Chromosomal Diversity in *Lactococcus lactis* and the Origin of Dairy Starter Cultures

William J. Kelly*[,1], Lawrence J. H. Ward†[,2], and Sinead C. Leahy[1]

[1]Rumen Microbial Genomics, AgResearch Limited, Grasslands Research Center, Private Bag 11008, Palmerston North 4442, New Zealand

[2]Fonterra Research Center, Private Bag 11029, Palmerston North 4442, New Zealand

†Present address: No. 1 Line, Palmerston North, New Zealand

*Corresponding author: E-mail: bill.kelly@agresearch.co.nz.

Accepted: 9 September 2010

Abstract

A large collection of *Lactococcus lactis* strains, including wild-type isolates and dairy starter cultures, were screened on the basis of their phenotype and the macrorestriction patterns produced from pulsed-field gel electrophoresis (PFGE) analysis of *Sma*I digests of genomic DNA. Three groups of dairy starter cultures, used for different purposes in the dairy industry, and a fourth group made up of strains isolated from the environment were selected for analysis of their chromosomal diversity using the endonuclease I-*Ceu*I. Chromosome architecture was largely conserved with each strain having six copies of the rRNA genes, and the chromosome size of individual strains ranged between 2,240 and 2,688 kb. The origin of *L. lactis* strains showed the greatest correlation with chromosome size, and dairy strains, particularly those with the cremoris phenotype, had smaller chromosomes than wild-type strains. Overall, this study, coupled with analysis of the sequenced *L. lactis* genomes, provides evidence that defined strain dairy starter cultures have arisen from plant *L. lactis* strains. Adaptation of these strains to the dairy environment has involved loss of functions resulting in smaller chromosomes and acquisition of genes (usually plasmid associated) that facilitate growth in milk. We conclude that dairy starter cultures generally and the industrially used cremoris and diacetylactis phenotype strains in particular comprise a specialized group of *L. lactis* strains that have been selected to become an essential component of industrial processes and have evolved accordingly, so that they are no longer fit to survive outside the dairy environment.

Key words: *L. lactis* subsp. *cremoris*, *L. lactis* subsp. *lactis*, *L. lactis* subsp. *lactis* biovar *diacetylactis*, dairy starter cultures, PFGE, chromosome size.

Introduction

Lactococcus lactis can be isolated from various environments but is predominantly studied because of its role as the main constituent of many industrial and artisanal starter cultures used for the manufacture of a vast range of fermented dairy products including fermented milks, sour cream, soft and hard cheeses, and lactic casein (Ward et al. 2002). For large-scale commercial production, the starter cultures used are commonly defined strains, which have been selected for their desirable properties especially in relation to acid production, flavor development, and bacteriophage resistance (Limsowtin et al. 1996).

The taxonomy of *L. lactis* has changed many times but is currently phenotypically based (Schleifer et al. 1985; van Hylckama Vlieg et al. 2006; Rademaker et al. 2007) and includes two subspecies (subsp. *lactis* and subsp. *cremoris*) and one biovar (subsp. *lactis* biovar *diacetylactis*). The lactis and cremoris phenotypes are differentiated on the basis of arginine utilization, growth temperature, and salt tolerance, whereas the biovar *diacetylactis* strains have the additional ability to metabolize citrate. Numerous studies including DNA–DNA hybridization, 16S rRNA, and gene sequence analysis have demonstrated the existence of two main genotypes. These two genotypic groups have also been called *L. lactis* subsp. *lactis* and *L. lactis* subsp. *cremoris*, but unfortunately the genotype and phenotype designations do not necessarily correspond, thus introducing a degree of confusion into the taxonomy of this species (Tailliez et al. 1998). An extensive study of 102 *L. lactis* isolates of dairy and plant origin using various genomic fingerprinting methods and

multilocus sequence analysis has clearly demonstrated that two major lineages exist (Rademaker et al. 2007). One of these comprises those strains with a *L. lactis* subsp. *cremoris* genotype and includes strains with both lactis and cremoris phenotypes. The other comprises those strains with a *L. lactis* subsp. *lactis* genotype that includes strains with the lactis phenotype as well as biovar *diacetylactis*. Comparative genome hybridization (CGH) using 39 *L. lactis* strains of dairy or plant origin (Bayjanov et al. 2009) provides further evidence confirming the unusual taxonomic structure in this species. As a result, it is necessary to specify a genotype (*cremoris* or *lactis*) and a phenotype (cremoris, diacetylactis, or lactis) to adequately describe individual strains.

Strains that show both the subsp. *cremoris* genotype and phenotype cluster closely together and form a definite subgroup that shows limited diversity relative to the other *L. lactis* strains examined (Rademaker et al. 2007; Taïbi et al. 2010). These *L. lactis* subsp. *cremoris* strains are favored for use as defined strain starter cultures for Cheddar cheese production because they are less likely to cause bitterness and other flavor defects (Heap 1998). The citrate-metabolizing biovar *diacetylactis* strains contribute to the flavor and aroma profile of a range of fermented dairy products and are also a component of the starter blends used for lactic casein manufacture (Heap and Lawrence 1984). These strains have long been distinguished taxonomically (Kempler and McKay 1981), but with the description of the genus *Lactococcus* (Schleifer et al. 1985), they were incorporated into *L. lactis* subsp. *lactis*. Biovar *diacetylactis* dairy starter strains have been genotypically (Köhler et al. 1991; Beimfohr et al. 1997) and phenotypically (Bachmann et al. 2009) distinguished from other *L. lactis* cultures, suggesting that these cultures may also form a separate subgroup.

Both *L. lactis* subspecies have been isolated from a variety of environmental sources but are most commonly associated with fresh or fermented plant material or with milk and milk products. Strains that show the *lactis* subspecies genotype can be readily isolated from these environments, whereas isolations of cultures with the *cremoris* subspecies genotype are comparatively rare (Klijn et al. 1995; Salama et al. 1995). Attempts to isolate new cremoris or diacetylactis phenotype strains from environmental sources have met with little success as wild-type strains of both subspecies show the lactis phenotype (Klijn et al. 1995; Salama et al. 1995; Ward et al. 1998).

Because of its industrial relevance, *L. lactis* has become the best studied of the lactic acid bacteria and regarded as a model organism for this bacterial group, although most work has been focused on a small number of laboratory strains of dairy origin. Complete genome sequences have been published for four strains. These include the two plasmid-cured strains (IL1403 and MG1363) on which much of the detailed biochemical and genetic knowledge of *L. lactis* is based (Bolotin et al. 2001; Wegmann et al. 2007). Both IL1403 and MG1363 belong to *L. lactis* subsp. *lactis* phenotypically, but the parent strain of IL1403 (CNRZ157) has a citrate permease plasmid and is able to metabolize citrate placing it with *L. lactis* subsp. *lactis* biovar *diacetylactis*, whereas MG1363 has a *lactis* phenotype and a *cremoris* genotype. The third genome-sequenced strain (SK11) has been used as a cheese starter culture and belongs to the subgroup of strains with both the subsp. *cremoris* genotype and phenotype (Makarova et al. 2006). The fourth genome is from a *L. lactis* subsp. *lactis* strain of plant origin (KF147), and a partial sequence is also available for a second plant strain (KF282) (Siezen et al. 2008, 2010). Comparison of the genomes from plant and dairy isolates has highlighted the differences in gene content that can occur between individual strains in the same species (Siezen et al. 2008) and shows that sequencing of one representative genome does not give a complete picture of the genetic repertoire of a species. Attempts to describe this intraspecies diversity have led to the terms species genome (Lan and Reeves 2000) and pangenome (Medini et al. 2005) being defined to cover all the genes present in the characterized strains of a species. Under both definitions, the genome has a core of genes responsible for the basic aspects of the biology of the species and a set of auxiliary or dispensable genes that contribute to species diversity and may provide a selective advantage in certain environments. As a result, there is much to learn about diversity at the intraspecies level, and the aim of this work was to examine the chromosomal diversity of a large collection of *L. lactis* strains to provide a framework for future comparative genomic work with this industrially important bacterial species.

Materials and Methods

Bacterial Cultures and Growth Media

In initial screening, 558 *L. lactis* strains were examined by pulsed-field gel electrophoresis (PFGE) analysis of *Sma*I digests of genomic DNA. These included 289 strains with the cremoris phenotype and genotype (*L. lactis* subsp. *cremoris*), 197 strains with the lactis phenotype (*L. lactis* subsp. *lactis*), and 72 strains that were able to metabolize citrate (*L. lactis* subsp. *lactis* biovar *diacetylactis*). The cremoris phenotype strains and the citrate-metabolizing strains were all either used as or isolated from dairy starter cultures. The lactis phenotype strains included cultures with both *cremoris* and *lactis* genotypes and came from diverse origins. These included dairy starter cultures and individual strains isolated from raw milk, pasture, soil, plant material, the rumen, and insect gut. Some of these have been described elsewhere (table 1), and the strains isolated from plant sources were included in the *L. lactis* diversity study

Table 1
Sizes of I-*Ceu*I Restriction Fragments, Total Chromosome Size (kb), Genotype, and Origin of *Lactococcus lactis* Cultures

Strain	Ce1	Ce2	Ce3	Ce4	Ce5	Ce6	Total (kb)	Genotype	Origin (Reference)[a]
1. *L. lactis* subsp. *cremoris*									
LW1477	1,340	540	220	75	45	22	2,242	*cremoris*	Dairy starter culture, *Scr*FI producer
KH	1,380	530	240	95	45	38	2,328	*cremoris*	Dairy starter culture (1)
AM2	1,440	520	240	75	45	38	2,358	*cremoris*	Dairy starter culture (1,2)
112	1,440	530	250	75	45	22	2,362	*cremoris*	Dairy starter culture
LW1494	1,440	530	240	75	45	45	2,375	*cremoris*	Dairy starter culture
HP	1,900	65	260	80	45	28	2,378	*cremoris*	Dairy starter culture (1,2)
2188	1,900	80	250	80	45	38	2,393	*cremoris*	Dairy starter culture
166	1,440	530	260	80	45	38	2,393	*cremoris*	Dairy starter culture (2)
KF322	1,440	460	340	75	45	38	2,398	*cremoris*	Isolated from mixed strain dairy starter culture
2128	1,440	560	230	90	45	35	2,400	*cremoris*	Dairy starter culture
FG2	1,800	380	70	80	45	35	2,410	*cremoris*	Dairy starter culture (3)
LW1499	1,440	530	300	75	45	22	2,412	*cremoris*	Dairy starter culture
LW1489	1,480	530	260	85	45	22	2,422	*cremoris*	Dairy starter culture
AM1[b]	1,520	530	240	75	45	38	2,448	*cremoris*	Dairy starter culture (1,2)
SK11	1,520	530	240	75	45	38	2,448	*cremoris*	Dairy starter culture, phage-resistant AM1 (1,2)
448	1,440	560	300	75	45	22	2,442	*cremoris*	Dairy starter culture (2)
BK5	1,520	520	250	80	45	38	2,453	*cremoris*	Dairy starter culture (1,2)
LW1492	1,480	560	280	90	45	22	2,477	*cremoris*	Dairy starter culture
168	1,520	600	240	90	45	48	2,543	*cremoris*	Dairy starter culture
E8	1,520	610	250	75	45	52	2,552	*cremoris*	Dairy starter culture (1,2)
2. *L. lactis* subsp. *lactis* biovar *diacetylactis*									
CNRZ157	1,440	540	240	80	45	20	2,365	*lactis*	Dairy starter culture (4)
LW1807	1,380	570	270	80	45	20	2,365	*lactis*	Dairy starter culture
LW1688	1,380	570	270	80	45	20	2,365	*lactis*	Dairy starter culture
LW3081	1,440	540	240	95	45	20	2,380	*lactis*	Dairy starter culture
D10	1,380	570	300	80	45	20	2,395	*lactis*	Dairy starter culture
LW1690	1,380	600	270	80	45	20	2,395	*lactis*	Dairy starter culture
LW3079	1,900	115	240	80	45	35	2,415	*lactis*	Isolated from mixed strain dairy starter culture
LW1503	1,440	600	240	95	45	20	2,440	*lactis*	Dairy starter culture
LW1811	1,440	600	240	95	45	22	2,442	*lactis*	Dairy starter culture
DRC3	1,440	600	240	95	45	22	2,442	*lactis*	Dairy starter culture (5)
DRC1	1,440	600	270	80	45	20	2,455	*lactis*	Dairy starter culture (2,5)
LW3074	1,440	600	270	80	45	20	2,455	*lactis*	Dairy starter culture
LW3087	1,440	600	270	95	45	20	2,470	*lactis*	Dairy starter culture
LW2333	1,440	650	240	95	45	22	2,492	*lactis*	Dairy starter culture
DRC2	1,640	440	270	80	45	20	2,495	*lactis*	Dairy starter culture (2,5)
LW1505	1,440	650	240	95	45	35	2,505	*lactis*	Dairy starter culture
LW840	1,540	570	270	05	45	20	2,540	*lactis*	Dairy starter culture
D6	1,540	600	285	80	45	20	2,570	*lactis*	Dairy starter culture
LW3076	1,540	600	285	80	45	20	2,570	*lactis*	Dairy starter culture
LW3077	1,540	600	285	80	45	35	2,585	*lactis*	Isolated from mixed strain dairy starter culture
3. *L. lactis* subsp. *lactis* (dairy cultures)									
LW1509	1,520	530	270	80	90	25	2,515	*cremoris*	Dairy starter culture
MG1363	1,640	530	240	80	50	25	2,565	*cremoris*	Dairy starter culture, plasmid-free NCDO712 (6)
NCDO712	1,640	530	270	80	50	25	2,595	*cremoris*	Dairy starter culture (6)
LW1515	1,640	530	290	75	50	20	2,605	*cremoris*	Dairy starter culture
GL17	1,640	550	300	110	50	25	2,675	*cremoris*	Dairy starter culture
IL1403	1,440	540	240	80	45	20	2,365	*cremoris*	Dairy starter culture, plasmid-free CNRZ157 (4)
LW1444	1,480	530	220	90	45	35	2,400	*lactis*	Dairy starter culture
BA2	1,440	550	260	90	45	20	2,405	*lactis*	Dairy starter culture
LW1448	1,540	500	240	80	50	22	2,432	*lactis*	Dairy starter culture
KF324	1,540	500	240	90	45	22	2,437	*lactis*	Dairy starter culture
ATCC7962	1,540	530	230	80	45	20	2,445	*lactis*	Dairy starter culture, nisin producer
NCDO895	1,440	550	230	170	45	20	2,455	*lactis*	Dairy starter culture, nisin producer
LW1512	1,440	550	300	90	45	35	2,460	*lactis*	Dairy starter culture
C10	1,520	530	240	90	45	35	2,460	*lactis*	Dairy starter culture

Table 1
Continued

Strain	Ce1	Ce2	Ce3	Ce4	Ce5	Ce6	Total (kb)	Genotype	Origin (Reference)[a]
LW1449	1,440	650	220	120	45	20	2,495	*lactis*	Dairy starter culture
ML8	1,540	530	260	90	45	35	2,500	*lactis*	Dairy starter culture (1,2)
NCDO1404	1,620	530	200	80	50	20	2,500	*lactis*	Dairy starter culture, nisin producer
LW2004	1,540	540	260	100	45	20	2,505	*lactis*	Dairy starter culture (7)
LW1514	1,540	600	260	100	45	22	2,567	*lactis*	Dairy starter culture
U	1,640	550	260	110	45	28	2,633	*lactis*	Dairy starter culture
4. *L. lactis* subsp. *lactis* (wild-type cultures)									
KW8	1,440	570	270	95	45	22	2,442	*cremoris*	Kaanga wai (fermented corn) (8)
KW2	1,590	520	240	80	48	22	2,500	*cremoris*	Kaanga wai (fermented corn) (8)
KF343	1,700	540	220	90	48	28	2,626	*cremoris*	Cow's milk
KF355	1,700	500	270	80	48	38	2,636	*cremoris*	Cow's milk
LW1190	1,700	500	270	80	48	38	2,636	*cremoris*	Sheep's milk (9)
KF292	1,590	500	220	90	45	22	2,467	*cremoris*	Soya sprouts, nisin producer (10)
511	1,640	470	220	90	45	20	2,485	*lactis*	Rumen, nisin producer
KF196	1,590	520	220	90	45	20	2,485	*lactis*	Radish sprouts, nisin producer (10)
KF363	1,640	470	240	90	45	20	2,505	*lactis*	Soil
KF201	1,640	470	240	80	45	38	2,513	*lactis*	Sliced mixed vegetables (10)
KF146	1,590	580	220	90	45	20	2,545	*lactis*	Alfalfa and radish sprouts, nisin producer (10)
LW1320	1,540	630	250	90	45	20	2,575	*lactis*	Goat's milk (9)
KF181	1,590	520	320	80	45	22	2,577	*lactis*	Alfalfa and onion sprouts (10)
KF67	1,590	500	340	90	45	20	2,585	*lactis*	Grapefruit juice, nisin producer (10)
LW1,180	1,640	500	270	110	45	22	2,587	*lactis*	Sheep's milk (9)
N1	1,760	450	240	75	45	20	2,590	*lactis*	Moth larval midgut, nisin producer (11)
KF5	1,590	520	290	120	45	28	2,593	*lactis*	Alfalfa sprouts (10)
KF165	1,740	500	220	80	45	20	2,605	*lactis*	Mung bean sprouts, nisin producer (10)
KF282	1,740	520	240	80	45	20	2,645	*lactis*	Mustard and cress, nisin producer (10)
KF147	1,740	520	270	90	48	20	2,688	*lactis*	Mung bean sprouts (10)

[a] References: (1) Lawrence and Pearce (1972); (2) Jarvis and Wolff (1979); (3) Davidson et al. (1996); (4) Chopin et al. (1984); (5) Czulak and Hammond (1954); (6) Gasson (1983); (7) Ward et al. (2004); (8) Kelly et al. (1994); (9) Ward et al. (1998); (10) Kelly et al. (1998a); (11) Shannon et al. (2001).

[b] Cultures for which genome sequences are available and their parent strains are shown in bold.

(Rademaker et al. 2007). Forty-eight of the *lactis* phenotype strains were nisin producers. Dairy starter strains were mainly from the culture collection of the Fonterra Research Center, Palmerston North, New Zealand, with additional cultures obtained from other culture collections or isolated from mixed strain dairy starters.

Lactococci were grown at 28 °C in M17 broth (Merck) (Terzaghi and Sandine 1975) supplemented with 0.5% w/v glucose for growth of the plasmid-free dairy strains (IL1403 and MG1363) and the wild-type strains. The stock cultures were maintained at −85 °C in M17 broth supplemented with 20% (v/v) glycerol. Tests for arginine and citrate metabolism were used to confirm the phenotype of the strains, and strain genotypes were determined using the polymerase chain reaction (PCR) primers and conditions described previously (Ward et al. 1998).

Pulsed-Field Gel Electrophoresis

Cells were harvested from 1.5 ml of an overnight culture by centrifugation (10,000 × g, 10 min), washed twice with 1 M NaCl:10 mM Tris–Cl (pH 7.6), and 300 μl aliquots were mixed with an equal volume of 2% (w/v) low melt agarose (Bio-Rad Laboratories). Embedded cells were lysed by treatment with lysozyme (1 mg/ml in EC buffer, 6 mM Tris–Cl:1 M NaCl:100 mM ethylenediaminetetraacetic acid [EDTA]:1% [w/v] sarkosyl, pH 7.6) overnight at 37 °C and proteinase K (0.5 mg/ml in lysis buffer, 50 mM Tris–Cl:50 mM EDTA:1% [w/v] sarkosyl, pH 8.0) for 24 h at 50 °C. Agarose plugs containing intact genomic DNA were washed three times with Tris–EDTA buffer (10 mM Tris–Cl:1 mM EDTA, pH 8.0) before storage in 10 mM Tris–Cl:100 mM EDTA (pH 8.0) at 4 °C. DNA embedded in agarose was digested for 16 h with 1.0 U of *Apa*I, *Sma*I, or I-*Ceu*I (New England Biolabs) in 100 μl of restriction enzyme buffer, loaded into wells of 1% (w/v) agarose gels (pulsed-field certified agarose, Bio-Rad), and run at 200 V for 20 h at 14 °C in 0.5× Tris–borate buffer (Sambrook et al. 1989) using a CHEF DR III PFGE apparatus and model 1000 mini chiller (Bio-Rad). Pulse times used were 1–30 s for *Apa*I or *Sma*I and 5–60 s for I-*Ceu*I. To determine the size of the largest fragments from I-*Ceu*I digests, gels were prepared from 0.8% chromosomal grade agarose (Bio-Rad) and run with the pulse time ramped between 150 and 400 s. Fragments smaller than 100 kb were

also resolved and measured using a FIGE Mapper electrophoresis system (Bio-Rad). DNA was visualized by staining with ethidium bromide and the image captured using a Gel Doc 1000 system (Eastman Kodak). Partial digestion with I-*Ceu*I was used to establish the *rrn* chromosomal skeleton as described by Liu et al. (1999). Genomic DNA prepared from *L. lactis* subsp. *lactis* IL1403 and phage lambda concatamers or *Saccharomyces cerevisiae* pulsed-field gel (PFG) markers (New England Biolabs) were used as size standards.

Plasmid Analysis

Plasmid DNA was isolated by the method of Anderson and McKay (1983) and the size of individual plasmid bands determined following electrophoresis in 0.7% agarose gels in Tris–acetate buffer (Sambrook et al. 1989) for 3 h at 4 V/cm and staining as above. Strains were also screened for the presence of large linear plasmids by running undigested genomic DNA in PFGs as described above.

DNA Methodology

To determine the relatedness of citrate-metabolizing strains based on the presence and chromosomal location of their prophage, the primers described by Chopin et al. (2001) were used in PCR reactions to amplify the chromosome–prophage junction regions from IL1403 for use as probes for PFGs. Hybridizations were done using the North2South Direct HRP Labeling and Detection Kit (Pierce) using the conditions recommended by the manufacturer. To determine the presence of plasmids encoding citrate permease (*cit*P), plasmid gels were hybridized with a probe to the gene for citrate permease that was constructed by PCR amplification from strain LW1503 using primers described by Klijn et al. (1995). To determine if cultures, which failed to cut with *Sma*I, contained the *Scr*FI R/M system, PFGs were hybridized with a probe specific for the *scr*FIAM methylase, which was constructed by PCR amplification from strain UC503 using primers described by Szatmari et al. (2006).

Statistical Analysis

Differences in average chromosome size among subspecific groups of *L. lactis* from each origin (dairy and wild type) were tested using a one-way analysis of variance and fitting a single factor comprising each genotype–phenotype origin combination. To test whether regions of the chromosome differ in degree of variability, we compared the variances in the lengths of different chromosomal regions based on I-*Ceu*I fragments. The variance in the length of a given region is expected to increase linearly with fragment size assuming that the number of insertions, duplications, and deletions per fragment increase with fragment length. Therefore, the variances in the sizes of the I-*Ceu*I fragments were standardized by dividing the variance of a given fragment by its average size before subjecting the values to pairwise *F* tests to test the equality of variances (Sokal and Rohlf 1981).

Genomic Analysis

Genomic data for the four publicly available and completed *L. lactis* genome sequences (GenBank accession numbers AM406671, CP000425, AE005176, and CP001834) were downloaded from the National Center for Biotechnology Information (NCBI) Web site (http://www.ncbi.nlm.nih.gov). Genome alignments of chromosomal sequences were performed using Mauve software (Version 2.3.1) (Darling et al. 2004).

Phylogenetic Analysis

Phylogenetic relationships were determined using a supertree approach. Protein coding gene sets for the four *L. lactis* strains, three *Streptococcus thermophilus* strains (GenBank accession numbers CP000023, CP000024, and CP000419), five *Lactobacillus* species (GenBank accession numbers AL935263, CR936503, CR954253, CP000416, and CP000517), as well as two outgroups (*Listeria* species; GenBank accession numbers AL591824 and AL592022) were downloaded from the NCBI Web site. The *S. thermophilus*, *L. delbrueckii* subsp. *bulgaricus* and *L. helveticus* genomes were included for comparison because of their use as dairy starter cultures. For a full description of the supertree methodology used please refer to Fitzpatrick et al. (2006).

Results and Discussion

PFGE Patterns of *L. lactis* Subsp. *cremoris* Dairy Starter Cultures

PFGE analysis of *Sma*I digests of genomic DNA from the 289 *L. lactis* subsp. *cremoris* strains showed that 230 (80%) could be linked into 12 groups of related strains. Representatives of these groups are included in table 1. Three (E8, HP, and SK11) of the four strains compared by Taïbi et al. (2010) were representative of groups of strains, whereas the fourth (Wg2) had a unique PFGE profile. The observation that many strains were related was not surprising because studies based on phage host range had previously indicated that a relatively small number of significantly different cremoris starter strains exist (Lawrence and Pearce 1972; Lawrence et al. 1978). The result is that many cremoris strains of diverse origin are unknowingly related and an example of this is shown in figure 1 where strains related to SK11 are compared. SK11 was isolated in New Zealand as a phage-resistant derivative of strain AM1 (DRI 1962), which had been obtained from Professor Collins from University of California Davis and originally named LT8 (DRI 1960). Two other cultures were introduced to New Zealand at the same time, AM2 (FC4) and AM3 (4B) (DRI 1960; Collins 1961),

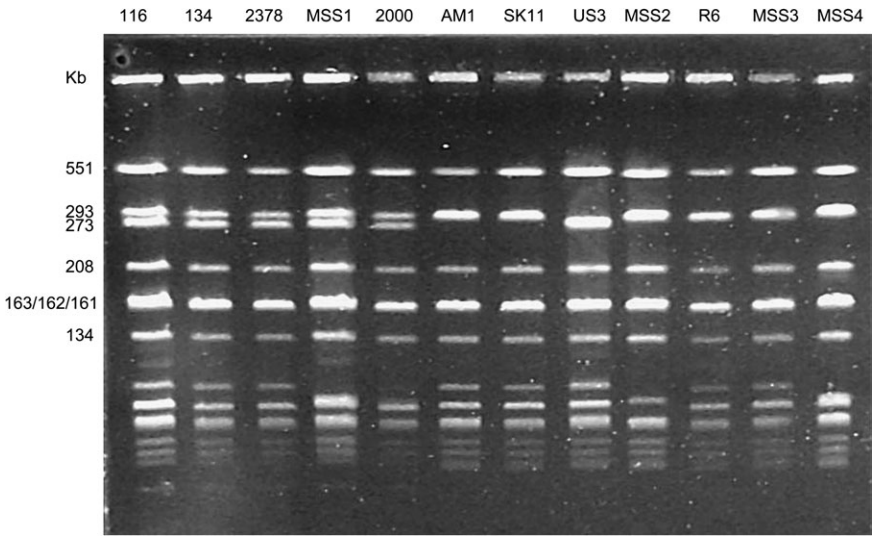

Fig. 1.—PFGE patterns of *Sma*I-digested genomic DNA from *Lactococcus lactis* subsp. *cremoris* SK11 and related strains. The mixed strain starter isolates (MSS1–4) were isolated from various mixed strain dairy starter cultures.

and all three had been isolated from commercial mixed strain cultures. AM1 and AM2 were regarded as slower starters and were found to make consistently good-flavored Cheddar cheese (Martley and Lawrence 1972). Figure 1 shows that AM1 and SK11 have the same PFGE *Sma*I digest pattern but that several other strains are similar. These include US3 and R6, which had both been in use as defined strain starters since the early 1950s (DRI 1951), strain 134 (originally described as a phage-resistant derivative of AM2, Limsowtin and Terzaghi 1976), and several isolates from mixed strain starter cultures.

The predicted PFGE pattern from the sequenced SK11 strain differs from that for SK11 shown in figure 1 but matches to that of strain 134. The major PFGE bands are very similar in these strains with the main difference being a deletion of ~20 kb from one of the largest PFGE bands (a 293-kb doublet which separates into 293- and 273-kb bands). When the 273-kb *Sma*I fragment from the SK11 genome is compared with the corresponding region from MG1363, the only major difference is the presence of a 19-kb prophage sequence (MG-1) in MG1363 (Ventura et al. 2007). A prophage (bIL310) is integrated in the same genomic region in IL1403, and there is strong homology and synteny between bIL310 and MG-1. Consequently, it is likely that a similar prophage has been lost from the sequenced SK11 strain.

DNA from 13 cremoris strains failed to cut with the enzyme *Sma*I, indicative of the presence of a restriction/modification system operating in these strains. Of the known lactococcal R/M systems, only the *Scr*FI methylase potentially blocks the *Sma*I recognition site (Szatmari et al. 2006). Unlike most lactococcal R/M systems, *Scr*FI is chromosomally encoded and a single 75-kb *Apa*I digest fragment of chromosomal DNA from these strains hybridized to a probe for the *scr*FIAM methylase gene. When the histories of these strains were examined, most were found to have been isolated from mixed strain starter cultures.

PFGE Patterns of Citrate-Utilizing Dairy Starter Cultures

The PFGE patterns of *Sma*I-digested genomic DNA from 72 Cit^+ strains showed that most strains had several bands in common, and representatives are shown in figure 2*A*. The relatedness of these strains is supported by the presence of prophage, and these were detected using probes to the chromosome–prophage junctions in strain IL1403 (Chopin et al. 2001). An example is shown in figure 2*B* where the left hand junction fragment between the IL1403 chromosome and the bIL309 prophage hybridizes strongly with all the Cit^+ strains. Six prophages have been identified on the IL1403 chromosome (Bolotin et al. 2001), and junction fragments for five of these (bIL285, bIL309, bIL310, bIL311, and bIL312) could be detected in all strains. The sixth prophage (bIL286) was found only in IL1403 and CNRZ157 and may be less stable than the others as we isolated a derivative of CNRZ157, which was spontaneously cured of this prophage. In a study using minisatellite polymorphism to distinguish closely related *L. lactis* strains, a sequence from within the bIL286 prophage was used as a strain-specific minisatellite. This sequence was only found in the genome of IL1403 and not in nine other *L. lactis* strains (Quénée et al. 2005). The observation by McGrath et al. (2002) that the genomes of IL1403 and the citrate-utilizing strains IL409 (DRC1) and F7/2 contain identical

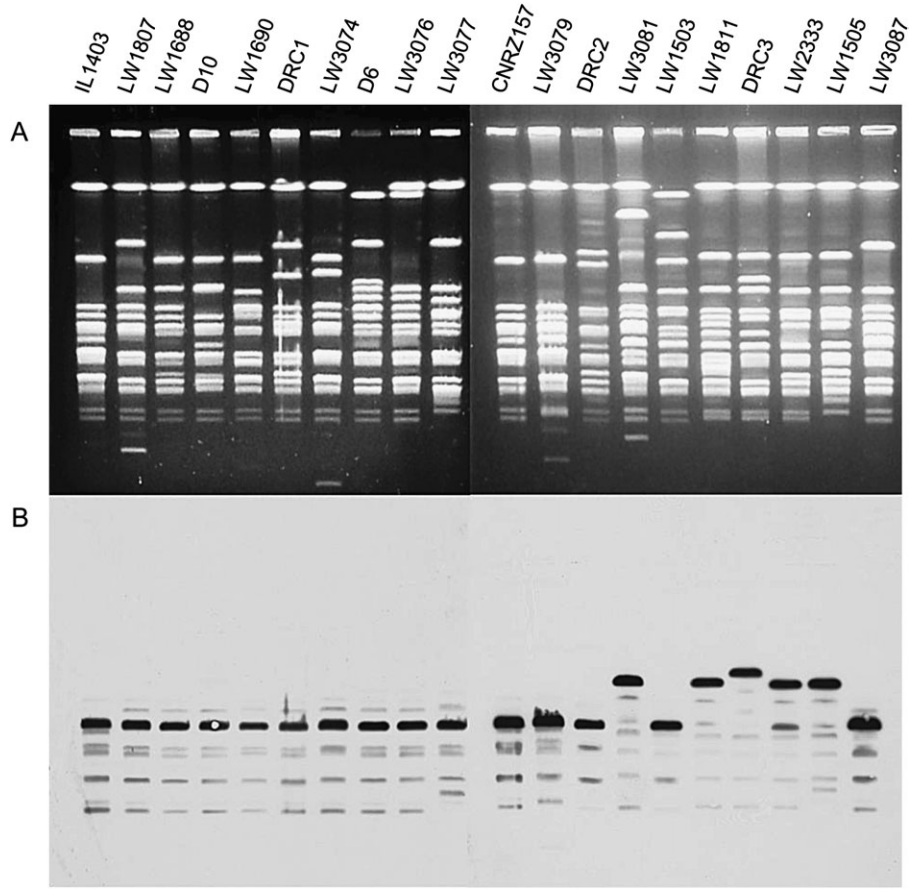

Fig. 2.—(*A*) PFGE patterns of *Sma*I digests of genomic DNA from Cit$^+$ *Lactococcus lactis* strains. (*B*) Southern blot of (*A*) hybridized with a PCR-amplified product of the left hand junction between the IL1403 chromosome and the prophage bIL309.

prophage-encoded bacteriophage resistance genes is in agreement with our observation that these strains are closely related and harbor related prophage. The gene identified (sie_{IL409}) showed 100% amino acid identity with orf2 of bIL309.

The plasmid-free strain IL1403 was originally derived from the citrate-utilizing strain *L. lactis* subsp. *lactis* biovar *diacetylactis* CNRZ157 (IL594) following protoplast-induced curing (Chopin et al. 1984; Bourel et al. 1996). Citrate utilization requires citrate permease to transport citrate into the cell and citrate lyase to initiate citrate breakdown (Drider et al. 2004). In lactococci, these activities are genetically separate with citrate permease being plasmid encoded on an 8-kb plasmid (Kempler and McKay 1981), whereas citrate lyase and other genes involved in citrate breakdown are chromosomal (Bolotin et al. 2001). The distribution of the chromosomal citrate-utilizing genes among different lactococcal strains is not known, but citrate lyase activity was not present in cell-free extracts of 24 dairy lactococcal strains (Harvey and Collins 1961), and the gene cluster *mae-maeP-citRCDEFXG*, which includes the genes required for the synthesis of active citrate lyase, is missing from the other sequenced *L. lactis* strains (Wegmann et al. 2007; Siezen et al. 2008).

PFGE Patterns of *L. lactis* Subsp. *lactis* Cultures

PFGE was also used to compare 197 strains with the lactis phenotype (*L. lactis* subsp. *lactis*) made up of 110 dairy starter strains and 87 isolated from various sources. With the exception of a group of cultures used in lactic casein manufacture that have been described previously (Ward et al. 2004), the *L. lactis* subsp. *lactis* strains showed much more diversity than the other groups and the majority of strains gave unique PFGE patterns.

Two of the sequenced *L. lactis* strains have the lactis phenotype, MG1363 (Wegmann et al. 2007) and KF147 (Siezen et al. 2010). MG1363 was made plasmid free by UV treatment and protoplast-induced curing (Gasson 1983) and belongs to a group of related strains, which includes NCDO712, C2, ML3, LM0230, and 952 (Davies et al. 1981; Lucey et al. 1993; Le Bourgeois et al. 2000). All have

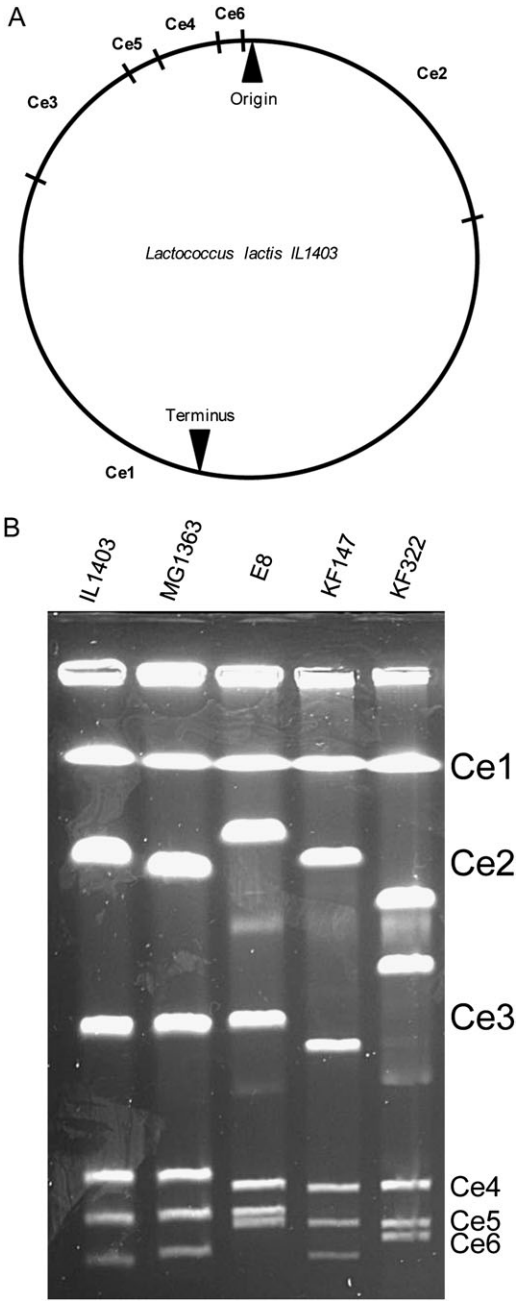

Fig. 3.—(*A*) Locations of I-*Ceu*I recognition sites on the *Lactococcus lactis* IL1403 chromosome. I-*Ceu*I cleaves at sites within the six 23S rRNA genes whose map positions are indicated. The resulting restriction fragments are designated Ce1 through Ce6. Their order in IL1403 and the majority of other strains is Ce2-Ce1-Ce3-Ce5-Ce4-Ce6. (*B*) PFGE patterns of genomic DNA from *L. lactis* strains.

a lactis phenotype but a *cremoris* genotype. NCDO712, C2, and MG1363 have a chromosomally integrated sex factor not found in other lactococcal strains. Strains belonging to this group have been widely used as conjugation recipients, and large DNA fragments are known to be capable of integration at several sites on the chromosome, but techniques developed for MG1363 have been difficult to transfer to dairy starter strains (Johansen 2003). KF147 is one of a group of *L. lactis* cultures isolated from minimally processed fruit and vegetable products (Kelly et al. 1998a) and has several novel properties not found in dairy starters.

Selection of Bacterial Strains for Chromosomal Analysis

Based on the *Sma*I PFGE patterns, 80 strains (table 1) were chosen as representative of the *L. lactis* species. These included 20 strains belonging to each of four groups chosen for comparison because of their origin or because of their use for different purposes in the dairy industry and also included the four strains whose genome sequence has been determined. These are 1) *L. lactis* subsp. *cremoris* dairy starter cultures, 2) *L. lactis* subsp. *lactis* biovar *diacetylactis* dairy starter cultures, 3) *L. lactis* subsp. *lactis* dairy starter cultures, and 4) wild-type *L. lactis* subsp. *lactis* strains. Based on PFGE patterns and strain history data, the cultures selected were believed to be unrelated to one another except for SK11, which is a bacteriophage-resistant derivative of *L. lactis* subsp. *cremoris* AM1, and IL1403 and MG1363, which are plasmid-free derivatives of *L. lactis* subsp. *lactis* biovar *diacetylactis* CNRZ157 and *L. lactis* subsp. *lactis* NCDO712, respectively.

Characteristics of the *L. lactis* Chromosome and Chromosomal Rearrangements

Chromosomal mapping has shown that *L. lactis* has a circular chromosome (fig. 3*A*) with six ribosomal operons that are transcribed divergently from the origin of chromosomal replication (Davidson et al. 1996). Whereas it is expected that in any one strain, all six 16S rRNA copies will have the same nucleotide sequence, work by Pillidge et al. (2009) has highlighted an additional level of complexity. A small number of *L. lactis* subsp. *cremoris* strains, some of which show PFGE patterns similar to SK11, appear to be genotypic hybrids and have both *cremoris*-like and *lactis*-like 16S rRNA types in their genome. Some *L. lactis* subsp. *cremoris* strains contain plasmids with a *lactis*-like 16S rRNA pseudogene, and it is proposed that these chimeric strains are the result of homologous recombination between the pseudogene and the corresponding chromosome gene (Pillidge et al. 2009).

Because plasmid DNA contributes to the PFGE patterns resulting from *Sma*I digests (Ward et al. 1993), the homing endonuclease I-*Ceu*I, which cuts only within the 23S rRNA gene (Liu et al. 1999), was used to produce a PFGE pattern based on chromosomal DNA alone. Macrorestriction patterns produced by PFGE of I-*Ceu*I digests of genomic DNA (fig. 3*B*) provide information on chromosomal size, the number and position of rRNA operons, and an indication of chromosomal rearrangements or insertions and deletions. All the 80 *L. lactis* strains examined gave six fragments

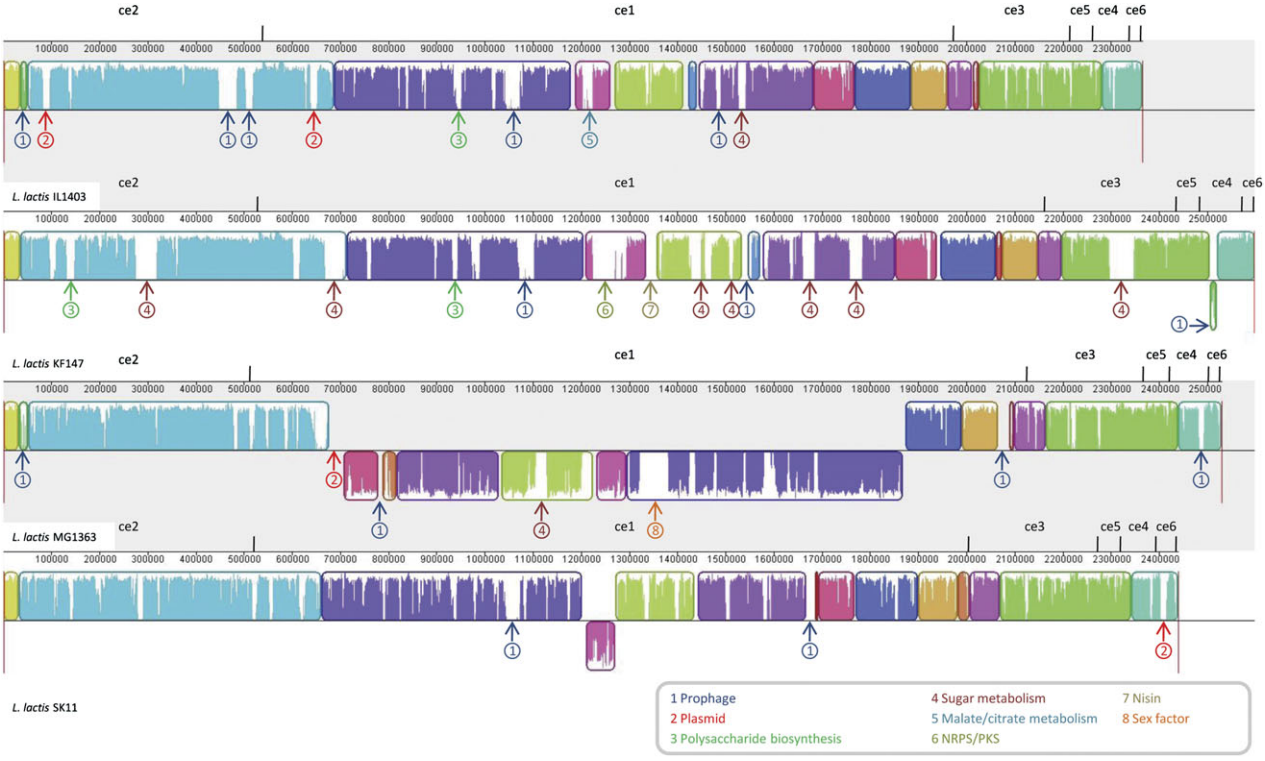

Fig. 4.—Alignment of the chromosomes of *Lactococcus lactis* KF147, IL1403, MG1363, and SK11. Colored blocks surround a section of the genome sequence that aligns to part of another genome. Inverted regions are depicted as blocks below the genome's center line. Inside each block, Mauve draws a similarity profile of the genome sequence. The height of the similarity profile corresponds to the average level of conservation in that region of the genome sequence. Regions outside the blocks, or shown as white space, lack detectable homology with the other genomes and contain sequence elements specific to that strain. The locations of the six I-*Ceu*I cut sites that indicate the locations of the 23S rRNA genes are shown above each strain.

when their genomic DNA was digested with I-*Ceu*I, indicating that the copy number of the rRNA genes is conserved in this species (table 1). These fragments were designated as Ce1–Ce6 following the nomenclature used by Le Bourgeois et al. (1992). Partial digests were used to determine the order of the I-*Ceu*I fragments, and for most strains, the relative size and order of the various fragments are the same as in IL1403, suggesting that chromosomal structure is conserved in most cases. A minority of strains showed chromosomal rearrangements, and these were of two types typified by strains HP and FG2. From the 289 cremoris strains investigated, 21 had PFGE patterns similar to HP and 15 had patterns similar to FG2. The rearrangement in FG2 has been described previously during chromosome mapping studies (Davidson et al. 1996). Curiously, strains HP and FG2 both carry plasmids that specify the same uncommon type of cell envelope proteinase (lactocepin) linked to a partially deleted copy of *abi*B (Christensson et al. 2001). These two strains also cluster together and separate from strains SK11 and AM2 in the CGH study reported by Bayjanov et al. (2009), and both HP and Z8 (a culture with the same atypical PFGE pattern as FG2) were the only *L. lactis* strains shown to lack the *bus*A operon in the osmolality study described by Obis et al. (2001). The *L. lactis* subsp. *lactis* biovar *diacetylactis* strain LW3079 isolated from a mixed strain starter culture has a rearrangement similar to that found in the HP-like strains. These chromosomal changes have no observable effect on cell growth, morphology, or phenotype.

With the availability of four different *L. lactis* genome sequences, it is possible to get an indication of the larger scale events that shape the lactococcal genome. Figure 4 compares the four genomes and shows that overall there is a high degree of conservation. A large inversion involving approximately half the chromosome has been described in MG1363 (Daveran-Mingot et al. 1998), although this occurs within the Ce1 fragment and does not result in a change to the I-*Ceu*I pattern (fig. 4) or alteration in chromosomal symmetry. Major genome insertions in these strains are highlighted in figure 4 and are predominantly associated with prophages, the integration of plasmid genes, polysaccharide biosynthesis, or the ability to metabolize plant-derived carbohydrates. Prophages are an important feature, as phage resistance has been a major driver for strain selection programs for dairy cultures, and phage challenge has been a continual selective stress in their environment. Strain-specific genes of significance include the

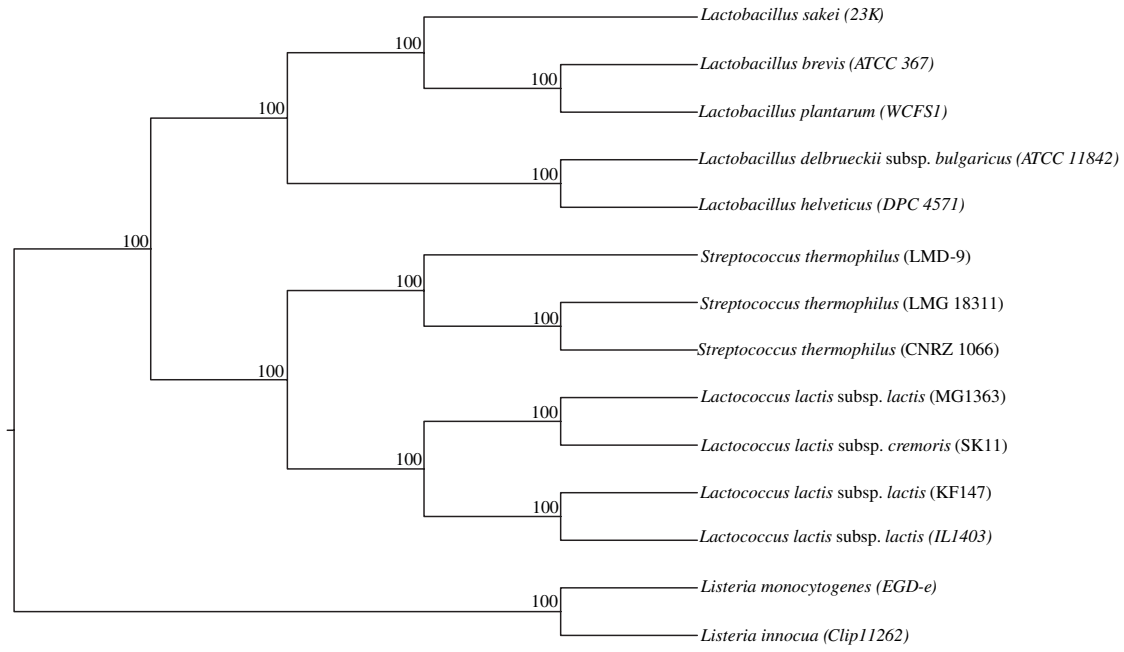

Fig. 5.—maximum representation with parsimony (MRP) supertree for *Lactococcus lactis* and other lactic acid bacteria derived from 1,160 single gene families. *Listeria* species were selected as an outgroup. Bootstrap scores for all nodes are displayed.

malate–citrate metabolism genes in IL1403 and the integrated sex factor in MG1363. KF147 also contains genes for nisin biosynthesis, although this strain does not produce nisin (Kelly et al. 1998a; Siezen et al. 2008), and for nonribosomal peptide and polyketide synthesis. The large insertion in the Ce3 fragment in KF147 (fig. 4) is a chromosomally integrated conjugative element that encodes the ability to metabolize alpha-galactosides such as melibiose and raffinose. Transfer of this element to a derivative of MG1363 and its integration at two different chromosomal sites have been described previously (Kelly et al. 1998b), and similar conjugative elements were found in several *L. lactis* strains of plant origin.

Phylogenetic Relationship between *L. lactis* Strains

The availability of four complete genome sequences covering most genotype/phenotype combinations makes it possible to produce a phylogeny truly representative of the entire genome. Supertree methods (Fitzpatrick et al. 2006) were used to derive phylogenies from 1160 single gene families with the results shown in figure 5. This analysis strongly supports the conclusion from the study by Rademaker et al. (2007) that two main lineages exist in *L. lactis*. These correspond to the two genotypes. Strains with a *L. lactis* subsp. *cremoris* genotype include strains with both lactis (MG1363) and cremoris (SK11) phenotypes, whereas strains with the *L. lactis* subsp. *lactis* genotype includes strains with the both diacetylactis (CNRZ157, the parent strain of IL1403) and lactis (KF147) phenotypes. From the supertree results, it appears that a similar situation may exist in *S. thermophilus*, but this awaits further study.

Differences in Chromosome Size between Subgroups of *L. lactis*

The average chromosome size of the 80 strains of *L. lactis* was 2,483 kb, with the chromosomes of individual strains ranging in size from 2,242 to 2,688 kb. This variation in chromosomal length (~20% of the size of the smallest chromosome) is similar to that found in natural isolates of *Escherichia coli* (Bergthorsson and Ochman 1998). Table 2

Table 2
Mean Chromosome Lengths (kb) of *Lactococcus lactis* Strains Belonging to the Various Groups

Genotype	***cremoris***	***lactis***	***lactis***	***cremoris***	***lactis***	***cremoris***
Phenotype	**Cremoris**	**Diacetylactis**	**Lactis**	**Lactis**	**Lactis**	**Lactis**
Source	**Dairy**	**Dairy**	**Dairy**	**Wild type**	**Wild type**	**Dairy**
Mean chromosome length (kb)	**2,412[a]**	**2,457[b]**	**2,471[b]**	**2,568[c]**	**2,568[c]**	**2,591[c]**
Number of strains	**20**	**20**	**15**	**5**	**15**	**5**

Treatments that share the same letter are not significantly different at $P < 0.05$.

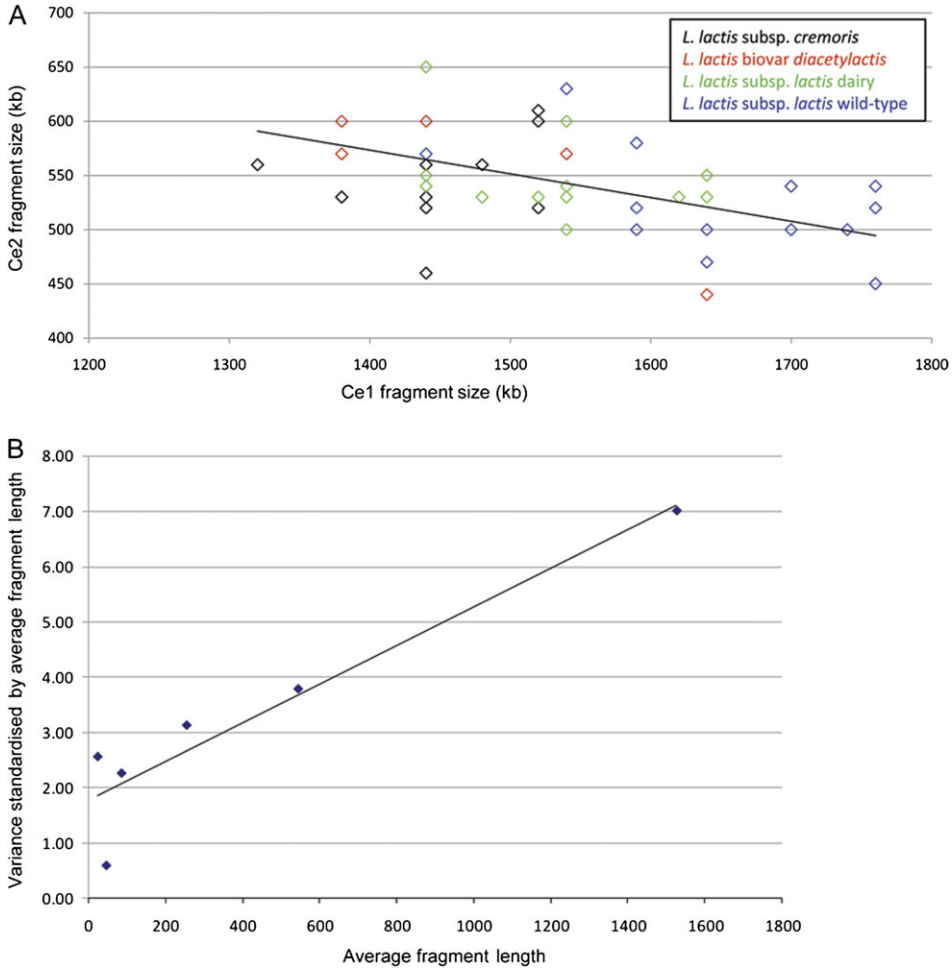

Fig. 6.—(*A*) Relationship between the lengths of the Ce1 and Ce2 chromosomal regions of *Lactococcus lactis*. (*B*) Relationship between variances of different I-*Ceu*I fragments standardized by their average size and the average size of the corresponding fragments.

shows the comparison between the mean chromosomal lengths of *L. lactis* strains from different genotype–phenotype origin combinations. The origin of the strains has the greatest influence on chromosome length with dairy strains having smaller chromosomes than the wild-type strains. This smaller chromosome size may be the result of a process of reductive genome evolution as a consequence of the adaptation to growth in milk. Among the dairy strains, those with the cremoris genotype and phenotype are significantly different and have the smallest chromosomes, whereas there is no significant difference between strains with the diacetylactis and lactis phenotypes. In this analysis, the five cremoris genotype/lactis phenotype dairy strains grouped with the wild-type strains. This could be related to the small sample size or could indicate that they are not as strongly adapted to the dairy environment and are closer to strains of plant origin. That MG1363 has much greater ability than either IL1403 or SK11 to grow on plant-derived carbohydrates (Wegmann et al. 2007) supports the latter option.

Variation among Chromosomal Regions

All the I-*Ceu*I fragments show some degree of length variation (table 1), although, except for the strains that show chromosomal rearrangements, the sizes of the individual fragments do not overlap. To test whether some chromosomal regions are more variable than others, we compared the standardized variances for each fragment across strains in pairwise *F* tests. The four strains that showed major chromosomal rearrangements (HP, 2188, FG2, and LW3079) were not included so this comparison was based on 76 strains. There were no significant differences when most of the fragments were compared against each other; however, the largest fragment (Ce1) was significantly more variable ($P \leq 0.005$), and it can be seen from figure 4 that the majority of chromosomal insertions are found in this region. Ce5 was significantly less variable ($P \leq 0.001$) than the other fragments. A feature of the size variation in the I-*Ceu*I fragments is a correlation between the two largest fragments (fig. 6*A*). This suggests that

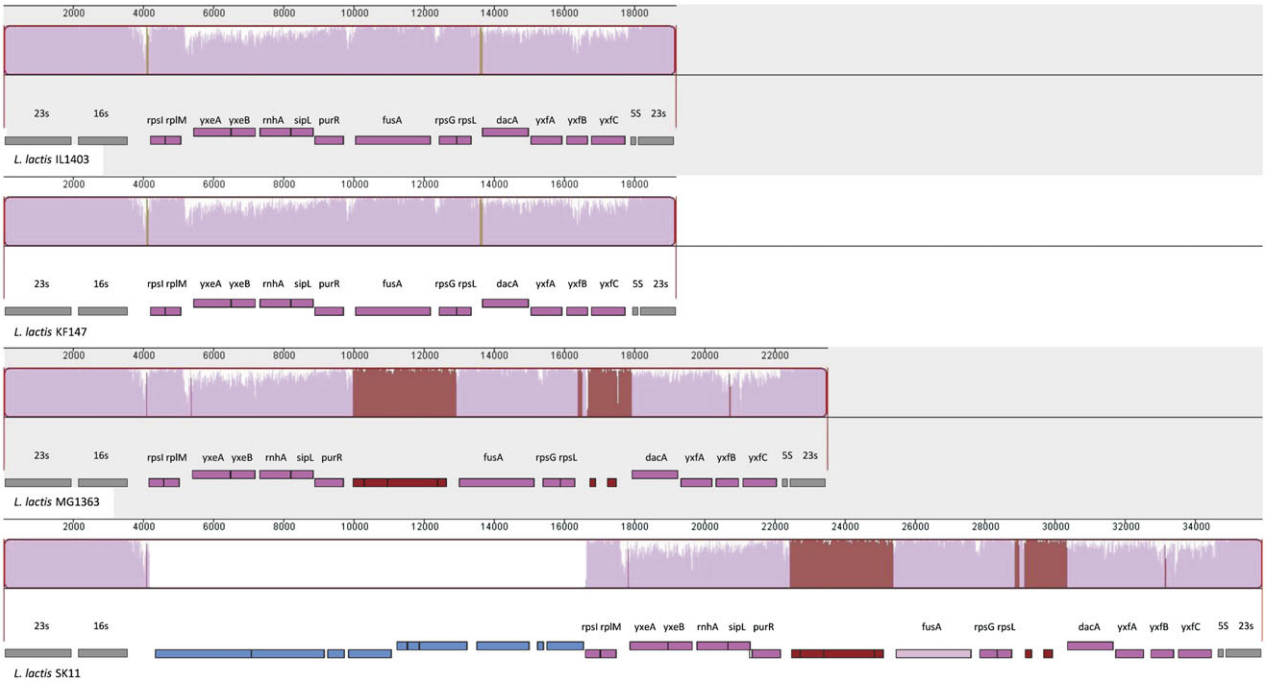

Fig. 7.—Alignment of the Ce6 region of the chromosomes of *Lactococcus lactis* IL1403, KF147, MG1363, and SK11 and identification of the genes present. Insertions common to the cremoris strains MG1363 and SK11 are shown in red, and the genes found only in SK11 are shown in blue. The *fus*A pseudogene in SK11 is shown in mauve.

chromosomal insertions and deletions in these regions are not entirely independent events and that maintenance of chromosomal symmetry may be an important consideration.

There is a significant correlation between the I-*Ceu*I fragment size and the standardized variance of each fragment (fig. 6*B*) with the two smallest fragments showing the greatest departure from this relationship. The smallest fragment (Ce6) shows greater variance than expected, whereas the variance in size of the second smallest fragment (Ce5) is less than for any of the other fragments. The size of Ce5 is strongly conserved (45–50 kb), and we have found only one strain (LW1509) that contains a large insertion of DNA in this region.

Gene Content of Ce6

Because Ce6 exhibits larger variation than expected for a fragment of this size, we examined the genes present in this region from the four available *L. lactis* genome sequences (fig. 7). MG1363 and SK11 (both *L. lactis* subsp. *cremoris* genotype) show two common insertions relative to IL1403 and KF147. These are of genes of unknown function between *pur*R and *fus*A and between *rps*L and *dac*A. In IL1403, 5 of the 14 genes in this region are predicted to be highly expressed (Karlin et al. 2004). These are *fus*A (predicted to be the most highly expressed gene in the IL1403 genome) and four genes encoding ribosomal proteins (*rps*I, *rpl*M, *rps*G, and *rps*L). Curiously, in the SK11 genome sequence, the *fus*A homologue (LACR_2595) is identified as a pseudogene.

The main difference in the genome sequences for this region is a 15-kb insertion in SK11 between the ribosomal protein S9 gene (*rps*I) and the 16S rRNA gene. This insertion contains two genes with high homology to type III restriction–modification systems found in other lactic acid bacteria, several hypothetical proteins, an integrase, and genes related to plasmid replication. It is probable that this insertion increases the phage resistance of SK11, and it may have been acquired by horizontal gene transfer, and subsequent chromosomal integration, of a small plasmid. It is apparent that many dairy starter strains have an insertion of similar size in the Ce6 region (table 1), and it will be of interest to determine if similar genes are found in other strains and whether they have an influence on cell growth rate given their location close to the chromosomal origin.

Contribution of Extrachromosomal Elements to Genome Size

Many technologically important properties (lactose metabolism, lactocepin proteinase, citrate permease, and bacteriophage resistance) are plasmid encoded in *L. lactis* strains used as dairy cultures and can be exchanged between strains by conjugation and between replicons by insertion sequence elements. Examination of plasmid gels

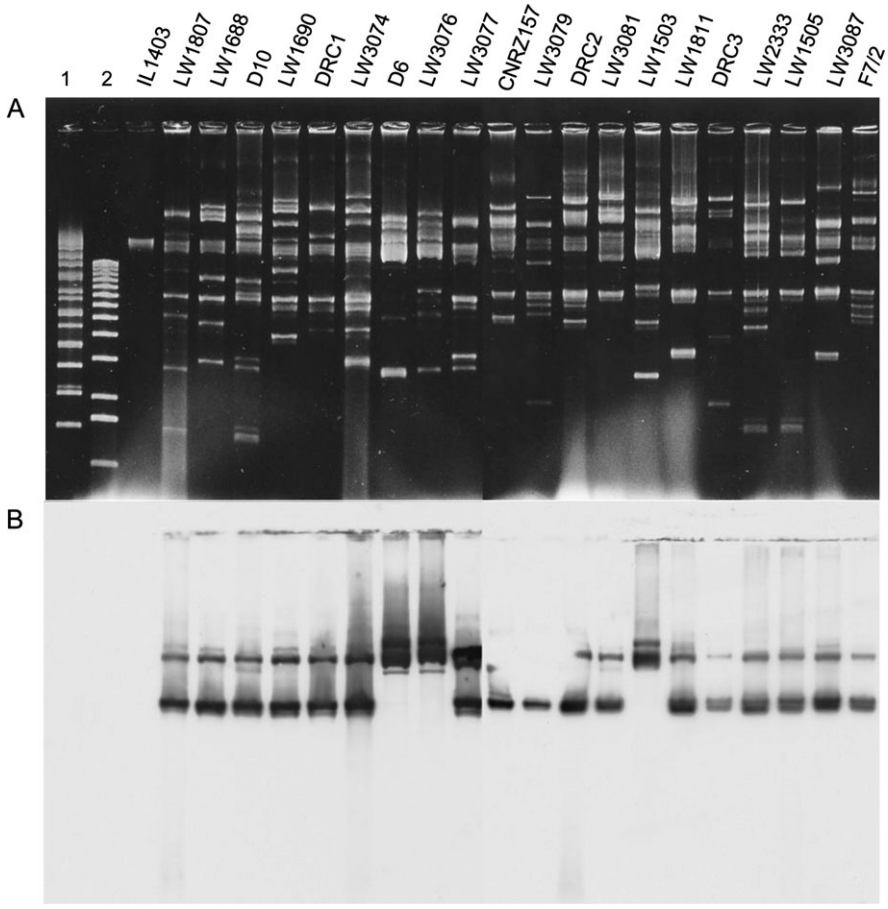

Fig. 8.—(*A*) Plasmid profile of Cit$^+$ *Lactococcus lactis* strains. Lane 1, Invitrogen supercoiled DNA ladder; lane 2, Invitrogen 1 kb DNA ladder. (*B*) Southern blot of (*A*) hybridized with a PCR-amplified product of the *cit*P gene. Where two bands are present, they represent open and closed circular forms of the same plasmid.

for 150 dairy starter cultures (90 *L. lactis* subsp. *cremoris*, 30 *L. lactis* subsp. *lactis*, and 30 *L. lactis* subsp. *lactis* biovar *diacetylactis*) gave an average of seven plasmids per strain (range 2–14) with up to 200 kb of plasmid DNA. By contrast, in plant strains, the average was <2 plasmids per strain (range 0–4). It was notable that small plasmids (<10 kb) were prevalent in dairy cultures but rare in the plant strains. Intact genomic DNA was run on PFGs but showed no evidence of large linear plasmids in these strains.

Plasmid profiles for 20 Cit$^+$ strains are shown in figure 8*A* and show that these strains have acquired a range of different extrachromosomal elements. A plasmid-encoded citrate permease is required for citrate utilization in *L. lactis*, and to determine its location, a Southern blot of the plasmid gel was probed with a PCR-amplified product of the *cit*P gene (fig. 8*B*). The citrate permease plasmid is conserved in size (8 kb) in most strains, but three strains (D6, LW3076, and LW1503) have an enlarged citrate plasmid of ~15 kb. We observed that the plasmid complement can be very unstable in some strains, and this spontaneous instability has previously been reported for CNRZ157 (Chopin et al. 1984).

Origin of Defined Strain Dairy Starter Cultures

Most of the *L. lactis* strains used in this study were dairy starter cultures and are representative of the cultures employed by the dairy industry since the concept of using defined strain starters was developed in the 1930s (Limsowtin et al. 1996). During this period, the best cultures were freely shared between laboratories, and this coupled with the repeated isolation of individual strains with particular characteristics from commercial mixed strain starters, and the development of bacteriophage-insensitive cultures has resulted in many closely related strains coexisting. Consequently, the relationship between strains is generally not known, although it has been speculated that the pool of strains with certain properties is small (Lawrence et al. 1978). Because of their differing histories of industrial use, even closely related strains may differ in some of their

important characteristics and in their plasmid complement (Ward et al. 2004).

The aim of this work was to use PFGE to gain a measure of the chromosomal diversity present in a large collection of *L. lactis* cultures including both plant and wild-type strains, with emphasis on the *L. lactis* strains with the cremoris and diacetylactis phenotypes that are of particular importance to the dairy industry. The diacetylactis phenotype is unusual in that it depends on the inheritance of both plasmid and chromosomal components for citrate to be transported into the cell and metabolized. The chromosomal genes are not found in the other sequenced strains and show homology with the plasmid-encoded citrate genes from *Leuconostoc* and *Weissella* species isolated from the dairy environment. Therefore, it can be hypothesized that the diacetylactis strains used as defined strain dairy starters represent a single lineage within *L. lactis* in which both chromosomal and plasmid elements are maintained and essential for an industrially significant phenotype. It should be noted that this observation is restricted to the Cit^+ strains used as defined strain dairy starter cultures. Wild-type *L. lactis* may differ in their citrate metabolism, and strains with a 23-kb plasmid-encoding genes that match those found in citrate-metabolizing leuconostocs have been isolated from Algerian dromedary's milk (Drici et al. 2010).

The cremoris phenotype is rather more complex to determine because it is measured as negative attributes including the inability to metabolize arginine and inability to grow at higher temperatures or at higher salt levels. From investigations of the genome sequences, it appears that these properties have arisen through the accumulation of mutations and that they are a response to the nutrient rich milk environment where certain gene functions are not longer required. The arginine deiminase–negative phenotype of SK11 correlates with a single base pair deletion (Wegmann et al. 2007), whereas various defects including the absence of the *bus*A operon have been shown to influence salt tolerance (Obis et al. 2001). The taxonomy of *L. lactis* is currently based on phenotype, but as the genomic basis for these phenotypic differences is elucidated, a case could be made for review and use of genotypic data to define the two subspecies.

Comparison of our data with the analysis of the genome-sequenced *L. lactis* strains strongly supports a plant-associated origin for dairy starter strains. A significant proportion of the KF147 genome is devoted to genes involved in the degradation and metabolism of plant-derived carbohydrates, and there is clear evidence for remnants of these capabilities in the genomes of dairy strains, most notably MG1363. This reductive evolution can be seen as part of the adaptation to growth in milk and accounts for the smaller chromosome size in dairy starters as compared with wild-type strains. Similar observations have been made for *Lactobacillus bulgaricus* (Van de Guchte et al. 2006) and *S. thermophilus* (Hols et al. 2005) dairy cultures. We conclude that dairy starter cultures generally and especially strains with cremoris and diacetylactis phenotypes comprise a specialized group of *L. lactis* strains. This is in accordance with the results from the recent studies of *L. lactis* (Rademaker et al. 2007; Bayjanov et al. 2009) and fits with knowledge of the origin of industrially used dairy starter cultures. Consequently, the world's dairy industry is based on the same small group of good starter strains, and these cultures have transitioned from free-living organisms associated with plant material. We hypothesize that these specialized dairy starters are no longer fit to survive outside the dairy environment and have evolved to become essential components of industrial processes. Whereas wild-type strains may have genes that can be used to enhance the metabolism of dairy strains (van Hylckama Vlieg et al. 2006), it is unlikely that there is an environmental source for new dairy starter cultures similar to those currently in use.

This analysis together with other studies (Nomura et al. 2006; Rademaker et al. 2007; Siezen et al. 2008; Bachmann et al. 2009; Bayjanov et al. 2009; Liu et al. 2010) begins to illustrate the genomic and phenotypic diversity present within *L. lactis*. Currently, too few genome sequences are available to delineate sets of core and auxillary genes and describe the pangenome of *L. lactis*, but in the future, it will be interesting to examine strains from other environments in more detail and to further define the genes necessary to make a good dairy starter culture.

Funding

New Zealand Foundation for Research, Science, and Technology (Contract C10X0239).

Acknowledgments

We thank Zaneta Park for assistance with the statistical analysis and Dr David Fitzpatrick from the Genome Evolution Laboratory, National University of Ireland, Maynooth, Ireland, for assistance with the supertree phylogenetic analysis. We also wish to thank Howard Heap and the members of the Microbial Fermentation Unit, Fonterra, Palmerston North, for their support and advice on lactococcal dairy starter cultures.

Literature Cited

Anderson DG, McKay LL. 1983. Simple and rapid method for isolating large plasmid DNA from lactic streptococci. Appl Environ Microbiol. 46:549–552.

Bachmann H, et al. 2009. Regulatory phenotyping reveals important diversity within the species *Lactococcus lactis*. Appl Environ Microbiol. 75:5687–5694.

Bayjanov JR, et al. 2009. PanCGH: a genotype-calling algorithm for pangenome CGH data. Bioinformatics. 25:309–314.

Beimfohr C, Ludwig W, Schleifer KH. 1997. Rapid genotypic differentiation of *Lactococcus lactis* subspecies and biovar. Syst Appl Microbiol. 20:216–221.

Bergthorsson U, Ochman H. 1998. Distribution of chromosome length variation in natural isolates of *Escherichia coli*. Mol Biol Evol. 15:6–16.

Bolotin A, et al. 2001. The complete genome sequence of the lactic acid bacterium *Lactococcus lactis* ssp. *lactis* IL1403. Genome Res. 11:731–753.

Bourel G, Bekal S, Diviès C, Prévost H. 1996. Citrate permease gene expression in *Lactococcus lactis* subsp. *lactis* strains IL1403 and MG1363. FEMS Microbiol Lett. 145:367–370.

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

Chopin A, Chopin M-C, Moillo-Batt A, Langella P. 1984. Two plasmid-determined restriction and modification systems in *Streptococcus lactis*. Plasmid. 11:260–263.

Christensson C, Pillidge CJ, Ward LJH, O'Toole PW. 2001. Nucleotide sequence and characterization of the cell envelope proteinase plasmid in *Lactococcus lactis* subsp. *cremoris* HP. J Appl Microbiol. 91:334–343.

Collins EB. 1961. Domination among strains of lactic streptococci with attention to antibiotic production. Appl Microbiol. 9:200–205.

Czulak J, Hammond LA. 1954. Compatibility among strains of lactic streptococci. Aust J Dairy Technol. 9:15–18.

Darling ACE, Mau B, Blattner FR, Perna NT. 2004. Mauve: multiple alignment of conserved genomic sequence rearrangements. Genome Res. 14:1394–1403.

Daveran-Mingot M-L, Campo N, Ritzenthaler P, Le Bourgeois P. 1998. A natural large chromosomal inversion in *Lactococcus lactis* is mediated by homologous recombination between two insertion sequences. J Bacteriol. 180:4834–4842.

Davidson BE, Kordias N, Dobos M, Hillier AJ. 1996. Genomic organization of lactic acid bacteria. Antonie van Leeuwenhoek. 70:161–183.

Davies FL, Underwood HM, Gasson MJ. 1981. The value of plasmid profiles for strain identification in lactic streptococci and the relationship between Streptococcus lactis 712, ML3 and C2. J Appl Bacteriol. 51:325–337.

[DRI] Dairy Research Institute. 1951. Starter news. Dairy Research Institute (NZ) News Circular 6(4).

[DRI] Dairy Research Institute. 1960. Cheese starters. Dairy Research Institute (NZ) News Circular 15(3).

[DRI] Dairy Research Institute. 1962. Cheese starters. Dairy Research Institute (NZ) News Circular 17(3).

Drici H, Gilbert C, Kihal M, Atlan D. 2010. Atypical citrate-fermenting *Lactococcus lactis* strains isolated from dromedary's milk. J Appl Microbiol. 108:647–657.

Drider D, Bekal S, Prévost H. 2004. Genetic organization and expression of citrate permease in lactic acid bacteria. Genet Mol Res. 3:273–281.

Fitzpatrick DA, Logue ME, Stajich JE, Butler G. 2006. A fungal phylogeny based on 42 complete genomes derived from supertree and combined gene analysis. BMC Evol Biol. 6:99.

Gasson MJ. 1983. Plasmid complements of *Streptococcus lactis* NCDO 712 and other lactic streptococci after protoplast-induced curing. J Bacteriol. 154:1–9.

Harvey RJ, Collins EB. 1961. Role of citritase in acetoin formation by *Streptococcus diacetilactis* and *Leuconostoc citrovorum*. J Bacteriol. 82:954–959.

Heap HA. 1998. Optimising starter culture performance in NZ cheese plants. Aust J Dairy Tech. 53:74–78.

Heap HA, Lawrence RC. 1984. The development of a defined starter system for casein manufacture. N Z J Dairy Sci Technol. 19:119–123.

Hols P, et al. 2005. New insights in the molecular biology and physiology of *Streptococcus thermophilus* revealed by comparative genomics. FEMS Microbiol Rev. 29:435–463.

Jarvis AW, Wolff JM. 1979. Grouping of lactic streptococci by gel electrophoresis of soluble cell extracts. Appl Environ Microbiol. 37:391–398.

Johansen E. 2003. Challenges when transferring technology from Lactococcus laboratory strains to industrial strains. Genet Mol Res. 2:112–116.

Karlin S, Theriot J, Mrázek J. 2004. Comparative analysis of gene expression among low G+C gram-positive genomes. Proc Natl Acad Sci USA. 101:6182–6187.

Kelly WJ, Asmundson RV, Moore DG, Huang CM. 1994. Microbiological investigation of Kaanga wai. In: Sanchez PC, editor. Advances in microbial processes for the utilization of tropical raw materials in the production of food products. Bangkok: UNESCO Regional Network for Microbiology in Southeast Asia. pp. 135–146.

Kelly WJ, Davey GP, Ward LJH. 1998a. Characterization of lactococci isolated from minimally processed fresh fruit and vegetables. Int J Food Microbiol. 45:85–92.

Kelly WJ, Davey GP, Ward LJH. 1998b. Conjugative transfer of raffinose metabolism in *Lactococcus lactis*. FEMS Microbiol Lett. 167:145–149.

Kempler GM, McKay LL. 1981. Biochemistry and genetics of citrate utilization in *Streptococcus lactis* ssp. *diacetylactis*. J Dairy Sci. 64:1527–1539.

Klijn N, Weerkamp AH, De Vos WE. 1995. Detection and characterization of lactose-utilizing *Lactococcus* spp. in natural ecosystems. Appl Environ Microbiol. 61:788–792.

Köhler G, Ludwig W, Schleifer KH. 1991. Differentiation of lactococci by rRNA gene restriction analysis. FEMS Microbiol Lett. 84:307–312.

Lan R, Reeves PR. 2000. Intraspecies variation in bacterial genomes: the need for a species genome concept. Trends Microbiol. 8:396–401.

Lawrence RC, Heap HA, Limsowtin G, Jarvis AW. 1978. Cheddar cheese starters: current knowledge and practices of phage characteristics and strain selection. J Dairy Sci. 61:1181–1191.

Lawrence RC, Pearce LE. 1972. Cheese starters under control. Dairy Ind. 37:73–78.

Le Bourgeois P, Daveran-Mingot M-L, Ritzenthaler P. 2000. Genome plasticity among related *Lactococcus* strains: identification of genetic events associated with macrorestriction polymorphisms. J Bacteriol. 182:2481–2491.

Le Bourgeois P, Lautier M, Mata M, Ritzenthaler P. 1992. Physical and genetic map of the chromosome of *Lactococcus lactis* subsp. *lactis* IL1403. J Bacteriol. 174:6752–6762.

Limsowtin GKY, Powell IB, Parente E. 1996. Types of starters. In: Cogan TM, Accolas J-P, editors. Dairy starter cultures. New York: VCH Publishers. pp. 101–129.

Limsowtin GKY, Terzaghi BE. 1976. Phage resistant mutants: their selection and use in cheese factories. N Z J Dairy Sci Technol. 11:251–256.

Liu M, Bayjanov JR, Renckens B, Nauta A, Siezen RJ. 2010. The proteolytic system of lactic acid bacteria revisited: a genomic comparison. BMC Genomics. 11:36.

Liu S-L, Schryvers AB, Sanderson KE, Johnston RN. 1999. Bacterial phylogenetic clusters revealed by genome structure. J Bacteriol. 181:6747–6755.

Lucey M, Daly C, Fitzgerald G. 1993. Relationship between *Lactococcus lactis* subsp. *lactis* 952, an industrial lactococcal starter culture and strains 712, ML3 and C2. J Appl Bacteriol. 75:326–335.

Makarova K, et al. 2006. Comparative genomics of the lactic acid bacteria. Proc Natl Acad Sci USA. 103:15611–15616.

Martley FG, Lawrence RC. 1972. Cheddar cheese flavour. II Characteristics of single strain starters associated with good or poor flavour development. N Z J Dairy Sci Technol. 7:38–44.

McGrath S, Fitzgerald GF, van Sinderen D. 2002. Identification and characterization of phage-resistance genes in temperate lactococcal bacteriophages. Mol Microbiol. 43:509–520.

Medini D, Donati C, Tettelin H, Masignani V, Rappuoli R. 2005. The microbial pan-genome. Curr Opin Genet Dev. 15:589–594.

Nomura M, Kobayashi M, Narita T, Kimoto-Nira H, Okamoto T. 2006. Phenotypic and molecular characterization of *Lactococcus lactis* from milk and plants. J Appl Microbiol. 101:396–405.

Obis D, Guillot A, Mistou MY. 2001. Tolerance to high osmolality of *Lactococcus lactis* subsp. *lactis* and *cremoris* is related to the activity of a betaine transport system. FEMS Microbiol Lett. 202:39–44.

Pillidge CJ, et al. 2009. Intragenomic 16S rRNA gene heterogeneity in *Lactococcus lactis* subsp. *cremoris*. Int Dairy J. 19:222–227.

Quénée P, Lepage E, Kim WS, Vergnaud G, Gruss A. 2005. Minisatellite polymorphism as a tool to distinguish closely related *Lactococcus lactis* strains. FEMS Microbiol Lett. 248:101–109.

Rademaker JLW, et al. 2007. Diversity analysis of dairy and nondairy *Lactococcus lactis* isolates, using a novel multilocus sequence analysis scheme and $(GTG)_5$-PCR fingerprinting. Appl Environ Microbiol. 73:7128–7137.

Salama MS, Musafija-Jeknic T, Sandine WE, Giovannoni SJ. 1995. An ecological study of lactic acid bacteria: isolation of new strains of *Lactococcus* including *Lactococcus lactis* subspecies *cremoris*. J Dairy Sci. 78:1004–1017.

Sambrook J, Fritsch EF, Maniatis T. 1989. Molecular cloning. A laboratory manual, 2nd ed. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press.

Schleifer KH, et al. 1985. Transfer of *Streptococcus lactis* and related streptococci to the genus *Lactococcus* gen. nov. Syst Appl Microbiol. 6:183–195.

Siezen RJ, et al. 2008. Genome-scale genotype-phenotype matching of two *Lactococcus lactis* isolates from plants identifies mechanisms of adaptation to the plant niche. Appl Environ Microbiol. 74:424–436.

Siezen RJ, et al. 2010. Complete genome sequence of *Lactococcus lactis* subsp. *lactis* KF147, a plant-associated lactic acid bacterium. J Bacteriol. 192:2649–2650.

Shannon AL, Attwood G, Hopcroft DH, Christeller JT. 2001. Characterization of lactic acid bacteria in the larval midgut of the keratinophagous lepidopteran, *Hofmannophila pseudospretella*. Lett Appl Microbiol. 32:36–41.

Sokal RR, Rohlf FJ. 1981. Biometry. San Fransisco (CA): Freeman.

Szatmari G, et al. 2006. In vitro expression of the restriction endonucleases LlaMI and ScrFI isolated from *Lactococcus lactis* M19 and UC503. J Biotechnol. 121:144–153.

Taïbi A, Dabour N, Lamoureux M, Roy D, LaPointe G. 2010. Evaluation of the genetic polymorphism among *Lactococcus lactis* subsp. *cremoris* strains using comparative genomic hybridization and multilocus sequence analysis. Int J Food Microbiol. doi:10.1016/j.ijfoodmicro.2010.02.020.

Tailliez P, Tremblay J, Ehrlich SD, Chopin A. 1998. Molecular diversity and relationship within *Lactococcus lactis*, as revealed by randomly amplified polymorphic DNA (RAPD). Syst Appl Microbiol. 21:530–538.

Terzaghi BE, Sandine WE. 1975. Improved medium for lactic streptococci and their bacteriophages. Appl Microbiol. 29:807–813.

Van de Guchte M, et al. 2006. The complete genome sequence of *Lactobacillus bulgaricus* reveals extensive and ongoing reductive evolution. Proc Natl Acad Sci USA. 103:9274–9279.

van Hylckama Vlieg JET, et al. 2006. Exploiting natural diversity and adaptive responses of *Lactococcus lactis* for innovations in food fermentations. Curr Opin Biotechnol. 17:183–190.

Ventura M, et al. 2007. Comparative analyses of prophage-like elements present in two *Lactococcus* strains. Appl Environ Microbiol. 73:7771–7780.

Ward AC, Hillier AJ, Davidson BE, Powell IB. 1993. Stability analysis of the *Lactococcus lactis* DRC1 lactose plasmid using pulsed-field gel electrophoresis. Plasmid. 29:70–73.

Ward LJH, Brown JCS, Davey GP. 1998. Two methods for the genetic differentiation of *Lactococcus lactis* ssp. *lactis* and *cremoris* based on differences in the 16S rRNA gene sequence. FEMS Microbiol Lett. 166:15–20.

Ward LJH, Davey GP, Heap HA, Kelly WJ. 2002. Lactococcus lactis. In: Roginski H, Fuquay JW, Fox PF, editors. Encyclopedia of dairy sciences. London: Elsevier Science Ltd. pp. 1511–1516.

Ward LJH, Heap HA, Kelly WJ. 2004. Characterization of closely related lactococcal starter strains which show differing patterns of bacteriophage sensitivity. J Appl Microbiol. 96:144–148.

Wegmann U, et al. 2007. Complete genome sequence of the prototype lactic acid bacterium *Lactococcus lactis* subsp. *cremoris* MG1363. J Bacteriol. 189:3256–3270.

Associate editor: Takashi Gojobori