple compilations. Bioinformatics **15:**471–479.
38. **Horng, J. S., K. M. Polzin, and L. L. McKay.** 1991. Replication and temperature-sensitive maintenance functions of lactose plasmid pSK11L from *Lactococcus lactis* subsp. *cremoris*. J. Bacteriol. **173:**7573–7581.
39. **Kehres, D. G., C. H. Lawyer, and M. E. Maguire.** 1998. The CorA magnesium transporter gene family. Microb. Comp. Genomics **3:**151–169.

40. **Khunajakr, N., C. Q. Liu, P. Charoenchai, and N. W. Dunn.** 1999. A plasmid-encoded two-component regulatory system involved in copper-inducible transcription in Lactococcus lactis. Gene **229:**229–235.
41. **Kiewiet, R., S. Bron, K. de Jonge, G. Venema, and J. F. Seegers.** 1993. Theta replication of the lactococcal plasmid pWVO2. Mol. Microbiol. **10:**319–327.
42. **Klaenhammer, T., E. Altermann, F. Arigoni, A. Bolotin, F. Breidt, J. Broadbent, R. Cano, S. Chaillou, J. Deutscher, M. Gasson, M. van de Guchte, J. Guzzo, A. Hartke, T. Hawkins, P. Hols, R. Hutkins, M. Kleerebezem, J. Kok, O. Kuipers, M. Lubbers, E. Maguin, L. McKay, D. Mills, A. Nauta, R. Overbeek, H. Pel, D. Pridmore, M. Saier, D. van Sinderen, A. Sorokin, J. Steele, D. O'Sullivan, W. de Vos, B. Weimer, M. Zagorec, and R. Siezen.** 2002. Discovering lactic acid bacteria by genomics. Antonie Leeuwenhoek **82:**29–58.
43. **Kleerebezem, M., J. Boekhorst, R. van Kranenburg, D. Molenaar, O. P. Kuipers, R. Leer, R. Tarchini, S. A. Peters, H. M. Sandbrink, M. W. Fiers, W. Stiekema, R. M. Lankhorst, P. A. Bron, S. M. Hoffer, M. N. Groot, R. Kerkhoven, M. de Vries, B. Ursing, W. M. de Vos, and R. J. Siezen.** 2003. Complete genome sequence of Lactobacillus plantarum WCFS1. Proc. Natl. Acad. Sci. USA **100:**1990–1995.
44. **Kok, J.** 1990. Genetics of the proteolytic system of lactic acid bacteria. FEMS Microbiol. Rev. **7:**15–42.
45. **Kok, J., and W. M. de Vos.** 1994. The proteolytic system of lactic acid bacteria, p. 169–210. *In* M. Gasson and W. M. de Vos (ed.), Genetics and biotechnology of lactic acid bacteria. Blackie Academic and Professional, London, United Kingdom.
46. **Liu, C. Q., P. Charoechai, N. Khunajakr, Y. M. Deng, Widodo, and N. W. Dunn.** 2002. Genetic and transcriptional analysis of a novel plasmid-encoded copper resistance operon from Lactococcus lactis. Gene **297:**241–247.
47. **Marugg, J. D., R. van Kranenburg, P. Laverman, G. A. Rutten, and W. M. de Vos.** 1996. Identical transcriptional control of the divergently transcribed *prtP* and *prtM* genes that are required for proteinase production in *Lactococcus lactis* SK11. J. Bacteriol. **178:**1525–1531.
48. **McKay, L. L.** 1983. Functional properties of plasmids in lactic streptococci. Antonie Leeuwenhoek **49:**259–274.
49. **Mills, D. A., L. L. McKay, and G. M. Dunny.** 1996. Splicing of a group II intron involved in the conjugative transfer of pRS01 in lactococci. J. Bacteriol. **178:**3531–3538.
50. **Monnet, C., M. Nardi, P. Hols, M. Gulea, G. Corrieu, and V. Monnet.** 2003. Regulation of branched-chain amino acid biosynthesis by alpha-acetolactate decarboxylase in Streptococcus thermophilus. Lett. Appl. Microbiol. **36:**399–405.
51. **Monnet, V., M. Nardi, A. Chopin, M. C. Chopin, and J. C. Gripon.** 1994. Biochemical and genetic characterization of PepF, an oligopeptidase from Lactococcus lactis. J. Biol. Chem. **269:**32070–32076.
52. **Nardi, M., P. Renault, and V. Monnet.** 1997. Duplication of the *pepF* gene and shuffling of DNA fragments on the lactose plasmid of *Lactococcus lactis*. J. Bacteriol. **179:**4164–4171.
53. **O'Sullivan, D., R. P. Ross, D. P. Twomey, G. F. Fitzgerald, C. Hill, and A. Coffey.** 2001. Naturally occurring lactococcal plasmid pAH90 links bacteriophage resistance and mobility functions to a food-grade selectable marker. Appl. Environ. Microbiol. **67:**929–937.
54. **Romero, D. A., and T. R. Klaenhammer.** 1993. Transposable elements in Lactococci: a review. J. Dairy Sci. **76:**1–19.
55. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
56. **Schultz, J. E., and A. Matin.** 1991. Molecular and functional characterization of a carbon starvation gene of Escherichia coli. J. Mol. Biol. **218:**129–140.
57. **Seegers, J. F., S. Bron, C. M. Franke, G. Venema, and R. Kiewiet.** 1994. The majority of lactococcal plasmids carry a highly related replicon. Microbiology **140:**1291–1300.
58. **Sharp, P. M., and W. H. Li.** 1987. The codon Adaptation Index—a measure of directional synonymous codon usage bias, and its potential applications. Nucleic Acids Res. **15:**1281–1295.
59. **Shearman, C., J. J. Godon, and M. Gasson.** 1996. Splicing of a group II intron in a functional transfer gene of Lactococcus lactis. Mol. Microbiol. **21:**45–53.
60. **Simons, A. F. M., F. A. Exterkate, S. Visser, and W. M. de Vos.** 1985. Proteinases of Streptococcus cremoris: physiological and genetic analysis. Antonie Leeuwenhoek **51:**565–566.
61. **Simons, G., and W. M. de Vos.** 1987. Gene expression in lactic streptococci, p. 458–460. *In* O. M. Neyssel, R. R. van der Meer, and K. C. A. M. Luyben (ed.), Proceedings of the Fourth European Congress on Biotechnology, vol. 1. Elsevier, Amsterdam, The Netherlands.
62. **Smith, R. L., E. Gottlieb, L. M. Kucharski, and M. E. Maguire.** 1998. Functional similarity between archaeal and bacterial CorA magnesium transporters. J. Bacteriol. **180:**2788–2791.
63. **Sterkenburg, A., P. Van Leeuwen, and J. T. Wouters.** 1988. Loss of phage resistance encoded by plasmid pSK112 in chemostat cultures of Lactococcus lactis ssp. cremoris SK110. Biochimie **70:**451–456.
64. **Tatusov, R. L., N. D. Fedorova, J. D. Jackson, A. R. Jacobs, B. Kiryutin, E. V. Koonin, D. M. Krylov, R. Mazumder, S. L. Mekhedov, A. N. Nikolskaya, B. S. Rao, S. Smirnov, A. V. Sverdlov, S. Vasudevan, Y. I. Wolf, J. J. Yin, and D. A. Natale.** 2003. The COG database: an updated version includes eukaryotes. BMC Bioinformatics **4:**41.
65. **Tynkkynen, S., G. Buist, E. Kunji, J. Kok, B. Poolman, G. Venema, and A. Haandrikman.** 1993. Genetic and biochemical characterization of the oligopeptide transport system of *Lactococcus lactis*. J. Bacteriol. **175:**7523–7532.
66. **van Asseldonk, M., G. Rutten, M. Oteman, R. J. Siezen, W. M. de Vos, and G. Simons.** 1990. Cloning of usp45, a gene encoding a secreted protein from Lactococcus lactis subsp. lactis MG1363. Gene **95:**155–160.
67. **van Kranenburg, R., and W. M. de Vos.** 1998. Characterization of multiple regions involved in replication and mobilization of plasmid pNZ4000 coding for exopolysaccharide production in *Lactococcus lactis*. J. Bacteriol. **180:**5285–5290.
68. **van Kranenburg, R., M. Kleerebezem, and W. M. de Vos.** 2000. Nucleotide sequence analysis of the lactococcal EPS plasmid pNZ4000. Plasmid **43:**130–136.
69. **van Rooijen, R. J., and W. M. de Vos.** 1990. Molecular cloning, transcriptional analysis, and nucleotide sequence of lacR, a gene encoding the repressor of the lactose phosphotransferase system of Lactococcus lactis. J. Biol. Chem. **265:**18499–18503.
70. **van Rooijen, R. J., S. van Schalkwijk, and W. M. de Vos.** 1991. Molecular cloning, characterization, and nucleotide sequence of the tagatose 6-phosphate pathway gene cluster of the lactose operon of Lactococcus lactis. J. Biol. Chem. **266:**7176–7181.
71. **Vos, P., G. Simons, R. J. Siezen, and W. M. de Vos.** 1989. Primary structure and organization of the gene for a procaryotic, cell envelope-located serine proteinase. J. Biol. Chem. **264:**13579–13585.
72. **Vos, P., M. van Asseldonk, F. van Jeveren, R. Siezen, G. Simons, and W. M. de Vos.** 1989. A maturation protein is essential for production of active forms of *Lactococcus lactis* SK11 serine proteinase located in or secreted from the cell envelope. J. Bacteriol. **171:**2795–2802.
73. **Wouters, J. A., J. W. Sanders, J. Kok, W. M. de Vos, O. P. Kuipers, and T. Abee.** 1998. Clustered organization and transcriptional analysis of a family of five csp genes of Lactococcus lactis MG1363. Microbiology **144:**2885–2893.
74. **Yu, W., K. Gillies, J. K. Kondo, J. R. Broadbent, and L. L. McKay.** 1996. Loss of plasmid-mediated oligopeptide transport system in lactococci: another reason for slow milk coagulation. Plasmid **35:**145–155.
75. **Yu, W., K. Gillies, J. K. Kondo, J. R. Broadbent, and L. L. McKay.** 1995. Plasmid-mediated oligopeptide transport system in lactococci. Dev. Biol. Stand. **85:**509–521.
76. **Yu, W., I. Mierau, A. Mars, E. Johnson, G. Dunny, and L. L. McKay.** 1995. Novel insertion sequence-like element IS982 in lactococci. Plasmid **33:**218–225.