Multireplicon genome architecture of *Lactobacillus salivarius*

Marcus J. Claesson*†‡, **Yin Li***†‡, **Sinead Leahy***†, **Carlos Canchaya***†, **Jan Peter van Pijkeren***†, **Ana M. Cerdeño-Tárraga**§, **Julian Parkhill**§, **Sarah Flynn***, **Gerald C. O'Sullivan**¶, **J. Kevin Collins***†, **Des Higgins**∥, **Fergus Shanahan**†**, **Gerald F. Fitzgerald***†, **Douwe van Sinderen***†, and **Paul W. O'Toole***†,††

Departments of *Microbiology and **Medicine and †Alimentary Pharmabiotic Centre, University College Cork, Cork, Ireland; §Pathogen Sequencing Unit, Wellcome Trust Sanger Centre, Cambridge CB10 1SA, United Kingdom; ¶Cork Cancer Research Centre, Mercy University Hospital, Cork, Ireland; and ∥Conway Institute, University College Dublin, Dublin 4, Ireland

Lactobacillus salivarius subsp. *salivarius* strain UCC118 is a bacteriocin-producing strain with probiotic characteristics. The 2.13-Mb genome was shown by sequencing to comprise a 1.83 Mb chromosome, a 242-kb megaplasmid (pMP118), and two smaller plasmids. Megaplasmids previously have not been characterized in lactic acid bacteria or intestinal lactobacilli. Annotation of the genome sequence indicated an intermediate level of auxotrophy compared with other sequenced lactobacilli. No single-copy essential genes were located on the megaplasmid. However, contingency amino acid metabolism genes and carbohydrate utilization genes, including two genes for completion of the pentose phosphate pathway, were megaplasmid encoded. The megaplasmid also harbored genes for the Abp118 bacteriocin, a bile salt hydrolase, a presumptive conjugation locus, and other genes potentially relevant for probiotic properties. Two subspecies of *L. salivarius* are recognized, *salivarius* and *salicinius*, and we detected megaplasmids in both subspecies by pulsed-field gel electrophoresis of sizes ranging from 100 kb to 380 kb. The discovery of megaplasmids of widely varying size in *L. salivarius* suggests a possible mechanism for genome expansion or contraction to adapt to different environments.

megaplasmid | probiotic | heterofermentation

Lactobacilli are widely used for fermenting foods products or as adjuncts to foodstuffs. Many *Lactobacillus* species have been shown to confer probiotic properties, meaning that, upon ingestion, they confer a range of benefits on the host (1). Lactobacilli are part of the normal human gastrointestinal (GI) microbiota, and they may also be found in the GI tracts of other mammalian species (2–4). Bacteria belonging to the species investigated in this study, *Lactobacillus salivarius*, have been isolated from the intestinal mucosa of 9% of human subjects examined (5), the tongue and rectum of 12% of healthy adults sampled (6), and feces of infants (7). *L. salivarius* subsp. *salivarius* strain UCC118 was isolated from the terminal ileum of a (otherwise healthy) patient undergoing urinary tract reconstructive surgery (8). This strain has been extensively studied for its probiotic properties in human trials and animal models (8–11). In addition, *L. salivarius* UCC118 produces a two-component bacteriocin Abp118 (12), which has broad spectrum activity against Gram-positive bacteria, including methicillin-resistant *Staphylococcus aureus* (8).

The application of genomic technologies recently has led to major advances in our understanding of lactobacilli (13, 14), through genome sequence determination for *Lactobacillus plantarum*, *Lactobacillus johnsonii*, *Lactobacillus acidophilus*, and *Lactobacillus sakei* (15–18). *L. plantarum* has the largest *Lactobacillus* genome sequenced to date (3.3 Mb) and has a commensurate biochemical complexity. For example, *L. plantarum* possesses enzymes for the biosynthesis of all amino acids except leucine, isoleucine, and valine, whereas *L. johnsonii* is predicted to be incapable of synthesizing any amino acids (19). *L. sakei*, although displaying the highest levels of orthology with *L. plantarum*, is unable to synthesize 18 amino acids (18). *L. johnsonii* and *L. acidophilus* are more related to each other than to *L. plantarum* (see also below) and have smaller genomes, reflected in nutritional fastidiousness. *L. acidophilus* appears to be auxotropic for 14 amino acids (17). *L. johnsonii* lacks the ability for *de novo* synthesis of purines and cofactors (16, 19). In common with *Streptococcus thermophilus* (20), which has been passed for centuries in the nutrient-rich medium of milk, *L. johnsonii* and *L. acidophilus* appear to have undergone genome reduction in adapting to a lifestyle of close host association.

The genus *Lactobacillus* is very diverse, which is evident from the lack of long-range synteny observed between the four available complete *Lactobacillus* genomes and the assembled genome from *Lactobacillus gasseri* (refs. 13, 14, and 19 and unpublished analyses). The *Lactobacillus* 16S rRNA gene phylogeny (available as Fig. 4, which is published as supporting information on the PNAS web site) shows that *L. salivarius* is part of a distinct clade at the periphery of the genus that is not represented by completed or in-progress genome sequence projects. The genome of *L. salivarius* UCC118 described here reveals the presence of a 242-kb megaplasmid, which, although apparently dispensable for viability based on gene content, confers on the strain a large number of contingency metabolic capabilities and traits directly related to GI tract survival or competitiveness.

Results and Discussion

General Genome Features. The genome sequence of *L. salivarius* subsp. *salivarius* strain UCC118 consists of 2,133,977 nucleotides with an average GC content of 33.04%. The genome comprises four replicons (Fig. 1), a circular chromosome of 1,827,111 nucleotides, a previously undescribed megaplasmid of 242,436 nucleotides designated pMP118, and two previously described plasmids of 20.4 kb (pSF118–20) and 44 kb (pSF118–44) (21). The two smaller plasmids have elevated GC contents (>39%), whereas the GC content of pMP118 is very close to that of the chromosome. The coding densities of the chromosome and pMP118 are 84% and 77%, respectively. Functions could be predicted for 71% of the chromosomal genes, whereas half of the genes on the megaplasmid had no predicted functions, and many

---

Fig. 1. Genome atlas of *L. salivarius* UCC118. The color coding of the genomic features in circle 1 and 2 represents different Clusters of Orthologous Groups categories.

of these genes were clustered in two distinct regions (Fig. 1). Seven rRNA operons and 78 tRNAs, representing all 20 amino acids, were detected in the genome and arranged in 25 clusters.

Multireplicon genome architecture increasingly has become recognized in bacteria (reviewed in refs. 22 and 23) and may include minichromosomes, which are difficult to formally distinguish from megaplasmids (24). We designated pMP118 as a megaplasmid for the following reasons: It contains neither tRNA nor rRNA genes; it has plasmid-related replication and partition proteins; and it does not contain the only copy in the genome of any known essential gene. Plasmid-encoded Rep proteins bind to specific sequences in the plasmid replication origin, initiating replication and facilitating binding of host proteins (25). The *repA* gene of pMP118 (LSL_1739) coincides with the position of a switch in GC skew that is characteristic of replication origins (23, 26). The *repA* gene product has the highest identity (33%) with the RepA/RepE protein of the *Enterococcus faecalis* theta-replicating plasmid pS86 (27). The pMP118 *ori* region includes a second Rep protein, LSL_1740, that is weakly related to staphylococcal plasmid replication proteins. The *ori* region extends to a gene LSL_1741, whose product contains the Pfam motif associated with presumptive chromosome partitioning ATPases, ParA. LSL_1741 is 41% identical to a copy number control protein of the *Listeria innocua* plasmid pLI100. The origin is disrupted by three pseudogenes between LSL_1740 and LSL_1741. RepA proteins usually bind to repetitive motifs called RepA boxes or iterons (25). We could not identify corresponding sequences in pMP118. The megaplasmid also has an asymmetric GC skew pattern (Fig. 1), which may be indicative of recent DNA acquisition or loss (28). Smaller plasmids like pSF118-20 and pSF118-44 are relatively common in lactobacilli (29).

Four extensive regions of bacteriophage-related DNA were detected in the genome (Fig. 1), two of which appeared to be intact prophage and two of which were remnants. The bacteriophage content of *L. salivarius* UCC118 will be described in detail elsewhere (M. Ventura, C.C., V. Bernini, E. Altermann, R. Barrangou, S. McGrath, M.J.C., Y.L., S.L., C. D. Walker, R. Zink, E. Neviani, J. Steele, J. Broadbent, T. R. Klaenhammer, G.F.F., P.W.O., and D.v.S., unpublished data). The other extended region of extensive GC content deviation was related to exopolysaccharide (EPS) production (EPS cluster 2; LSL_1547–

LSL_1574; see below). Alignment of the concatenated *L. salivarius* UCC118 genome against those of *L. plantarum*, *L. johnsonii*, *L. acidophilus*, and *L. sakei* revealed no long-range synteny (Fig. 5, which is published as supporting information on the PNAS web site). The greatest overall number of orthologs was shared with *L. plantarum* (61.1% of *L. salivarius* proteins), followed by *E. faecalis* (50.4%), *L. sakei* (49.8%), *L. acidophilus* (46.3%), and *L. johnsonii* (45.1%).

Intergenic shuffling that modulates restriction-modification (R/M) activity is a recognized phenomenon in lactic acid bacteria and other groups (30, 31). We identified a type I R/M system (LSL_0915–LSL_0920) in the *L. salivarius* genome that appeared to be subject to phase variation controlled by DNA inversion events at intragenic inverted repeat sites (Fig. 6, which is published as supporting information on the PNAS web site). The 9,360-bp shufflon contains genes for two extra complete HsdS subunits, downstream and outside of the R/M operon, that allow for nine combinations of the active HsdS subunit. Interestingly, sequence reads confirmed the presence of all nine HsdS combinations in the clonal growth used for DNA preparation. Based on sequence reads, the combination A″B″ was most abundant at 38%, followed by AB and A′B′ (15% each), AB′ and A′B (6% each), and almost equal representations of other combinations. This sequence read distribution indicates that a culture of *L. salivarius* UCC118 contains a heterogenous collection of different restriction enzymes and differently methylated chromosomes, thus providing a highly effective barrier against bacteriophage infection.

Pseudogenes and Transposable Elements. The genome includes 73 pseudogenes (Table 1, which is published as supporting information on the PNAS web site), defined as genes that have frameshifts, stop codons, large deletions, insertion sequence (IS) insertions, or truncations of at least 10% of the sequence. A disproportionately high fraction (27%) are located on pMP118. Among pseudogenes, genes with unknown function are the most common, followed by transposase-encoding genes, genes predicted to encode surface proteins, restriction/modification systems, and ABC transporters. Frame shifting is the most common cause of gene inactivation. For example, the megaplasmid harbors a full-length *prtP* gene homolog (putatively encoding the

MICROBIOLOGY

major caseinolytic cell-surface protease PrtP; LSL_1774b) that contains a single frameshift.

Transposable elements and, in particular, IS element are common features in bacterial genomes, and transposons have been shown to be means of gene transfer. Scattered across the two larger replicons, we identified 16 different IS elements, present in a total of 43 copies (11 of which are nonfunctional because of frameshifts, truncations, and in-frame stop codons) representing 10 IS families (Table 2, which is published as supporting information on the PNAS web site). The largest number of IS elements belonged to the ISLasa3 homology group, with 21 copies of this IS element uniformly distributed over the chromosome and one additional copy on the megaplasmid. Remaining IS elements were found in either single or double copies. Two of these tandem pairs were present as almost identical arrangements, together with flanking DNA, in both the chromosome and the megaplasmid (LSL_732–LSL_733 corresponding with LSL_1779–LSL_1780, and LSL_0049–LSL_0049b corresponding with LSL_1957–LSL_1958). The megaplasmid contains almost 25% of the IS element content of the genome. It is possible that IS elements contributed to formation of pMP118 by a cointegration mechanism, which was suggested for the formation of an archaeal megaplasmid (24). Overall, *L. salivarius* possesses both the highest number and diversity of IS elements in *Lactobacillus* genomes sequenced to date.

Biosynthetic Capabilities. Major pathways are provided as Figs. 7–14, which are published as supporting information on the PNAS web site. *L. salivarius* has an alanine dehydrogenase gene (EC1.4.1.1; LSL_1768) located on the megaplasmid, which is unique among published *Lactobacillus* genomes but present in *Lactobacillus casei* and *Lactobacillus delbrueckii* draft genomes (www.jgi.doe.gov). This enzyme catalyzes the NAD^+-dependent reversible reductive amination of pyruvate into alanine. It is most frequently found in the genus *Bacillus* and has been exploited for engineering of *L. lactis* to produce L-alanine (32). By virtue of the pMP118-encoded enzyme, L-alanine can be synthesized from pyruvate. Linked to this alanine dehydrogenase gene is a gene encoding a putative alanine permease (LSL_1767); both genes show elevated GC content (40.37% and 41.5% for LSL_1767 and LSL_1768, respectively), indicating lateral transfer. The megaplasmid also encodes a paralog (LSL_1927) for one of the two enzymes required for conversion of pyruvate to L-aspartate.

The megaplasmid also harbors two genes (LSL_1931 and LSL_1932) predicted to encode the alpha and beta subunits of L-serine dehydratase (EC 4.3.1.17), which catalyzes pyruvate-serine interconversion. Serine formed from pyruvate can subsequently be converted to glycine by a chromosomally encoded enzyme. Serine may be thiolated to cysteine by CysK (EC 2.5.1.47; LSL_0026 and LSL_1718), and cysteine can be converted to methionine by using four chromosomally encoded enzymes. Genes whose products are predicted to synthesize aspartate from pyruvate, and lysine and threonine from aspartate, were also annotated. Unlike *L. plantarum*, *L. salivarius* appears to lack the genes required for synthesis of tryptophan and related amino acids. In summary, *L. salivarius* UCC118 can synthesize *de novo* or by interconversion nine amino acids, can convert glutamine to three more, and is theoretically auxotrophic for eight amino acids. This level of auxotrophy makes it clearly more dependent on uptake of extracellular amino acids than *L. plantarum*, less dependent than *L. acidophilus* (auxotrophic for 14), and far less dependent than *L. johnsonii* or *L. sakei* (auxotrophic for 18). However, *L. sakei* has numerous other adaptations for living in the particular nutrient-rich meat environment (18) where extracellular compounds, including amino acids, are freely available.

In *L. salivarius* UCC118, the pentose phosphate pathway (see below) may be used for phosphoribosyl pyrophosphate (PRPP) generation. The *L. salivarius* UCC118 chromosome encodes a complete pathway for the biosynthesis of UMP that can be further converted to UTP and CTP via a pathway apparently requiring the megaplasmid gene LSL_1936 (nucleoside-diphosphate kinase; Dnk). However, as discussed in ref. 33, other enzymes, including pyruvate kinase, can catalyze this reaction. *L. salivarius* UCC118 contains a complete purine biosynthetic pathway for the creation of IMP and the pathway branches AMP and GMP. *L. acidophilus* and *L. johnsonii* are auxotrophic for pyrimidines and purines, respectively (16, 17), whereas *L. plantarum* can synthesize pyrimidines and purines *de novo* (19) from PRPP, as can *L. sakei* (18). Thus, although other "small-genome" intestinal lactobacilli are auxotrophic for one or other nucleotide class, *L. salivarius* UCC118 appears to be independent of a host provision of both.

Peptidoglycan and Exopolysaccharide Biosynthesis. All of the genes involved in peptidoglycan biosynthesis are located on the *L. salivarius* chromosome. Exopolysaccharide (EPS) genes in the *L. salivarius* UCC118 genome are mainly clustered in two regions, at positions 0.99–1.101 Mb and 1.62–1.65 Mb in the chromosome. This arrangement is in contrast to species such as *L. johnsonii* (16), *L. acidophilus* (17), and *Lactobacillus rhamnosus* (34), in which most of the EPS genes are clustered in a single operon. Cluster 1 (LSL_0977–LSL_0997) contains 21 genes that include two putative chain length determinators, an oligosaccharide translocase or flippase, and 12 glycosyltransferases (Fig. 9). A homolog encoding a priming glucose phosphotransferase appears to be absent from this cluster. Cluster 1 shows low level synteny with other bacterial EPS clusters (e.g., *Oceanobacillus iheyensis*), which is restricted to a minority of the glycosyltransferase genes and biosynthesis genes at either end of the cluster (data not shown). Cluster 2 (LSL_1547-LSL_1574) has lower GC content (29%) compared with the rest of the genome and comprises 21 EPS-related genes, interrupted by 7 genes of unknown or unrelated function; 14 of the genes in this cluster show significant homology to genes in the major EPS cluster of *L. plantarum*, but there is no extended synteny. This region of the *L. plantarum* genome also shows aberrant base composition and has recently been subdivided into three clusters, the second of which is up-regulated in a quorum-sensing mutant (35). *L. salivarius* EPS cluster 2 includes genes for a LytR-type transcriptional regulator, a flippase, and seven glycosyltransferases, including a priming glucose phosphotransferase (EpsE homolog; LSL_1550), and, thus, represents a more complete unit than Cluster 1, relative to characterized EPS operons.

Carbohydrate Metabolism and Transport. Genes in the Clusters of Orthologous Groups top-level category of metabolism were enriched in the half of the *L. salivarius* chromosome closest to the origin (Fig. 1), as previously reported for *L. plantarum* (15), although carbon metabolism and transport genes are more clustered at positions 400 kb, 500 kb, and 1.35 Mb. *L. salivarius* is currently regarded as homofermentative (36), meaning that sugars can be fermented only via the Embden-Meyerhof-Parnas pathway, and genes for the complete glycolysis pathway are present in the chromosome. Interestingly, genes for the pentose phosphate pathway also were found in the *L. salivarius* UCC118 genome, suggesting that it should be grouped among the facultatively heterofermentative lactobacilli. The genome sequence suggested that *L. salivarius* UCC118 would be able to assimilate ribose (Fig. 2). This suggestion was experimentally confirmed by growth on ribose as a sole carbon source, and, furthermore, we detected lactate, acetate, and ethanol by HPLC in the culture medium, confirming the heterofermentative status of this strain. The two key enzymes of the pentose phosphate pathway, transketolase (LSL_1946) and transaldolase (LSL_1888, LSL_1947),
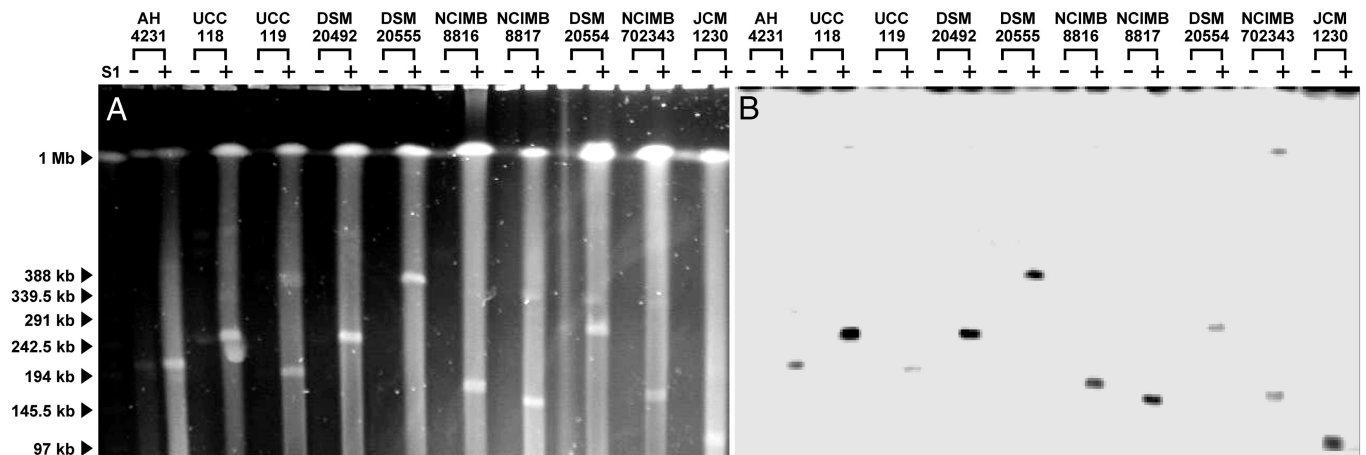
RIBOSE NUCLEOTIDES

NADP+ NADPH+H++CO2

LSL_0645
PHOSPHO-GLUCONATE DEHYDROGENASE EC 1.1.1.44

RIBULOSE-5-P

RIBOKINASE EC 2.7.1.15 LSL_0423

LSL_0318 LSL_1610
RIBOSE-P-PYRO-PHOSPHOKINASE EC 2.7.6.1

5-P-RIBOSYL-PP (PRPP)

6-P-GLUCONATE

LSL_1439 LSL_1806 RIBOSE-5-P ISOMERASE EC 5.3.1.6

LSL_1807 LSL_1948 RIBULOSE-3-P EPIMERASE EC 5.1.3.1

LSL_1953 RIBITOL-5-P 2-DEHYDROGENASE EC 1.1.1.137

RIBITOL-5-P

LSL_1142 6-PHOSPHO-GLUCONO-LACTONASE EC 3.1.1.31

RIBOSE-5-P XYLULOSE-5-P

TRANS-KETOLASE EC 2.2.1.1 LSL_1946

LSL_1509 LSL_1956 PHOSPHO-KETOLASE EC 4.1.2.9

6-P-GLUCONO-1,5-LACTONE

SEDOHEPTULOSE-7-P GLYCERALDEHYDE-3-P

NADPH+H+

LSL_0462 GLUCOSE-6-P DEHYDROGENASE EC 1.1.1.49

NADP+

LSL_1888 LSL_1947 TRANSALDOLASE EC 2.2.1.2

LSL_1509 LSL_1956 PHOSPHO-KETOLASE EC 4.1.2.9

ERYTHROSE-4-P

LSL_1946 TRANSKETOLASE EC 2.2.1.1

GLUCOSE-6-P

LSL_0464 GLUCOSE PHOSPHATE ISOMERASE EC 5.3.1.9

FRUCTOSE-6-P

LSL_0866 6-PHOSPHO-FRUCTO-1-KINASE EC 2.7.1.11

FRUCTOSE-1,6-P2

LSL_0406 FRUCTOSE BIPHOSPHATE ALDOLASE EC 4.1.2.13

GLYCERAL-DEHYDE-3-P

LSL_0552 GLUCOKINASE EC 2.7.1.2

LSL_1903 FRUCTOSE BIPHOSPHATASE EC 3.1.3.11

GLUCOSE GLYCOLYSIS

Fig. 2. Genes for completing the pentose phosphate pathway are on pMP118. Numbers and arrows in red or green represent genes and pathways encoded by pMP118 or the chromosome, respectively. Features in amber have genes coded by both replicons.

are encoded by the megaplasmid (Fig. 2). The gene products of pMP118 are not essential for nucleotide biosynthesis when *L. salivarius* UCC118 is grown on glucose, but the presence of an additional copy of a gene for ribose-5-phosphate isomerase on pMP118 (LSL_1806) may increase the flexibility and flux of this pathway. The presence of pMP118 would be essential for the growth of UCC118 if pentose were used as the sole carbon source. Growth on ribose, aided by genes resident on pMP118, may confer a competitive advantage on the strain when living in the human GI tract, because ribose might be an abundant carbon source in the GI tract because of RNA degradation.

A gene on pMP118 encoding fructose biphosphatase (LSL_1903) and a chromosomally encoded phosphoenolpyruvate carboxykinase (LSL_0395) were identified. The first enzyme catalyzes the formation of fructose-6-phosphate, which is required for the synthesis of glucosamine-6-phosphate and its derivatives involved in peptidoglycan formation. The second enzyme catalyzes the formation of phosphoenolpyruvate by decarboxylation of oxaloacetate, while hydrolyzing ATP, a rate-limiting step in gluconeogenesis. Although *L. sakei* 23k has fructose biphosphatase and *L. plantarum* WCFS1 has phosphoenolpyruvate carboxykinase, *L. salivarius* UCC118 and *L. casei* ATCC334 (www.jgi.doe.gov) seem to be the only sequenced *Lactobacilli* that have both enzymes (with the assistance of pMP118 in *L. salivarius*), suggesting the presence of a functional gluconeogenesis pathway (Figs. 11 and 12). When glucose is exhausted, as can be encountered in certain regions of the GI tract, the gluconeogenesis pathway might be activated. The presence of a complete pentose phosphate pathway to generate glyceraldehyde-3-phosphate, coupled with gluconeogenesis from pyruvate, could be an adaptation to pentose-based growth.

In addition to glucose and ribose, data obtained from API 50 sugar fermentation tests showed that *L. salivarius* UCC118 is able to ferment a number of other monosaccharides (adonitol,

galactose, fructose, mannose, mannitol, sorbitol, and *N*-acetylglucosamine) and disaccharides (maltose, lactose, sucrose, and trehalose). Good growth of *L. salivarius* UCC118 (i.e., at least 10^9 colony-forming units/ml) was achieved in MRS broth supplemented with each of the listed carbohydrates (data not shown). Most of these carbohydrates are likely found in, or can be generated from, human dietary components, reflecting the adaptation of *L. salivarius* to the human GI tract. Putative phosphotransferase transporters or ABC families of transporters for using these sugars were identified in the *L. salivarius* UCC118 genome. In addition, genes encoding enzymes involved in rhamnose and *N*-acetylneuraminic acid (sialic acid) catabolism as well as sorbitol utilization are present on pMP118. Collectively, the presence of pMP118 appears likely to increase the metabolic flexibility and, thus, competitiveness of *L. salivarius* UCC118.

Secreted Proteins and Transporters. The predicted secreted protein complement of *L. salivarius* UCC118 contains 119 proteins (J.P.v.P., C.C., K. A. Ryan, Y.L., M.J.C., B. Sheil, L. Steidler, L. O'Mahony, G.F.F., D.v.S., and P.W.O., unpublished data). Among these 119 proteins, only 10 are likely sortase-dependent proteins. The megaplasmid is predicted to encode eight of the secreted proteins. There is a single sortase gene *srtA* (LSL_1606) in the genome and genes encoding putative signal peptidase I (LSL_0876) and signal peptidase II (LSL_0825). The genome contains the second lowest total number of predicted transporters among sequenced *Lactobacillus* genomes (Table 3, which is published as supporting information on the PNAS web site), among which there are 23 phosphotransferase transporters.

Megaplasmid-Encoded Properties and Plasmid Profiles of *L. salivarius*. There are no genes on the megaplasmid pMP118 that might be considered strictly essential for viability. For example, pMP118 harbors an additional copy of *rpsN* (LSL_1944, ribosomal protein S14P), which is a paralog of the chromosomal gene LSL_1422 and is homologous to *rpsN2* of *L. plantarum* and *L. johnsonii*. The gene on pMP118 encoding a bile-salt hydrolase (choloylglycine hydrolase; LSL_1801) is one of only two genes encoding bile-inactivating enzymes detected in the genome. The copy number of pMP118 was estimated by PCR to be 4.7 ± 0.6 copies per chromosome equivalent in stationary phase, so gene dosage effects would contribute to amplifying the contribution of LSL_1801 to bile resistance.

The *L. salivarius* UCC118 chromosome harbors one copy each of *ldhL* and *ldhD* genes, whereas pMP118 encodes an additional copy of the *ldhD* gene (LSL_1887), whose product is 42% identical to the *L. plantarum* enzyme. D-lactate is an important component of cell wall precursors in *L. plantarum* (37), and the additional pMP118-encoded LdhD could increase the efficiency of D-lactate production, provided that the gene is expressed and the gene product is catalytically active. In addition, LSL_1901 on pMP118 encodes a bifunctional acetaldehyde/alcohol dehydrogenase, which is the only enzyme in this strain that catalyzes the formation of ethanol from acetyl-CoA via acetaldehyde. Although not essential, the presence of this additional reductive pathway on pMP118 likely would improve the redox-balancing capability of strain UCC118.

L. salivarius UCC118 elaborates a two-component Class IIb bacteriocin, Abp118, that has been characterized in ref. 12. The cloned "chromosomal" EcoRI fragment containing the Abp118 locus that we previously sequenced was actually derived from pMP118, and the gene now designated LSL_1924 contains the EcoRI site at one end of the cloned fragment. This EcoRI fragment includes LSL_1921, which encodes a presalivaricin B homolog; we previously reasoned that this protein was likely nonfunctional, because an internal deletion introduced in the gene for Abpα eliminated all antibacterial activity from

Fig. 3. Megaplasmids of varying size are found in *L. salivarius*. (*A*) PFGE of genomic DNA of 10 *L. salivarius* strains. (*B*) Corresponding Southern hybridization with the pMP118 *repA* probe. (+) or (−) indicate treatment with S1 nuclease. Arrowheads to left indicate λ DNA concatamers used as size standard. Detailed strain information is provided in *Supporting Text*. The estimated size of megaplasmids are as follows: AH4231, 210 kb; UCC118, 242 kb; UCC119, 195 kb; DSM20492, 240 kb; DSM20555, 380 kb; NICMB8816, 180 kb; NCIMB8817, 145 kb; DSM20554, 260 kb; NCIMB702343, 160 kb; JCM1230, 100 kb.

the cloned fragment in *L. lactis* and *L. plantarum* (12). However, the gene LSL_1924 that was truncated in the original cloned fragment encodes a potential bacteriocin transport protein. From consideration of the complete sequence of this region, it is possible that the salivaricin gene is functional but requires specific transport or accessory proteins. Interestingly, we have recently detected the genetic locus for Abp118 production in *L. salivarius* strains isolated from chickens (38), suggesting this trait may be widely distributed, and its linkage to strain plasmid content is being investigated.

We used a combination of hybridization and S1 nuclease treatment, in combination with pulsed-field gel electrophoresis (PFGE) to investigate plasmid content of *L. salivarius* strains from varying sources (described in Table 4, which is published as supporting information on the PNAS web site). Nuclease S1 preferentially nicks and then linearizes megaplasmids because of their torsional stress (39). *L. salivarius* comprises two subspecies, *salivarius* and *salicinius* (36), and both were found to contain megaplasmids (Fig. 3), of sizes varying from 100 kb to 380 kb. All of these megaplasmids hybridized with a pMP118 *repA* gene probe. (Prominent bands in S1 nuclease-treated samples in Fig. 3*A* are chromosomal DNA, which caused nonspecific hybridization in Fig. 3*B* in some samples). It is noteworthy that pMP118 contains a tract of genes that show low relatedness to known or suspected conjugation genes (see Table 5, which is published as supporting information on the PNAS web site), thus representing a functional or remnant plasmid transfer locus. The ability to disseminate by conjugation would explain the apparently universal presence of pMP118-related plasmids (in strains so far tested), and this potential *tra* locus also might be involved in mobilization of smaller plasmids.

The megaplasmid pMP118 is the largest plasmid from lactic acid bacteria in current nucleotide databases. However, plasmids >100 kb in *Lactobacillus* spp. have been reported. Lactacin F production by *L. acidophilus* strain 88 was shown by conjugation analysis to be linked to a 110-kb plasmid, pPM68 (40). A plasmid of 150 kb was identified in *L. gasseri* by PFGE and was suggested to be linear on the basis of electrophoretic behavior (41). However, its size and conformation were not confirmed. Given that many plasmid-profile studies of lactic acid bacteria (29) predated PFGE technology or did not employ conditions required to separate undigested replicons from the chromosome (39), it is possible that megaplasmids have a wider distribution

in these bacteria than was previously recognized. The sequence of pMP118 reported here, in the context of the whole genome, uniquely illustrates the contribution of a megaplasmid to diverse metabolic and phenotypic properties of *L. salivarius* by both integrating with and extending chromosomally encoded features. It also provides a definitive platform for investigating the replication and dissemination of very large plasmids in this species.

The circular chromosome of *L. salivarius* UCC118 is the smallest *Lactobacillus* chromosome so far sequenced, 57.5 kb smaller than that of *L. sakei* and 165.5 kb smaller than that of *L. johnsonii* (Tables 6 and 7, which are published as supporting information on the PNAS web site). The presence of genes located on pMP118 confers numerous additional metabolic capabilities. The contribution of large numbers of pMP118 genes of unknown function to phenotype obviously cannot be evaluated. However, the fact that pMP118 contains a repertoire of genes that likely confer metabolic flexibility, seen in the context of significant megaplasmid size variation in other strains, strongly suggests that the multireplicon genome architecture of *L. salivarius* bestows on the species a dynamic and flexible genetic complement. This architecture could be in response to dietary fluctuations in host species, flexible niches in the GI tract, or adaptation to different hosts. In conclusion, the sequence of the *L. salivarius* UCC118 genome has provided a unique insight into the contribution of chromosomal and megaplasmid-encoded genes to the biology of this organism.

Materials and Methods

Bacterial Strains. Bacterial strains are listed in Table 4.

Sequencing, Assembly, and Analysis. The genome of *L. salivarius* UCC118 was sequenced by a standard shotgun strategy (see *Supporting Text*, which is published as supporting information on the PNAS web site, for details). The final assembly of the genome was confirmed by concordance of pulsed-field gel restriction patterns (ApaI, AsiSI, BglI, BssHII, MluI, RsrII, SacII, and SmaI) with the restriction map generated *in silico*. Standard procedures were used for genome annotation and analysis (*Supporting Text*). The genome sequence was integrated in the ERGO database (42), where a functional annotation was carried out, followed by manual verification and curation. Predicted proteins were functionally classified on the basis of Clusters of Orthologous Groups (43). The Transport

Classification Database (TCDB; ref. 44) was used to categorize transport proteins. Genomes were aligned by using the Artemis Comparison Tool (45). The annotated genome sequence has been deposited in GenBank under accession nos. CP000233 (chromosome) and CP000234 (pMP118).

Metabolic Reconstruction. A metabolic reconstruction was performed by using ERGO predictions, and Pathway Tools (46) based on EC numbers and enzymatic nomenclature. In addition, Pathway Hole Filler, a method using Bayesian statistics and BLAST comparisons, helped in identifying missing enzymes in the reconstructed pathways (47). Kyoto Encyclopedia of Genes and Genomes maps (48) were used to record and visualize this overview.

Copy Number Determination. The copy number of pMP118 relative to the chromosome was estimated by quantitative PCR. A range of gene fragments were amplified first to find equal PCR efficiencies. The genes *lspD* (LSL_1838) and a pseudogene (LSL_1320) on pMP118 and chromosome, respectively, were selected. The ratio of plasmid to chromosome was calculated by using the formula $2^{(Ct\ chromosomal\ gene\ -\ Ct\ plasmid\ gene)}$, where Ct is the crossing threshold value.

PFGE and Hybridization. Nuclease S1 treatment and PFGE was carried out as described in ref. 39. Southern hybridization followed standard protocols (49). The probe was a 765-bp amplicon corresponding to the *repA* gene of pMP118 (LSL_1739).

1. Klaenhammer, T. R. (2000) *J. Nutr.* **130,** 415S–416S.
2. Klaenhammer, T. R. & Russell, W. M. (2000) in *Encyclopedia Food Microbiology* (Elsevier, Amsterdam), Vol. 2, pp. 1151–1157.
3. Tannock, G. W. (2004) *Appl. Environ. Microbiol.* **70,** 3189–3194.
4. Vaughan, E. E., Heilig, H. G., Ben-Amor, K. & de Vos, W. M. (2005) *FEMS Microbiol. Rev.* **29,** 477–490.
5. Molin, G., Jeppsson, B., Johansson, M. L., Ahrne, S., Nobaek, S., Stahl, M. & Bengmark, S. (1993) *J. Appl. Bacteriol.* **74,** 314–323.
6. Ahrne, S., Nobaek, S., Jeppsson, B., Adlerberth, I., Wold, A. E. & Molin, G. (1998) *J. Appl. Microbiol.* **85,** 88–94.
7. Heilig, H. G., Zoetendal, E. G., Vaughan, E. E., Marteau, P., Akkermans, A. D. & de Vos, W. M. (2002) *Appl. Environ. Microbiol.* **68,** 114–123.
8. Dunne, C., Murphy, L., Flynn, S., O'Mahony, L., O'Halloran, S., Feeney, M., Morrissey, D., Thornton, G., Fitzgerald, G., Daly, C., *et al.* (1999) *Antonie Van Leeuwenhoek* **76,** 279–292.
9. Dunne, C., O'Mahony, L., Murphy, L., Thornton, G., Morrissey, D., O'Halloran, S., Feeney, M., Flynn, S., Fitzgerald, G., Daly, C., *et al.* (2001) *Am. J. Clin. Nutr.* **73,** 386S–392S.
10. McCarthy, J., O'Mahony, L., O'Callaghan, L., Sheil, B., Vaughan, E. E., Fitzsimons, N., Fitzgibbon, J., O'Sullivan, G. C., Kiely, B., Collins, J. K. & Shanahan, F. (2003) *Gut* **52,** 975–980.
11. Sheil, B., McCarthy, J., O'Mahony, L., Bennett, M. W., Ryan, P., Fitzgibbon, J. J., Kiely, B., Collins, J. K. & Shanahan, F. (2004) *Gut* **53,** 694–700.
12. Flynn, S., van Sinderen, D., Thornton, G. M., Holo, H., Nes, I. F. & Collins, J. K. (2002) *Microbiology* **148,** 973–984.
13. Siezen, R. J., van Enckevort, F. H., Kleerebezem, M. & Teusink, B. (2004) *Curr. Opin. Biotechnol.* **15,** 105–115.
14. Klaenhammer, T. R., Barrangou, R., Buck, B. L., Azcarate-Peril, M. A. & Altermann, E. (2005) *FEMS Microbiol. Rev.* **29,** 393–409.
15. Kleerebezem, M., Boekhorst, J., van Kranenburg, R., Molenaar, D., Kuipers, O. P., Leer, R., Tarchini, R., Peters, S. A., Sandbrink, H. M., Fiers, M. W., *et al.* (2003) *Proc. Natl. Acad. Sci. USA* **100,** 1990–1995.
16. Pridmore, R. D., Berger, B., Desiere, F., Vilanova, D., Barretto, C., Pittet, A.-C., Zwahlen, M.-C., Rouvet, M., Altermann, E., Barrangou, R., *et al.* (2004) *Proc. Natl. Acad. Sci. USA* **101,** 2512–2517.
17. Altermann, E., Russell, W. M., Azcarate-Peril, M. A., Barrangou, R., Buck, B. L., McAuliffe, O., Souther, N., Dobson, A., Duong, T., Callanan, M., *et al.* (2005) *Proc. Natl. Acad. Sci. USA* **102,** 3906–3912.
18. Chaillou, S., Champomier-Verges, M. C., Cornet, M., Crutz-Le Coq, A. M., Dudez, A. M., Martin, V., Beaufils, S., Darbon-Rongere, E., Bossy, R., Loux, V. & Zagorec, M. (2005) *Nat. Biotechnol.* 1527–1533.
19. Boekhorst, J., Siezen, R. J., Zwahlen, M. C., Vilanova, D., Pridmore, R. D., Mercenier, A., Kleerebezem, M., de Vos, W. M., Brussow, H. & Desiere, F. (2004) *Microbiology* **150,** 3601–3611.
20. Bolotin, A., Quinquis, B., Renault, P., Sorokin, A., Ehrlich, S. D., Kulakauskas, S., Lapidus, A., Goltsman, E., Mazur, M., Pusch, G. D., *et al.* (2004) *Nat. Biotechnol.* **22,** 1554–1558.
21. Flynn, S. (2001) Ph.D. thesis (University College Cork, Ireland).
22. Ochman, H. (2002) *Curr. Biol.* **12,** R427–R428.
23. Bentley, S. D. & Parkhill, J. (2004) *Annu. Rev. Genet.* **38,** 771–792.
24. Ng, W. V., Ciufo, S. A., Smith, T. M., Bumgarner, R. E., Baskin, D., Faust, J., Hall, B., Loretz, C., Seto, J., Slagel, J., *et al.* (1998) *Genome Res.* **8,** 1131–1141.
25. del Solar, G., Giraldo, R., Ruiz-Echevarria, M. J., Espinosa, M. & Diaz-Orejas, R. (1998) *Microbiol. Mol. Biol. Rev.* **62,** 434–464.
26. Lobry, J. R. & Sueoka, N. (2002) *Genome Biol.* **3,** RESEARCH0058.
27. Martinez-Bueno, M., Valdivia, E., Galvez, A. & Maqueda, M. (2000) *Curr. Microbiol.* **41,** 257–261.
28. Eppinger, M., Baar, C., Raddatz, G., Huson, D. H. & Schuster, S. C. (2004) *Nat. Rev. Microbiol.* **2,** 872–885.
29. Wang, T. T. & Lee, B. H. (1997) *Crit. Rev. Biotechnol.* **17,** 227–272.
30. Cerdeno-Tarraga, A. M., Patrick, S., Crossman, L. C., Blakely, G., Abratt, V., Lennard, N., Poxton, I., Duerden, B., Harris, B., Quail, M. A., *et al.* (2005) *Science* **307,** 1463–1465.
31. O'Sullivan, D., Twomey, D. P., Coffey, A., Hill, C., Fitzgerald, G. F. & Ross, R. P. (2000) *Mol. Microbiol.* **36,** 866–875.
32. Hols, P., Kleerebezem, M., Schanck, A. N., Ferain, T., Hugenholtz, J., Delcour, J. & de Vos, W. M. (1999) *Nat. Biotechnol.* **17,** 588–592.
33. Kilstrup, M., Hammer, K., Ruhdal Jensen, P. & Martinussen, J. (2005) *FEMS Microbiol. Rev.* **29,** 555–590.
34. Peant, B., LaPointe, G., Gilbert, C., Atlan, D., Ward, P. & Roy, D. (2005) *Microbiology* **151,** 1839–1851.
35. Sturme, M. H., Nakayama, J., Molenaar, D., Murakami, Y., Kunugi, R., Fujii, T., Vaughan, E. E., Kleerebezem, M. & de Vos, W. M. (2005) *J. Bacteriol.* **187,** 5224–5235.
36. Rogosa, M., Wiseman, R. F., Mitchell, J. A., Disraely, M. N. & Beaman, A. J. (1953) *J. Bacteriol.* **65,** 681–699.
37. Goffin, P., Deghorain, M., Mainardi, J. L., Tytgat, I., Champomier-Verges, M. C., Kleerebezem, M. & Hols, P. (2005) *J. Bacteriol.* **187,** 6750–6761.
38. Li, Y., Mulcahy, G., Van Sinderen, D. & O'Toole, P. W. (2005) in *LAB8 Symposium on Lactic Acid Bacteria* (Fed. Eur. Microbiol. Soc., Egmond aan Zee, Netherlands).
39. Barton, B. M., Harding, G. P. & Zuccarelli, A. J. (1995) *Anal. Biochem.* **226,** 235–240.
40. Muriana, P. M. & Klaenhammer, T. R. (1987) *Appl. Environ. Microbiol.* **53,** 553–560.
41. Roussel, Y., Colmin, C., Simonet, J. M. & Decaris, B. (1993) *J. Appl. Bacteriol.* **74,** 549–556.
42. Overbeek, R., Larsen, N., Walunas, T., D'Souza, M., Pusch, G., Selkov, E., Jr., Liolios, K., Joukov, V., Kaznadzey, D., Anderson, I., *et al.* (2003) *Nucleic Acids Res.* **31,** 164–171.
43. Tatusov, R. L., Fedorova, N. D., Jackson, J. D., Jacobs, A. R., Kiryutin, B., Koonin, E. V., Krylov, D. M., Mazumder, R., Mekhedov, S. L., Nikolskaya, A. N., *et al.* (2003) *BMC Bioinformatics* **4,** 41.
44. Busch, W. & Saier, M. H., Jr. (2002) *Crit. Rev. Biochem. Mol. Biol.* **37,** 287–337.
45. Carver, T. J., Rutherford, K. M., Berriman, M., Rajandream, M. A., Barrell, B. G. & Parkhill, J. (2005) *Bioinformatics* **21,** 3422–3423.
46. Karp, P. D., Paley, S. & Romero, P. (2002) *Bioinformatics* **18,** Suppl. 1, S225–S232.
47. Green, M. L. & Karp, P. D. (2004) *BMC Bioinformatics* **5,** 76.
48. Kanehisa, M. (1997) *Trends Genet.* **13,** 375–376.
49. Sambrook, J. & Russell, D. W. (2001) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab. Press, Woodbury, NY).

MICROBIOLOGY