The Complete Genome Sequence of the Lactic Acid Bacterium *Lactococcus lactis* ssp. *lactis* IL1403

Alexander Bolotin, 1 Patrick Wincker, 2 Stéphane Mauger, 1, 3 Olivier Jaillon, 2 Karine Malarme, 1 Jean Weissenbach, 2 S. Dusko Ehrlich, 1 and Alexei Sorokin 1, 4

1 Génétique Microbienne, Institut National de la Recherche Agronomique, Domaine de Vilvert, 78352 Jouy en Josas CEDEX, France; 2 Génoscope, Centre National De Séquençage, BP 191 91006 Evry CEDEX, France

*Lactococcus lactis* is a nonpathogenic AT-rich gram-positive bacterium closely related to the genus *Streptococcus* and is the most commonly used cheese starter. It is also the best-characterized lactic acid bacterium. We sequenced the genome of the laboratory strain IL1403, using a novel two-step strategy that comprises diagnostic sequencing of the entire genome and a shotgun polishing step. The genome contains 2,365,589 base pairs and encodes 2310 proteins, including 293 protein-coding genes belonging to six prophages and 43 insertion sequence (IS) elements. Nonrandom distribution of IS elements indicates that the chromosome of the sequenced strain may be a product of recent recombination between two closely related genomes. A complete set of late competence genes is present, indicating the ability of *L. lactis* to undergo DNA transformation. Genomic sequence revealed new possibilities for fermentation pathways and for aerobic respiration. It also indicated a horizontal transfer of genetic information from Lactococcus to gram-negative enteric bacteria of *Salmonella-Escherichia* group.

The questions addressed in research on useful bacteria are often antithetical to those involving pathogens, because one of the basic objectives is to improve rather than to limit bacterial growth. Efficient use of lacticocci by dairy industry requires understanding of many aspects of bacterial physiology, such as use of sugars and proteins from milk for growth, conversion of sugars to lactate, and synthesis of substances involved in cheese flavor, and thus of the relationship between different types of fermentation. The potential for new applications of LAB, such as oral vaccines (Steidler et al. 2000) or production of foreign proteins and metabolites, leads to questions concerning the protein secretion system, biosynthesis of cofactors, and regulation of central metabolism. In addition to questions related to the industrial use of lactococci, fundamental biological questions, such as retrohoming of introns (Cousineau et al. 1998), are also being addressed in *L. lactis*.

There are two subspecies of *L. lactis*, designated initially as *Streptococcus lactis* and *Streptococcus cremoris* and reclassified more recently as *L. lactis* ssp. *lactis* and *L. lactis* ssp. *cremoris*, respectively (Schleifer et al. 1985). The former is preferred for making of soft cheeses and the latter for the hard ones. The two subspecies have been intensely studied, mainly because of their industrial interest, and have became excellent models for research on metabolism, physiology, genetics, and molecular biology of LAB.

Lactic acid bacteria (LAB) are a heterogeneous group of microorganisms that convert carbohydrates into lactic acid. They comprise both pathogens (such as *Streptococcus pneumoniae* or *Streptococcus pyogenes*) and useful bacteria (such as *Streptococcus thermophilus* and *Lactococcus lactis*, which were used for millennia in milk fermentation). Determination and analysis of the genome sequence of a representative LAB is therefore of great interest, as it would provide information allowing us to combat the former and use the latter more efficiently. Until now, no complete and annotated genome sequence of either LAB class has been reported.

In nature, *L. lactis* occupies a niche related to plant or animal surfaces and the animal gastrointestinal tract. It is believed to be dormant on the plant surfaces and to multiply in the gastrointestinal tract after being swallowed by a ruminant. In contrast, “domesticated” species of *L. lactis*, used by dairy industry as starters in cheese fermentation, live in a different niche, which is defined by technological considerations, such as fast growth and rapid production of lactic acid in milk. The importance of *L. lactis* for humankind can be appreciated from the estimate that close to 10^12 tons of cheese are made annually (Fox 1989), leading to human consumption of close to 10^18 lacticocci.

The sequence data described in this paper has been submitted to the GenBank data library under accession no. AE005176.
nistic genome sequencing, has been reported (Bolotin et al. 1999). Here we present the analysis of the accurate sequence of the IL1403 genome, which is the first such report for any lactic acid bacterium. We focus mainly on features related to the importance of _L. lactis_ for humankind, which is its use in dairy fermentation. Also, several unexpected findings are reported, such as a putative chimerical structure of the genome, the possibility that _L. lactis_ can respire, the existence of genes required for DNA transformation, and a discovery of a transfer of genetic information from lactococci to gram-negative enteric bacteria.

**RESULTS AND DISCUSSION**

Two-Step Sequencing Strategy

The first step of our strategy, designated diagnostic genome sequencing, was described before (Bolotin et al. 1999). Briefly, it implies cloning of relatively short (1–20 kb) genome fragments in _Escherichia coli_ plasmid and phage vectors, and sequencing of a limited number of randomly chosen clones, to a redundancy of about one. A novel procedure, designated multiplex MLA PCR, developed and tested in the course of the long accurate PCR (MLA PCR), developed and tested in the course of the _Bacillus subtilis_ genome sequencing project (Sorokin et al. 1996; Kunst et al. 1997), is then applied for connecting the resulting contigs and synthesizing the missing genome regions, sequenced subsequently by standard methods. This approach allowed us to establish the entire _L. lactis_ genome sequence and assemble it in a unique contig, with a sequencing redundancy of less than two (Bolotin et al. 1999). Three- to fourfold fewer sequencing reactions were required to reach this goal than if the fully random approach were used. For comparison, only 10,235 reactions were needed to assemble _L. lactis_ genome sequence, whereas 40,020 were required for the genome of _Neisseria meningitidis_ (Tettelin et al. 2000), which is of a similar size. Diagnostic sequence allowed us to identify all _L. lactis_ genes that encode proteins sufficiently similar to those present in the databases. However, the elevated error rate, estimated to be ~1%, did not allow us to predict the genes unique for _L. lactis_ or the borders of coding region. To obtain a more complete and reliable description of the _L. lactis_ genome, we carried out a second step of our strategy. It involved random sequencing of additional clones until the overall redundancy of ~6.4 was reached and then primer walking on PCR-generated templates to ensure that each base was sequenced at least four times and at least once on each strand. We designated this step “shotgun polishing” and concluded that the strategy presented here can be a good alternative to the fully random strategy used in most cases (Fraser and Fleischmann 1997). Its advantages should increase even more when a greater number of completely sequenced and thoroughly annotated bacterial genomes becomes available. Carrying out the diagnostic step and polishing only a very little will then be sufficient to determine a reliable genome sequence of bacteria relatively close to the ones that were already sequenced and annotated.

**Gene Content**

The circular chromosome of _L. lactis_ IL1403 has 2,365,589 bp and an average G+C content of 35.4%. We detected 2310 open reading frames (ORFs) in the sequence, with an average length of 879 bp. Protein-coding genes represent 86% of the genome, stable RNA 1.4%, and noncoding regions 12.6%. These values are similar to those observed for genomes of other bacteria. We have assigned a biochemical or biological role to 64.2% (1482 ORFs) of the genes and classified them into functional categories (Table 1). There are 20.1% of genes (465 ORFs) that match hypothetical coding sequences of unknown function, and the remaining 15.7% (363 ORFs) represent genes with no similarity to known proteins, which can be considered specific for lactococci.

**Origin and Terminus of Replication**

Approximate position of the replication origin and terminus of the _L. lactis_ chromosome was determined previously, using the GC and AT skews (Fig. 1; Bolotin et al. 1999). It should be noted that the precision of the origin mapping is greater than that of the terminus, as there are conserved elements (dnaA and dnaN genes, DnaA boxes) in the vicinity of the former but not of the latter (rtp gene was not found). We choose as the coordinate 1 of the genome the middle of a _Hind_III site localized near the replication origin (Fig 1).

**RNA, IS Elements, and Prophages**

Location of six rRNA operons, 62 tRNA genes, the RNA component of RNase P gene (mpB), and the 10S RNA (ssrA) were determined earlier from the diagnostic sequence (Bolotin et al. 1999). There are six different IS elements in the IL1403 chromosome: IS981, IS982, IS983, IS904, IS905, and IS1077, present in 10, 1, 15, 9, 1, and 7 copies, respectively (Fig. 1) and totaling 42 kb. It is remarkable that one or two copies of IS904 always accompany IS1077 and that the relative orientation of the two is generally not the same. The former element might be a satellite of the latter. Another remarkable feature is that three of the IS elements are not randomly distributed over the chromosome (Fig. 1). Seven copies of IS1077 (and the associated IS904) occupy the region between 2150 and 840 kb, encompassing the replication origin, whereas 15 copies of IS983 occupy a different region, between 680 and 2270 kb. The two regions overlap by only ~150 kb. As the 10 copies of IS981 are distributed over the whole genome, the un-
### Table 1. Functional Classification of the *Lactococcus lactis* Protein-Coding Genes

| Amino Acid Family | Gene(s) | Protein Function |
|-------------------|---------|-----------------|
| **Aromatic amino acids** | **aroA** | 1802 3-phosphoshikimate 1-carboxyvinyltransferase |
| **L-glutamate** | **gltB** | 1319 glutamate synthase large subunit |
| **L-citrulline** | **gltA** | 668 citrate synthase |
| **L-glutamine** | **glnA** | 2283 glutamine synthetase |
| **L-arginine** | **argJ** | 806 ornithine acetyltransferase |
| **L-proline** | **proA** | 1651 gamma-glutamyl phosphate reductase |
| **L-proline** | **proB** | 1652 glutamate 5-kinase |
| **L-proline** | **proC** | 1953 pyrrole-5-carboxylate reductase |
| **Histidine** | **hisA** | 1236 phosphoribosylformiminoo-s-aminomimidazole-carboxamidase ribotide isomerase |
| **Histidine** | **hisB** | 1234 fructose-1,6-bisphosphate:phosphatase reductase |
| **Histidine** | **hisC** | 1229 histidinol-phosphate aminotransferase |
| **Histidine** | **hisD** | 1232 histidinol dehydrogenase |
| **Histidine** | **hisF** | 1237 cyclase HisF |
| **Histidine** | **hisG** | 1231 ATP phosphoribosyltransferase |
| **Histidine** | **hisH** | 1235 amidotransferase |
| **Histidine** | **hisI** | 1237 phosphorylase AMP cyclohydrolase |
| **Histidine** | **hisK** | 1238 histidinol phosphatase |
| **Histidine** | **hisL** | 1230 ATP phosphoribosyltransferase regulatory subunit |
| **Serine** | **cysD** | 77 O-acetylhomoserine sulffydrase |
| **Serine** | **cysE** | 1921 serine acetyltransferase |
| **Serine** | **cysK** | 792 cysteine synthase |
| **Serine** | **cysM** | 527 cysteine synthase |
| **Serine** | **glyA** | 592 serine hydroxymethyltransferase |
| **Serine** | **hisA** | 1236 phosphoribosylformiminoo-s-aminomimidazole-carboxamidase ribotide isomerase |
| **Serine** | **hisB** | 1234 fructose-1,6-bisphosphate:phosphatase reductase |
| **Serine** | **hisC** | 1229 histidinol-phosphate aminotransferase |
| **Serine** | **hisD** | 1232 histidinol dehydrogenase |
| **Serine** | **hisF** | 1237 cyclase HisF |
| **Serine** | **hisG** | 1231 ATP phosphoribosyltransferase |
| **Serine** | **hisH** | 1235 amidotransferase |
| **Serine** | **hisI** | 1237 phosphorylase AMP cyclohydrolase |
| **Serine** | **hisK** | 1238 histidinol phosphatase |
| **Serine** | **hisL** | 1230 ATP phosphoribosyltransferase regulatory subunit |
| **Folic acid** | **dfra** | 1163 dihydropholate reductase |
| **Folic acid** | **fhs** | 961 formlate:hydrogenolactone synthase |
| **Folic acid** | **folB** | 1166 dihydronopterin aldolase |
| **Folic acid** | **folC** | 1169 folic acid synthase |
| **Folic acid** | **folD** | 877 tetrahydrofolate dehydrogenase/cyclohydrolase |
| **Folic acid** | **folE** | 1167 GTP cyclohydrolase I |
| **Folic acid** | **folF** | 1168 dihydrodopterato synthase |
| **Folic acid** | **pabB** | 1348 para-amino benzoate synthase component II |
| **Folic acid** | **pabB** | 1348 para-amino benzoate synthase component I |
| **Heme and porphyrin** | **emH** | 1609 ferrohelaicase |
| **Heme and porphyrin** | **hemK** | 589 protoporphyrinogen oxidase |
| **Heme and porphyrin** | **hemN** | 1154 oxygen-independent coproporphyrogen III oxidase |
| **Menaquinone** | **folP** | 1168 dihydropteroate synthase |
| **Menaquinone** | **folE** | 1167 GTP cyclohydrolase I |
| **Menaquinone** | **folD** | 877 tetrahydrofolate dehydrogenase/cyclohydrolase |
| **Menaquinone** | **folC** | 1169 folic acid synthase |
| **Menaquinone** | **hbs** | 961 formlate:hydrogenolactone synthase |
| **Menaquinone** | **folB** | 1166 dihydronopterin aldolase |
| **Menaquinone** | **folC** | 1169 folic acid synthase |
| **Menaquinone** | **folD** | 877 tetrahydrofolate dehydrogenase/cyclohydrolase |
| **Menaquinone** | **folE** | 1167 GTP cyclohydrolase I |
| **Menaquinone** | **folF** | 1168 dihydrodopterato synthase |
| **Menaquinone** | **pabB** | 1348 para-amino benzoate synthase component II |
| **Menaquinone** | **pabB** | 1348 para-amino benzoate synthase component I |
| **MenAquinone** and ubiquinone | **ispA** | 1349 para-amino benzoate synthase component II |
| **MenAquinone** and ubiquinone | **ispB** | 1348 para-amino benzoate synthase component I |
| **MenAquinone** and ubiquinone | **folP** | 1168 dihydrodopterato synthase |
| **MenAquinone** and ubiquinone | **folC** | 1169 folic acid synthase |
| **MenAquinone** and ubiquinone | **hfs** | 961 formlate:hydrogenolactone synthase |
| **MenAquinone** and ubiquinone | **folB** | 1166 dihydronopterin aldolase |
| **MenAquinone** and ubiquinone | **folC** | 1169 folic acid synthase |
| **MenAquinone** and ubiquinone | **folD** | 877 tetrahydrofolate dehydrogenase/cyclohydrolase |
| **MenAquinone** and ubiquinone | **folE** | 1167 GTP cyclohydrolase I |
| **MenAquinone** and ubiquinone | **folF** | 1168 dihydrodopterato synthase |
| **MenAquinone** and ubiquinone | **pabB** | 1348 para-amino benzoate synthase component II |
| **MenAquinone** and ubiquinone | **pabB** | 1348 para-amino benzoate synthase component I |
| **MenAquinone** and ubiquinone | **ispA** | 1349 para-amino benzoate synthase component II |
| **MenAquinone** and ubiquinone | **ispB** | 1348 para-amino benzoate synthase component I |
| **MenAquinone** and ubiquinone | **folP** | 1168 dihydrodopterato synthase |
| **MenAquinone** and ubiquinone | **folC** | 1169 folic acid synthase |
| **MenAquinone** and ubiquinone | **hfs** | 961 formlate:hydrogenolactone synthase |
| **MenAquinone** and ubiquinone | **folB** | 1166 dihydronopterin aldolase |
| **MenAquinone** and ubiquinone | **folC** | 1169 folic acid synthase |
| **MenAquinone** and ubiquinone | **folD** | 877 tetrahydrofolate dehydrogenase/cyclohydrolase |
| **MenAquinone** and ubiquinone | **folE** | 1167 GTP cyclohydrolase I |
| **MenAquinone** and ubiquinone | **folF** | 1168 dihydrodopterato synthase |
| **MenAquinone** and ubiquinone | **pabB** | 1348 para-amino benzoate synthase component II |
| **MenAquinone** and ubiquinone | **pabB** | 1348 para-amino benzoate synthase component I |
| **MenAquinone** and ubiquinone | **isol** | 1349 para-amino benzoate synthase component II |
| **MenAquinone** and ubiquinone | **isol** | 1348 para-amino benzoate synthase component I |
| **MenAquinone** and ubiquinone | **isol** | 1349 para-amino benzoate synthase component II |
| **MenAquinone** and ubiquinone | **isol** | 1348 para-amino benzoate synthase component I |
| **MenAquinone** and ubiquinone | **isol** | 1349 para-amino benzoate synthase component II |
| **MenAquinone** and ubiquinone | **isol** | 1348 para-amino benzoate synthase component I |

(Table continues on pp. 734–746.)
### Table 1. (Continued)

| Gene | Description |
|------|-------------|
| cobQ | 1115 cobryic acid synthase |
| ribA | 1024 GTP cyclohydrolase II / 3,4-dihydroxy-2-butane-4-phosphate synthase |
| ribB | 1023 riboflavin synthase alpha chain |
| ribC | 1142 riboflavin kinase |
| ribD | 1023 riboflavin-specific deaminase |
| ribH | 1025 riboflavin synthase beta chain |
| thioredoxin, glutaredoxin, and glutathione |  |
| gpo | 1402 glutathione peroxidase |
| gshR | 864 glutathione reductase |
| trxA | 1692 thioredoxin |
| trxB1 | 966 thioredoxin reductase |
| trxB2 | 1695 thioredoxin reductase |
| trxH | 396 thioredoxin H-type |
| Thiamin |  |
| apbE | 1125 thiamine biosynthesis lipoprotein |
| thiD1 | 1295 phosphomethylpyrimidine kinase |
| thiD2 | 485 phosphomethylpyrimidine kinase |
| thiE | 1294 thiamin-phosphate pyrophosphorylase |
| thiM | 1295 hydroxyethylthiazole kinase |
| Pyridinenucleotides |  |
| nadE | 1110 NAD-synthetase |
| yvdG | 2139 pyridinenucleotide-disulfide oxidoreductase |
| CELL ENVELOPE |  |
| Membranes, lipoproteins, and porins |  |
| bmpA | 1462 basic membrane protein A |
| cdaA | 2200 phosphatidate cytidylyltransferase |
| clsA | 988 cardiolipin synthase |
| clsB | 1188 cardiolipin synthase |
| dgaA | 1095 diacylglycerol kinase |
| lgt | 606 prolipoprotein diacylglycerol transferase |
| pagA | 2047 CDP-diacylglycerol-phosphate phosphatidyltransferase |
| plpA | 318 outer membrane lipoprotein precursor |
| plpB | 319 outer membrane lipoprotein precursor |
| plpC | 320 outer membrane lipoprotein precursor |
| plpD | 321 outer membrane lipoprotein precursor |
| yfIC | 596 acylphosphate phosphohydrolase |
| Murein sacculus and peptidoglycan |  |
| acmA | 269 N-acetylmuramidase |
| acmB | 1977 N-acetylmuramidase |
| acmC | 1403 N-acetylmuramidase |
| acmD | 528 N-acetylmuramidase |
| asd | 1667 aspartate-semialdehyde dehydrogenase |
| dacA | 2356 D-alanyl-D-alanine carboxypeptidase |
| dacB | 976 D-alanyl-D-alanine carboxypeptidase |
| dal | 862 alanine racemase |
| ddl | 341 D-alanine-D-alanine ligase |
| glmU | 1952 UDP-N-acetylgalactosamine pyrophosphorylase |
| mraY | 892 phospho-N-acetylmuramoyl-pentapeptide transferase |
| mreC | 2316 cell shape determining protein |
| mreD | 2315 cell shape determining protein |
| murA1 | 1314 UDP-N-acetylglucosamine 1-carboxyvinyltransferase |
| murA2 | 535 UDP-N-acetylglucosamine 1-carboxyvinyltransferase |
| murB | 1175 UDP-N-acetylenolpyruvoylglucosamine reductase |
| murC | 2119 UDP-N-acetylmuramate-alanine ligase |
| murD | 1634 UDP-N-acetylmuramoylalaneine D-glutamate ligase |
| murE | 1871 UDP-MurNac-tripeptide synthetase |

### Table 1. (Continued)

| Gene | Description |
|------|-------------|
| murF | 342 D-Ala-D-Ala adding enzyme |
| murG | 1633 peptidoglycan synthesis protein MurG |
| murl | 1313 glutamate racemase |
| pbp1B | 393 penicillin-binding protein 1B |
| pbp2A | 2178 penicillin-binding protein 2a |
| pbp2B | 339 penicillin-binding protein 2B |
| pbpX | 890 penicillin-binding protein |
| ponA | 530 penicillin-binding protein 1A |
| racD | 2310 aspartate racemase |
| uppS | 2201 undecaprenyl pyrophosphate synthetase |
| Surface polysaccharides, lipopolysaccharides and antigens |  |
| dltA | 1293 D-alanine activating enzyme |
| dltB | 1291 peptidoglycan biosynthesis protein |
| dltC | 1290 D-alanyl carrier protein |
| dltD | 1290 D-alanine transfer protein DltD |
| dltE | 145 oxidoreductase |
| flOL | 746 flotillin-like protein |
| hasC | 1378 UTP-glucose-1-phosphate uridylyltransferase |
| icaA | 681 glycosyl transferase |
| icaB | 683 intercellular adhesion protein IcaB |
| icaC | 684 collagen adhesin |
| kdtB | 2239 lipopolysaccharide core biosynthesis protein |
| mvaA | 1611 hydroxymethylglutaryl-CoA reductase |
| mycA | 981 myosin-cross-reactive antigen |
| pspA | 2304 glucosyltransferase-S |
| pspB | 2306 glucosyltransferase-S |
| rggA | 202 rhamnosyltransferase |
| rggB | 203 rhamnosyltransferase |
| rggE | 207 glycosyltransferase |
| rggF | 209 polysaccharide biosynthesis protein |
| tagB | 953 teichoic acid biosynthesis protein B |
| tagD1 | 220 glycerol-3-phosphate cytidiltransferase |
| tagD2 | 951 glycerol-3-phosphate cytidiltransferase |
| tagF | 952 teichoic acid biosynthesis protein F |
| tagL | 936 exopolysaccharide biosynthesis protein |
| tagS | 948 teichoic acid biosynthesis protein |
| tagX | 945 teichoic acid biosynthesis protein |
| tagZ | 943 teichoic acid biosynthesis protein |
| ycbB | 212 glycosyltransferase |
| ycbD | 213 UDP-glucose 4-epimerase |
| ycbF | 215 LPS biosynthesis protein |
| ycbG | 216 LPS biosynthesis protein |
| ycbH | 217 sugar transferase |
| ycbI | 218 sugar transferase |
| ycbj | 219 LPS biosynthesis protein |
| ycbk | 214 polysaccharide biosynthesis export protein |
| yijG | 899 glycosyl transferase |
| yjef | 949 lipopolysaccharide biosynthesis protein |
| ymF | 1297 glycosyl transferase |
| ymF | 1299 UDP-N-acetylgalactosamine 2-epimerase |
| yohH | 1478 lipopolysaccharide biosynthesis protein |
| yohJ | 1479 lipopolysaccharide biosynthesis protein |
| ystC | 1853 polysaccharide biosynthesis protein |
| ywaG | 2206 glycosyltransferase |
| ywaG | 2207 lipopolysaccharide biosynthesis protein |
| CELLULAR PROCESSES |  |
| Cell division |  |
| ezzA | 2225 cell division regulator |
| ftsA | 1940 cell division protein FtsA |
| ftsE | 1000 cell division ATP-binding protein FtsE |
| ftsH | 27 cell division protein FtsH |
| ftsK | 1705 cell division protein FtsK |
| ftsQ | 1632 cell division protein FtsQ |
| ftsW1 | 663 cell division protein FtsW |
| ftsW2 | 908 cell division protein FtsW |
| ftsX | 1001 cell division protein |

734 Genome Research www.genome.org
### Table 1. (Continued)

| Gene | Function                                      |
|------|-----------------------------------------------|
| **FtsY** | 825  | cell division protein FtsY                  |
| **FtsZ** | 1938 | cell division protein FtsZ                  |
| **GidA** | 1915 | glucose inhibited division protein GidA     |
| **GidB** | 1381 | glucose-inhibited division protein GidB     |
| **GidC** | 1257 | glucose inhibited division protein GidC     |
| **MesJ** | 24   | cell cycle protein MesJ                     |
| **ParA** | 99   | chromosome partitioning protein             |
| **RodA** | 917  | rod-shape determining protein               |
| **Smc**  | 812  | chromosome segregation SMC protein          |

**Cell killing**

| Gene | Function                                      |
|------|-----------------------------------------------|
| **Hly** | 498  | hemolysin like protein                       |

**Chaperones**

| Gene | Function                                      |
|------|-----------------------------------------------|
| **DnaK** | 979  | DnaK protein                   |
| **GroEL** | 400  | 60 KD chaperonin                |
| **GroES** | 399  | 10 KD chaperonin                |
| **SugE** | 25   | protein                        |

**Detoxification**

| Gene | Function                                      |
|------|-----------------------------------------------|
| **AhpC** | 336  | alkyd hydroperoxide reductase              |
| **AhpF** | 337  | alkyd hydroperoxide reductase              |
| **SodA** | 413  | superoxide dismutase                     |

**Protein and peptide secretion**

| Gene | Function                                      |
|------|-----------------------------------------------|
| **Ffh** | 1658 | signal recognition particle Ffh            |
| **IspA** | 1026 | lipoprotein signal peptidase               |
| **SecA** | 118  | preprotein translocase SecA subunit        |
| **SecE** | 2175 | preprotein translocase SecE subunit        |
| **SecG** | 967  | protein-export protein SecG                 |
| **SecY** | 2159 | preprotein translocase SecY subunit        |
| **SipL** | 2351 | signal peptidase I                         |
| **Tig**  | 536  | trigger factor                              |

**Transformation**

| Gene | Function                                      |
|------|-----------------------------------------------|
| **CoiA** | 1785 | competence protein CoiA                    |
| **ComC** | 2104 | type 4 preplin-like protein specific leader peptidase |
| **ComEA** | 1833 | competence protein ComEA                   |
| **ComEC** | 1832 | competence protein ComEC                   |
| **ComFA** | 1098 | competence protein ComFA                   |
| **ComFC** | 1097 | competence protein ComFC                   |
| **ComGA** | 2189 | competence protein ComGA                   |
| **ComGB** | 2188 | competence protein ComGB                   |
| **ComGC** | 2187 | competence protein ComGC                   |
| **ComGD** | 2187 | competence protein ComGD                   |
| **ComX**  | 2224 | competence regulator ComX                  |
| **DprA** | 1254 | DNA processing Smp protein                 |
| **RadA** | 2150 | DNA repair protein RadA                    |
| **RecQ** | 1874 | ATP-dependent DNA helicase RecQ            |

**Central intermediary metabolism**

| Gene | Function                                      |
|------|-----------------------------------------------|
| **MetK** | 1971 | S-adenosylmethionine synthetase             |
| **PcaC** | 2052 | gamma-carboxysucinolactone decarboxylase    |

**Amino sugars**

| Gene | Function                                      |
|------|-----------------------------------------------|
| **FemD** | 436  | phosphoglucoamine mutase                    |
| **GlmS** | 1035 | glucosamine-fructose-6-phosphate aminotransferase |
| **NagA** | 1374 | N-acetylglucosamine-6-phosphate deacytelase |
| **NagB** | 1615 | glucosamine-6-P isomerase                   |
| **YlfH** | 1157 | N-acetylglucosamine catabolic protein        |
| **YpcD** | 1524 | endo-beta-N-acetylgalactosaminidase         |

**Degradation of polysaccharides**

| Gene | Function                                      |
|------|-----------------------------------------------|
| **Agl**  | 1732 | alpha-gluosidase                            |
| **AmyL** | 1278 | alpha-amylase                               |
| **AmyY** | 1734 | alpha-amylase                               |

**Amino acids and amines**

| Gene | Function                                      |
|------|-----------------------------------------------|
| **AnsB** | 743  | L-asparaginase                               |
| **AraT** | 57   | aromatic amino acid specific aminotransferase |
| **ArcA** | 2115 | aromatic amino acid specific aminotransferase |
| **ArcB** | 2114 | ornithine carbamoyltransferase              |
| Gene | Description |
|------|-------------|
| arcC1 | carbamate kinase |
| arcC2 | carbamate kinase |
| arcC3 | carbamate kinase |
| arcT | aminotransferase |
| argF | ornithine carbamoyltransferase |
| bcaT | branched-chain amino acid aminotransferase |
| gadB | glutamate decarboxylase |
| hicD | L-2-hydroxyisocaproate dehydrogenase |
| ipd | indole-3-pyruvate decarboxylase |
| otcA | ornithine carbamoyltransferase |
| pdc | phenolic acid decarboxylase |
| pfs | 5'-methylthioadenosine/S-adenosylhomocysteine nucleosidase |
| sdaA | alpha-subunit L-serine dehydratase |
| sdaB | beta-subunit of L-serine dehydratase |
| yciA | amino acid amidohydrolase |
| yjiB | amino acid amidohydrolase |
| yjE | aminotransferase |
| ylE | 3-hydroxyisobutyrate dehydrogenase |
| dhaK | dihydroxyacetone kinase |
| dhaL | dihydroxyacetone kinase |
| dhaM | dihydroxyacetone kinase |
| glpD | glycerol-3-phosphate dehydrogenase |
| gpdA | glycerol-3-phosphate dehydrogenase |
| lctO | L-lactate oxidase |
| yjiB | oxidoreductase |
| atpA | ATP synthase alpha subunit |
| atpB | ATP synthase beta subunit |
| atpD | ATP synthase epsilon subunit |
| atpE | ATP synthase beta subunit |
| atpF | ATP synthase beta subunit |
| atpG | ATP synthase gamma subunit |
| atpH | ATP synthase delta subunit |
| cydA | cytochrome D ubiquinol oxidase subunit |
| cdB | cytochrome D ubiquinol oxidase subunit |
| fer | ferredoxin |
| nrdH | glutaredoxin-like protein |
| nrdL | ribonucleotide reductase |
| nifJ | pyruvate-flavodoxin oxidoreductase |
| nifS | pyridoxal-phosphate dependent aminotransferase |
| nifU | NifU protein |
| nizD | pyridoxal-phosphate dependent aminotransferase |
| qor | quinone oxidoreductase |
| ylfJ | NADPH-flavin oxidoreductase |
| ylfE | flavodoxin |
| yvcC | FMN-binding protein |
| kdgA | 2-dehydro-3-deoxyphosphogluconate aldolase |
| kdgk | 2-dehydro-3-deoxygluconokinase |
| ackA1 | acetate kinase |
| ackA2 | acetate kinase |
| adhA | alcohol dehydrogenase |
| adhE | alcohol-acetaldehyde dehydrogenase |
| aclD | alpha-acetolactate decarboxylase |
| aclC | alpha-acetoglutarate decarboxylase |
| als | alpha-acetolactate synthase |

**Table 1.** (Continued)

| Gene | Description |
|------|-------------|
| butA | acetoin reductase |
| butB | 2,3-butanediol dehydrogenase |
| frdC | fumarate reductase flavoprotein subunit |
| mae | malate dehydrogenase |
| pfl | pyruvate-formate lyase |
| pflA | pyruvate-formate lyase activating enzyme |
| pta | phosphate acetyltransferase |
| yse | 2-nitropropane deoxygenase |
| fbp | fructose-1,6-bisphosphatase |
| enoA | enolase |
| enoB | 2-phosphoglycerate dehydratase |
| fbaA | fructose-bisphosphate aldolase |
| gapA | glyceroldehyde 3-phosphate dehydrogenase |
| gapB | glyceroldehyde 3-phosphate dehydrogenase |
| ldh | L-lactate dehydrogenase |
| pdhA | PDHE1 component alpha subunit |
| pdhB | PDHE1 component beta subunit |
| pdhD | lipoamide dehydrogenase component of PDH complex |
| bglA | phospho-beta-glucosidase |
| bglH | beta-glucosidase |
| bglS | beta-glucosidase A |
| galE | UDP-glucose 4-epimerase |
| galK | galactokinase |
| galM | aldose 1-epimerase |
| galT | galactose-1-phosphate uridylyltransferase |
| glk | glucose kinase |
| gntK | glucanase kinase |
| gntZ | 6-phosphogluconate dehydrogenase |
| lacC | tagatose-6-phosphate kinase |
| lacZ | beta-galactosidase |
| maa | maltose O-acetyltransferase |
| malQ | 4-alpha-glucanotransferase |
| mflD | mannitol 1-phosphate dehydrogenase |
| pmf | mannose-6-phosphate isomerase |
| rbsK | ribokinase |
| Gene | Description |
|------|-------------|
| scrK | 1518 fructokinase |
| thgA | 2058 thiogalactoside acetyltransferase |
| uxaC | 1674 glucuronate isomerase |
| uxuA | 1678 D-mannionate dehydratase |
| uxuB | 1679 fructuronate reductase |
| xylA | 1550 xylose isomerase |
| xylB | 1548 xylulose kinase |
| xylM | 1674 aldose 1-epimerase |
| xylX | 1544 beta-1,4-xyllosidase |
| xynB | 1543 acetyltransferase hypothetical protein |
| yeeB | 443 sugar hydrolase |
| ygjD | 694 4-alpha-glucanotransferase |
| yidC | 834 beta-glucosidase |
| yncA | 1321 acetyltransferase |
| ypbG | 1519 sugar kinase |
| ypcA | 1521 beta-glucosidase |
| ypdB | 1532 sugar hydrolase |
| ypdD | 1537 sugar hydrolase |
| yrcA | 1722 phospho-beta-glucosidase |
| citB | 670 aconitate hydratase |
| citC | 1207 acetate-SH-citratelyase ligase |
| citD | 1208 citratelyase acyl-carrier protein |
| citE | 1209 citratelyase beta chain |
| citF | 1210 citratelyase alpha chain |
| citG | 1211 CitG protein |
|  | 672 isocitrate dehydrogenase |
| accA | 790 acetyl-CoA carboxylase carboxyltransferase subunit alpha |
| accB | 786 biotin carboxyl carrier protein of acetyl-CoA carboxylase |
| accC | 788 biotin carboxylase |
| accD | 789 acetyl-CoA carboxylase carboxyltransferase subunit beta |
| acpA | 782 acyl carrier protein |
| acpD | 116 acyl carrier protein phosphodiesterase |
| acpS | 862 acyl carrier protein synthase |
| cfa | 1972 cyclopropane fatty acid synthase |
| fabD | 783 malonyl CoA-acyl carrier protein transacylase |
| fabF | 786 3-oxoacyl-acyl carrier protein synthase II |
| fabG1 | 784 3-oxoacyl-acyl carrier protein reductase |
| fabG2 | 1843 3-oxoacyl-acyl carrier protein reductase |
| fabH | 782 3-oxoacyl-acyl-carrier protein synthase III |
| fabI | 562 NADH-dependent enoyl-ACP reductase |
| fabZ1 | 561 hydroxymyristoyl-acyl carrier protein dehydrogenase |
| fabZ2 | 787 3R-hydroxymyristoyl-acyl carrier protein dehydrogenase |
| fadA | 1843 acetyl coenzyme A acyltransferase |
| fadD | 655 long-chain acyl-CoA synthetase |
| hmcM | 1614 hydroxymethylglutaryl-CoA synthase |
| lplL | 65 lipoate-protein ligase |
| plsX | 72 fatty acid/phospholipid synthesis protein |
| thiL | 1613 acetyl coenzyme A acyltransferase |
| ydiD | 386 acyl carrier protein phosphodiesterase |
| yeaG | 408 mevalonate kinase |
| yeaH | 410 diposphomevalonate decarboxylase |
| yebA | 411 mevalonate kinase |
| yscE | 1830 lipase |

### Purines, Pyrimidines, Nucleosides and Nucleotides

| Gene | Description |
|------|-------------|
| 2′-deoxyribonucleotide metabolism |
| dcmA | 1156 dCMP deaminase |
| dcdA | 1156 ribonucleoside-diphosphate reductase alpha chain |
| dcdF | 1156 ribonucleoside-diphosphate reductase beta chain |
| nrdE | 1004 nucleotide and nucleoside interconversions |
| cmk | 1761 cytidine monophosphate kinase |
| dukA | 494 deoxyribonucleoside kinase |
| dukB | 1171 deoxyribonucleoside kinase |
| nucA | 1101 nucleotidase |
| pyrH | 2088 UMP Kinase |
| ycJ | 301 phosphatase |
| purA1 | 2029 adenylsuccinate synthase |
| purB1 | 1689 adenylsuccinate lyase |
| purC | 1578 phosphoribosylaminomimidazole-succinocarboxamide synthetase |
| purD | 1554 phosphoribosylamine-glycinamide carboxylase |
| purE | 1553 phosphoribosylpyrophosphate amidotransferase |
| purF | 1572 phosphoribosylpyrophosphate amidotransferase |
| purH | 1560 bifunctional purine biosynthesis protein |
| purK | 1552 phosphoribosylaminomimidazole carboxylase |
| purL | 1575 phosphoribosylformylglycinamidine synthase II |
| purM | 1566 phosphoribosylaminomimidazole synthetase |
| purN | 1565 phosphoribosylglycinamide formyltransferase |
| purQ | 1577 phosphoribosylformylglycinamidine synthase I |
| pydA | 1593 dihydroorotate dehydrogenase A |
| pydB | 1383 dihydroorotate dehydrogenase B |
| pydC | 1082 dihydroorotate |
| pydE | 1081 orotate phosphoribosyltransferase |
| pyf | 1382 orotidine-phosphate decarboxylase |
| pyZ | 1384 dihydroorotate dehydrogenase electron transfer subunit |
| thyA | 1583 thymidylate synthase |
| yeaB | 404 thymidylate kinase |

### Salvage of nucleosides and nucleotides

| Gene | Description |
|------|-------------|
| add | 2158 adenine deaminase |
| adk | 623 adenine phosphoribosyltransferase |
| apt | 2158 aspartate carbamoyltransferase |
| apt | 623 adenine phosphoribosyltransferase |
| cdd | 1463 cytidine deaminase |
| deoB | 956 phosphopentomutase |
| deoC | 1464 deoxyribosyl-phosphate aldolase |
| deoD | 957 purine nucleoside phosphorylase |
| gmk | 1967 guanylate kinase |
| hpt | 25 hypoxanthine-guanine phosphoryltransferase |
| ndd | 272 anaerobic ribonucleoside-triphosphate reductase |
| ndg | 273 anaerobic ribonucleoside-triphosphate reductase activating protein |
Table 1. (Continued)

| Gene | Description                        | Function                        |
|------|------------------------------------|---------------------------------|
| pdp  | pyrimidine-nucleoside phosphorylase|                                  |
| prsA | ribose-phosphate pyrophosphokinase |                                  |
| prsB | ribose-phosphate pyrophosphokinase |                                  |
| udk  | uridine kinase                     |                                  |
| udp  | uridine phosphoribosyltransferase  |                                  |
| upp  | uracil phosphoribosyltransferase   |                                  |
| xpt  | xanthine phosphoribosyltransferase |                                  |
| yfiG | thymidine kinase                   |                                  |
| Sugar-nucleotide biosynthesis and interconversions |
| cpsM | dTDP-4-keto-6-deoxyglucose-3,5-epimerase |                                  |
| rmlA | glucose-1-phosphatethymidylyltransferase |                                  |
| rmlB | dTDP-glucose4,6-dehydratase        |                                  |
| rmlC | dTDP-L-rhamnose synthase           |                                  |
| REGULATORY FUNCTIONS |
| General |
| ahrC | transcriptional regulator          |                                  |
| aldB | regulatory protein AldR            |                                  |
| argR | arginine catabolic regulator       |                                  |
| birA | bifunctional protein BirA          |                                  |
| codY | transcriptional regulator          |                                  |
| codZ | transcriptional regulator          |                                  |
| copR | transcriptional regulator          |                                  |
| fur  | ferric uptake regulator            |                                  |
| gadA | positive regulator                 |                                  |
| glnA | nitrogen regulatory protein P-II    |                                  |
| glnB | glutamine synthetase repressor     |                                  |
| gntR | transcriptional regulator          |                                  |
| nadR | transcriptional regulator          |                                  |
| rciA | transcriptional regulator          |                                  |
| rcfA | transcriptional regulator          |                                  |
| rcfB | transcriptional regulator          |                                  |
| relA | transcriptional regulator          |                                  |
| meA  | transcriptional regulator          |                                  |
| meB  | transcriptional regulator          |                                  |
| meC  | transcriptional regulator          |                                  |
| meD  | transcriptional regulator          |                                  |
| tagR | transcriptional regulator          |                                  |
| tenA | transcriptional regulator          |                                  |
| yabA | transcriptional regulator          |                                  |
| yabB | transcriptional regulator          |                                  |
| ybdA | transcriptional regulator          |                                  |
| ybdG | transcriptional regulator          |                                  |
| ybeD | transcriptional regulator          |                                  |
| ycfA | transcriptional regulator          |                                  |
| ycfB | transcriptional regulator          |                                  |
| ycfD | transcriptional regulator          |                                  |
| ycfG | transcriptional regulator          |                                  |
| ycfH | transcriptional regulator          |                                  |
| ycfI | transcriptional regulator          |                                  |
| ycfJ | transcriptional regulator          |                                  |
| Two-component systems |
| kinA | sensor protein kinase              |                                  |
| kinB | sensor protein kinase              |                                  |
| kinC | sensor protein kinase              |                                  |
| kinD | sensor protein kinase              |                                  |
| KinE | sensor protein kinase              |                                  |
| KinF | sensor protein kinase              |                                  |
| KinG | sensor protein kinase              |                                  |
| KinH | sensor protein kinase              |                                  |
| LysR-family regulators |
| fhuR | transcriptional regulator          |                                  |
| mopR | malolactic fermentation system     |                                  |
| AraC-family regulators |
| adaA | methylphosphotriester-DNA alkyltransferase |                                  |
| xylR | xylose operon repressor            |                                  |
| yhiR | transcriptional regulator          |                                  |
| yhiS | transcriptional regulator          |                                  |
| yhiT | transcriptional regulator          |                                  |
| GntR-family regulators |
| busR | transcriptional regulator          |                                  |
| kdgR | transcriptional regulator          |                                  |
| Table 1. | (Continued) |
|----------|-------------|
| **rgrA** | 437 transcriptional regulator |
| **rgrB** | 1461 transcriptional regulator |
| **DeoR-family regulators** |
| **citR** | 1206 citrate lyase regulator |
| **lacR** | 984 lactose transport regulator |
| **rdrA** | 797 transcriptional regulator |
| **rdrB** | 1332 transcriptional regulator |
| **MarR-family regulators** |
| **rmaA** | 750 transcriptional regulator |
| **rmaB** | 715 transcriptional regulator |
| **rmaC** | 1503 transcriptional regulator |
| **rmaD** | 115 transcriptional regulator |
| **rmaE** | 1511 transcriptional regulator |
| **rmaF** | 1341 transcriptional regulator |
| **rmaG** | 781 transcriptional regulator |
| **rmaH** | 932 transcriptional regulator |
| **rmaI** | 1583 transcriptional regulator |
| **rmaJ** | 584 transcriptional regulator |
| **zitR** | 2185 zinc transport transcriptional regulator |
| **BglG-family regulators** |
| **bglR** | 1493 beta-glucoside operon antiterminator |
| **GTP-binding proteins** |
| **eraL** | 355 GTP-binding protein Era |
| **hflX** | 225 GTP-binding protein HflX |
| **obgL** | 1630 GTP-binding protein Obg |
| **thfD** | 2238 GTP-binding protein ThfD |
| **typA** | 2094 GTP-binding protein TypA/BipA |
| **yphL** | 1330 GTP-binding protein |
| **yqeL** | 224 GTP-binding protein |
| **ysxL** | 1165 GTP-binding protein |
| **yyaL** | 12 GTP-binding protein |
| **REPLICATION** |
| **Degradation of DNA** |
| **exoA** | 799 exodeoxyribonuclease A |
| **recB** | 5 subunit B of ATP-dependent exonuclease |
| **recC** | 8 subunit A of ATP-dependent exonuclease |
| **sbcA** | 1354 ATP-dependent dsDNA exonuclease |
| **sbcD** | 1357 exonuclease SbcD |
| **uvrA** | 1887 exonuclease ABC subunit A |
| **uvrB** | 557 exonuclease ABC subunit B |
| **uvrC** | 857 exonuclease ABC subunit C |
| **xseA** | 878 exonuclease VII large subunit |
| **xseB** | 879 exonuclease VII small subunit |
| **DNA replication, restriction, modification, recombination, and repair** |
| **cshA** | 100 chromosome segregation helicase |
| **dinG** | 1900 ATP-dependent helicase DinG |
| **dnaA** | 1 replication initiation protein DnaA |
| **dnaB** | 758 replication protein DnaB |
| **dnaC** | 754 replicative DNA helicase |
| **dnaD** | 1083 DNA replication protein DnaD |
| **dnaE** | 496 DNA polymerase III, alpha chain 2 |
| **dnaG** | 545 DNA primase |
| **dnaH** | 2279 DNA polymerase III, subunits beta and tau |
| **dnapA** | 759 primosomal protein Dnap |
| **dnapB** | 2308 Dnap protein |
| **dnaN** | 2 DNA polymerase III, beta chain |
| **dnaQ** | 1010 DNA polymerase III, epsilon chain |
| **gyrA** | 1123 DNA gyrase subunit A |
| **gyrB** | 929 DNA gyrase subunit B |
| **hexA** | 2294 mismatch repair protein MutS |
| **TRANSCRIPTION** |
| **Degradation of RNA** |
| **pnpA** | 1923 5' to 3' phosphodiesterase nucleotidyltransferase |
| **pac** | 810 ribonuclease III |
| **mnhA** | 2350 ribonuclease Hil |
| **mnhB** | 1329 ribonuclease Hil |
| **vacB** | 968 ribonuclease |
| **vacC** | 1227 ribonuclease |
| **RNA synthesis, modification, and DNA transcription** |
| **greA** | 626 transcription elongation factor GreA |
| **mfd** | 19 transcription-repair coupling factor |
| **nusA** | 774 transcription termination protein NusA |
| **nusB** | 693 transcription termination protein NusB |
| **nusG** | 2174 transcription antitermination protein |
| **queA** | 1617 S-adenosylhomocysteine tRNA ribosyltransferase |
| **rluA** | 2182 pseudouridylate synthetase |
| **rluB** | 1308 pseudouridine synthetase |
| **rluC** | 1390 pseudouridine synthetase |
| **rluD** | 1027 pseudouridine synthetase |
| **rluE** | 368 pseudouridine synthetase |
| **rpoA** | 2153 DNA-directed RNA polymerase alpha chain |
| **rpoB** | 1863 DNA-directed RNA polymerase beta chain |
| **rpoC** | 1859 DNA-directed RNA polymerase beta chain |
| **rpoD** | 547 major RNA polymerase sigma factor |
Table 1. Continued

| Protein | Accession | Function |
|---------|-----------|----------|
| rpoE    | 624       | DNA-directed RNA polymerase delta chain |
| rrmA    | 1365      | rRNA methyltransferase |
| rsuA    | 2327      | rRNA methyltransferase |
| sigX    | 2243      | RNA polymerase ECF sigma factor |
| smpB    | 1777      | tmRNA-binding protein SmpB |
| sunL    | 1958      | rRNA methylase |
| trmA    | 1607      | tRNA methyltransferase |
| yfjD    | 597       | tRNA/rRNA methyltransferase |
| rRNA processing |
| rheA    | 354       | ATP-dependent RNA helicase |
| rheB    | 416       | ATP-dependent RNA helicase |
| rimM    | 1607      | 16S rRNA methylase |

TRANSLATION

Aminoacyl tRNA synthetases

| Protein | Accession | Function |
|---------|-----------|----------|
| alaS    | 1780      | alanyl-tRNA synthetase |
| argS    | 2117      | arginyl-tRNA synthetase |
| asnS    | 1896      | asparaginyl-tRNA synthetase |
| aspS    | 2041      | aspartyl-tRNA synthetase |
| cysS    | 1919      | cysteinyl-tRNA synthetase |
| gltX    | 2141      | glutamyl-tRNA synthetase |
| glyS    | 1102      | glycyl-tRNA synthetase |
| hisS    | 2043      | histidyl-tRNA synthetase |
| ileS    | 1933      | isoleucyl-tRNA synthetase |
| leuS    | 829       | leucyl-tRNA synthetase |
| lysS    | 377       | lysyl-tRNA synthetase |
| metS    | 800       | methionyl-tRNA synthetase |
| pheS    | 2010      | phenylalanyl-tRNA synthetase |
| proS    | 2197      | prolyl-tRNA synthetase |
| serS    | 1768      | seryl-tRNA synthetase |
| traS    | 391       | tryptophanyl-tRNA synthetase |
| valS    | 2250      | valyl-tRNA synthetase |

Degradation of proteins, peptides, and glycopeptides

| Protein | Accession | Function |
|---------|-----------|----------|
| gcp     | 294       | O-sialoglycoprotein endopeptidase |
| htrA    | 2205      | Exported serine protease |
| pepA    | 394       | Glutamyl aminopeptidase |
| pepC    | 1948      | Aminopeptidase C |
| pepD    | 249       | Dipeptidase |
| pepDB   | 1601      | Dipeptidase |
| pepF    | 1784      | Oligopeptidase F |
| pepM    | 601       | Mathionine aminopeptidase |
| pepN    | 304       | Aminopeptidase N |
| pepO    | 1867      | Neutral endopeptidase |
| pepP    | 691       | Aminopeptidase P |
| pepQ    | 1698      | Proline dipeptidase |
| pepT    | 1878      | Tripeptidase |

Protein modification

| Protein | Accession | Function |
|---------|-----------|----------|
| def     | 555       | Polypeptide deformylase |
| pknB    | 1956      | Serine/threonine protein kinase |
| pmpA    | 1782      | Protein maturation protein |
| pmsR    | 2085      | Peptide methionine sulfoxide reductase |
| pmsX    | 1594      | Peptide methionine sulfoxide reductase |
| ppiA    | 3689      | Peptidyl-prolyl cis-trans isomerase A |
| ppiB    | 914       | Peptidyl-prolyl cis-trans isomerase C |
| ppl     | 1957      | Protein serine/threonine phosphatase |

Table 1. Continued

| Protein | Accession | Function |
|---------|-----------|----------|
| ptpL    | 2284      | Protein-tyrosine phosphatase |
| ytaD    | 1905      | Protein-tyrosine phosphatase |

Ribosomal proteins: synthesis and modification

| Protein | Accession | Function |
|---------|-----------|----------|
| rpmA    | 105       | Methyltransferase |
| rplA    | 2079      | 50S ribosomal protein L1 |
| rplB    | 2168      | 50S ribosomal protein L2 |
| rplC    | 2170      | 50S ribosomal protein L3 |
| rplD    | 2169      | 50S ribosomal protein L4 |
| rplE    | 2164      | 50S ribosomal protein L5 |
| rplF    | 2162      | 50S ribosomal protein L6 |
| rplG    | 753       | 50S ribosomal protein L9 |
| rplH    | 1302      | 50S ribosomal protein L10 |
| rplI    | 2080      | 50S ribosomal protein L11 |
| rplJ    | 1301      | 50S ribosomal protein L12 |
| rplK    | 2347      | 50S ribosomal protein L13 |
| rplL    | 2165      | 50S ribosomal protein L14 |
| rplM    | 2160      | 50S ribosomal protein L15 |
| rplN    | 2166      | 50S ribosomal protein L16 |
| rplO    | 2152      | 50S ribosomal protein L17 |
| rplP    | 2152      | 50S ribosomal protein L18 |
| rplQ    | 2153      | 50S ribosomal protein L19 |
| rplR    | 2151      | 50S ribosomal protein L20 |
| rplS    | 2151      | 50S ribosomal protein L21 |
| rplT    | 2167      | 50S ribosomal protein L22 |
| rplW    | 2169      | 50S ribosomal protein L23 |
| rplX    | 2165      | 50S ribosomal protein L24 |
| rpmA    | 1091      | 50S ribosomal protein L27 |
| rpmB    | 196       | 50S ribosomal protein L28 |
| rpmC    | 2166      | 50S ribosomal protein L29 |
| rpmD    | 2160      | 50S ribosomal protein L30 |
| rpmE    | 1640      | 50S ribosomal protein L31 |
| rpmF    | 96        | 50S ribosomal protein L32 |
| rpmG    | 622       | 50S ribosomal protein L33 |
| rpmGB   | 96        | 50S ribosomal protein L33 |
| rpmGC   | 2175      | 50S ribosomal protein L33 |
| rpmH    | 134       | 50S ribosomal protein L34 |
| rpmI    | 1912      | 50S ribosomal protein L35 |
| rpmJ    | 2154      | 50S ribosomal protein L36 |
| rpmK    | 854       | 30S ribosomal protein S1 |
| rpmL    | 2228      | 30S ribosomal protein S2 |
| rpmM    | 2166      | 30S ribosomal protein S3 |
| rpmN    | 284       | 30S ribosomal protein S4 |
| rpmO    | 2161      | 30S ribosomal protein S5 |
| rpmP    | 2275      | 30S ribosomal protein S6 |
| rpmQ    | 2355      | 30S ribosomal protein S7 |
| rpmR    | 2162      | 30S ribosomal protein S8 |
| rpmS    | 2347      | 30S ribosomal protein S9 |
| rpmT    | 2170      | 30S ribosomal protein S10 |
| rpmU    | 2153      | 30S ribosomal protein S11 |
| rpmV    | 2355      | 30S ribosomal protein S12 |
| rpmW    | 2154      | 30S ribosomal protein S13 |
| rpmX    | 2164      | 30S ribosomal protein S14 |
| rpmY    | 911       | 30S ribosomal protein S15 |
| rpmZ    | 2165      | 30S ribosomal protein S16 |
| rpmA2   | 2165      | 30S ribosomal protein S17 |
| rpmB2   | 2274      | 30S ribosomal protein S18 |
| rpmC2   | 2167      | 30S ribosomal protein S19 |
| rpmD2   | 2179      | 30S ribosomal protein S20 |
| rpmE2   | 2237      | 30S ribosomal protein S21 |
| rpmF2   | 293       | Acetyltransferase |
| rpmG2   | 293       | Acetyltransferase |
| rpmH2   | 740       | Acetyltransferase |
| rpmI2   | 798       | Acetyltransferase |
| rpmJ2   | 776       | Probable ribosomal protein |
| rpmK2   | 1962      | Methyl-5'-tRNA formyltransferase |
| gatA    | 166       | Glu-tRNA amido transferase subunit A |
| Table 1. (Continued) |
|----------------------|
| **Translation factors** |
| gatB 168 Glu-tRNA amidotransferase subunit B |
| gatC 165 Glu-tRNA amidotransferase subunit C |
| ksgA 690 kasugamycin dimethyltransferase |
| mpa 132 ribonuclease P protein component |
| tgt 156 queine tRNA-ribosyltransferase |
| trmH 1942 tRNA-guanosine methyltransferase |
| trnU 853 tRNA-methyltransferase |
| truA 485 tRNA pseudouridine synthase A |
| truB 1141 tRNA pseudouridine synthase B |
| **TRANSPORT AND BINDING PROTEINS** |
| **General** |
| ecsA 2075 ABC transporter ATP binding protein |
| ecsB 2074 ABC transporter permease protein |
| mscI 2171 large conductance mechanosensitive channel protein |
| yabE 16 ABC transporter ATP-binding protein |
| yahG 74 ABC transporter ATP binding protein |
| yaiE 87 transporter |
| yajA 90 transporter |
| ybaB 102 ABC transporter ATP binding protein |
| ycfB 251 ABC transporter ATP binding protein |
| ycfC 252 ABC transporter permease protein |
| ycfI 260 ABC transporter ATP binding protein |
| ycfG 262 ABC transporter ATP binding protein |
| ychD 276 ABC transporter ATP-binding protein |
| ychE 277 ABC transporter ATP-binding protein |
| ychf 278 ABC transporter permease protein |
| ydaG 310 ABC transporter ATP binding and permease protein |
| ydbA 312 ABC transporter ATP binding and permease protein |
| ydcE 325 ABC transporter ATP binding protein |
| ydcF 326 ABC transporter permease protein |
| ydaA 382 permease |
| yfaC 520 ABC transporter ATP binding protein |
| yfcB 521 ABC transporter permease protein |
| yfgE 563 ABC transporter ATP binding protein |
| ygfF 564 ABC transporter permease protein |
| ygfA 652 ABC transporter ATP-binding protein |
| ygfB 653 ABC transporter permease protein |
| yhcA 721 ABC transporter ATP-binding and permease protein |
| yif 886 transporter |
| yijC 894 ABC transporter permease protein |
| yijD 895 ABC transporter ATP binding protein |
| yjcA 921 ABC transporter ATP binding protein |
| yjcC 993 ABC transporter ATP-binding protein |
| yijD 994 ABC transporter permease protein |
| yjf 996 transporter |

| Table 1. (Continued) |
|----------------------|
| **Amino acids, peptides and amines** |
| arcD1 2112 arginine/ornitine antiporter |
| arcD2 2107 arginine/ornitine antiporter |
| bmQ 685 branched chain amino acid permease |
| busAA 1475 betaine ABC transporter ATP binding protein |
| busAB 1474 betaine ABC transporter permease and substrate binding protein |
| choQ 865 choline ABC transporter ATP binding protein |
| choS 867 choline ABC transporter permease and substrate binding protein |
| ctrA 113 cationic amino acid transporter |
| dtpT 705 di-/tri-peptidetransporter |
| gmdC 1326 glutamate-gamma-aminobutyrate antiporter |
| glnP 1818 glutamine ABC transporter permease and substrate binding protein |
| glnQ 1819 glutamine ABC transporter ATP-binding protein |
| glkP 1856 glutamate ABC transporter permease protein |
| glkQ 1855 glutamate ABC transporter ATP-binding protein |
| gltS 559 glutamate or arginine ABC transporter substrate binding protein |
| lysP 2277 lysine specific permease |
| lysQ 370 lysine specific permease |
| oppA 1906 oligopeptide ABC transporter substrate binding protein |
| oppB 1908 oligopeptide ABC transporter permease protein |
| oppC 1907 oligopeptide ABC transporter permease protein |
| oppD 1910 oligopeptide ABC transporter ATP binding protein |
| oppF 1909 oligopeptide ABC transporter ATP binding protein |
| Gene | Description |
|------|-------------|
| optA | oligopeptide ABC transporter substrate binding protein |
| optB | oligopeptide ABC transporter permease protein |
| optC | oligopeptide ABC transporter permease protein |
| optD | oligopeptide ABC transporter ATP binding protein |
| optF | oligopeptide ABC transporter ATP binding protein |
| optS | oligopeptide ABC transporter substrate binding protein |
| potA | spermidine/putrescine ABC transporter ATP-binding protein |
| potB | spermidine/putrescine ABC transporter permease protein |
| potC | spermidine/putrescine ABC transporter permease protein |
| potD | spermidine/putrescine ABC transporter substrate binding protein |
| yagE | amino acid permease |
| ydcB | amino acid ABC transporter ATP binding protein |
| ydcC | amino acid ABC transporter permease protein |
| ydcG | amino acid permease |
| yjgD | amino acid ABC transporter substrate binding protein |
| yjgE | amino acid ABC transporter ATP binding protein |
| ylcA | amino acid permease |
| yqfD | amino acid permease |
| ysfD | amino acid permease |
| yvfD | amino acid ABC transporter substrate binding protein |
| phnA | alkylphosphonate uptake protein |
| phnB | phosphonate ABC transporter permease protein |
| phnC | phosphonate ABC transporter ATP-binding protein |
| phnE | phosphonate ABC transporter permease protein |
| pstA | phosphate ABC transporter ATP-binding protein |
| pstB | phosphate ABC transporter ATP binding protein |
| pstC | phosphate ABC transporter permease protein |
| pstD | phosphate ABC transporter permease protein |
| pstE | phosphate ABC transporter substrate binding protein |
| pstF | phosphate ABC transporter substrate binding protein |
| yafB | sulfate transporter |
| glpF | glycerol uptake facilitator |
| gntP | glucose permease |
| lacS | lactose permease |
| maeP | malate permease |
| malE | maltose ABC transporter substrate binding protein |
| malF | maltose ABC transporter permease protein |
| malG | maltose ABC transporter permease protein |
| mleP | maltose ABC transporter ATP binding protein |
| mspK | multiple sugar ABC transporter ATP-binding protein |
| rbsA | ribose ABC transporter ATP binding protein |
| rbsB | ribose ABC transporter substrate binding protein |
| rbsC | ribose ABC transporter permease protein |
| rbsD | ribose ABC transporter permease protein |
| rggC | polysaccharide ABC transporter permease protein |
| tagG | teichoic acid ABC transporter permease protein |
| tagH | teichoic acid ABC transporter ATP binding protein |
| uuxT | Na-galactoside symporter |
| yxIT | D-xylene proton-symporter |
| xynT | xylose symporter |
| yngE | sugar ABC transporter ATP binding protein |
| yngF | sugar ABC transporter permease protein |
| yngG | sugar ABC transporter permease protein |
| ypgD | sugar transport symporter |
| ypcG | sugar ABC transporter substrate binding protein |
| ypcH | sugar ABC transporter permease protein |
| ypdA | sugar ABC transporter substrate binding protein |
| yqgE | transporter |
| yvdD | transporter |
| amtB | ammonium transporter |
| cadA | cadmium efflux ATPase |
| copA | copper/potassium-transporting ATPase |
| copB | copper-potassium-transporting ATPase B |
| feoA | ferrous iron transport protein A |
| feoB | ferrous iron transport protein B |
| fhuB | ferrichrome ABC transporter permease protein |
| fhuC | ferrichrome ABC transporter ATP binding protein |
| fhuD | ferrichrome ABC transporter substrate binding protein |
| fhuG | ferrichrome ABC transporter permease protein |
| kupA | potassium uptake protein |
| kupB | potassium uptake protein |
| magA | cation-transporting P-ATPase |
| mlsA | manganese ABC transporter substrate binding protein |
| mtsB | manganese ABC transporter ATP binding protein |
| mtsC | manganese ABC transporter permease protein |
| nah | Na+/H+ antiporter |
| pacL | cation-transporting ATPase |
| ydaE | cation transporter |
| yddA | transporter |
| ydiF | Na+/H+ antiporter |
| yfgQ | cation-transporting ATPase |
| ygfE | divalent cation transport-related protein |
| Gene | Description |
|------|-------------|
| yieF | Mercuric reductase |
| yjdJ | Potassium channel protein |
| yliI | Cation-transporting ATPase |
| yndG | Metal ABC transporter substrate binding protein |
| yoaB | Cation-transporting ATPase |
| yogJ | Cation transporter |
| ypbB | Cationic transporter |
| yqeI | Cation transporter |
| yqgG | Cation transport ATPase |
| ysdE | Cation transporter |
| ytdB | Manganese transporter |
| yuiA | Metal transport ATPase |
| yxdC | Cation-transporting ATPase |
| zitP | Zinc ABC transporter permease protein |
| zitQ | Zinc ABC transporter ATP binding protein |
| Nucleosides, purines and pyrimidines |
| pbuX | Xanthine permease |
| pnuC1 | Nicotinamide mononucleotide transporter |
| pnuC2 | Nicotinamide mononucleotide transporter |
| pyrP | Uracil permease |
| PTS system |
| celB | Cellulose-specific PTS system IIC component |
| fruA | Fructose-specific PTS system enzyme IIIBC component |
| mtlA | Mannitol-specific PTS system IIIBC component |
| mtlF | Mannitol-specific PTS system IIA component |
| ptbA | Beta-glucoside-specific PTS system IIABC component |
| ptcA | Cellulose-specific PTS system IIAB component |
| ptcB | Cellulose-specific PTS system IIIB component |
| ptcC | Cellulose-specific PTS system IIC component |
| ptnA8 | Mannose-specific PTS system component IIAB |
| ptnC | Mannose-specific PTS system component IIC |
| ptnD | Mannose-specific PTS system component IIB |
| ptsH | Phosphocarrier protein Hpr |
| ptsI | Phosphoenolpyruvate protein |
| ptsK | Hpr(Ser) kinase |
| yidB | Cellulose-specific PTS system IIC component |
| yldD | Succrose-specific PTS system IIABC component |
| yleE | Beta-glucoside-specific PTS system IIABC component |
| Multidrug resistance |
| bld | Multidrug efflux transporter |
| cydC | Cytochrome D ABC transporter ATP binding and permease protein |
| cydD | Cytochrome D ABC transporter ATP binding and permease protein |
| 

| Table 1. (Continued) |

| Lactococcus lactis IL1403 Genome Sequence |

| Gene | Description |
|------|-------------|
| ImrP | Integral membrane protein |
| napC | Multidrug-efflux transporter |
| pmrA | Multidrug resistance efflux pump |
| pmrB | Multidrug resistance efflux pump |
| yidF | Transporter |
| ycdH | Transporter |
| ydiC | Efflux pump antibiotic resistance protein |
| ylfF | Membrane-bound transport protein |
| yjeE | Multidrug resistance protein |
| yniG | Drug-export protein |
| ypeE | Transporter |
| ypiB | Transporter |
| yqiI | Multidrug transporter |
| yweA | Membrane protein |
| yxbD | Transporter |
| OTHER CATEGORIES |
| 

| Gene | Description |
|------|-------------|
| ArsC | Arsenate reductase |
| ClpB | ClpB protein |
| ClpC | ATP-dependent protease ATP-binding subunit |
| ClpE | ATP-dependent protease ATP-binding subunit |
| ClpP | ATP-dependent Clp protease proteolytic subunit |
| ClpX | ATP-dependent Clp protease |
| Cpo | Non-heme chloride peroxidase |
| CspD | Cold shock protein D |
| CspE | Cold shock protein E |
| CtsA | Carbon starvation protein |
| CtsR | Transcriptional regulator |
| DinF | Damage-inducible protein DinF |
| DinP | DNA-damage-inducible protein P |
| DpsA | Non-heme iron-binding ferritin |
| GrpE | Stress response protein GrpE |
| HrcA | Heat-inducible transcription repressor HrcA |
| OsmC | Osmotically inducible protein |
| PhoL | Phosphate starvation inducible protein |
| Tpx | Thiol peroxidase |
| YbjA | Reductase |
| YbiE | General stress protein GSP13 |
| Drug and analog sensitivity |
| BacA | Undecaprenol kinase |
| Bar | Acryltransferase |
| PacA | Penicillin acylase |
| PacB | Penicillin acylase |
| YmdC | Kanamycin kinase |
| Phage related functions and prophages |
| pi101 | Prophage pi1 protein 01, integrase |
| pi102 | Prophage pi1 protein 02 |
| pi103 | Prophage pi1 protein 03, transcriptional regulator |
| pi104 | Prophage pi1 protein 04, transcriptional regulator |
| pi105 | Prophage pi1 protein 05 |
| pi106 | Prophage pi1 protein 06 |
| pi107 | Prophage pi1 protein 07 |
| pi108 | Prophage pi1 protein 08 |
| pi109 | Prophage pi1 protein 09 |
| pi110 | Prophage pi1 protein 10, transcriptional regulator |
| pi111 | Prophage pi1 protein 11, recombinase |
| pi112 | Prophage pi1 protein 12 |
| pi113 | Prophage pi1 protein 13, replisome organisser |
| pi114 | 456 | prophage pi1 protein 14, DNA replication protein |
| pi115 | 457 | prophage pi1 protein 15 |
| pi116 | 457 | prophage pi1 protein 16 |
| pi117 | 458 | prophage pi1 protein 17 |
| pi118 | 458 | prophage pi1 protein 18 |
| pi119 | 459 | prophage pi1 protein 19 |
| pi120 | 459 | prophage pi1 protein 16, deoxyuridine 5'-triphosphate nucleotidohydrolase |
| pi121 | 459 | prophage pi1 protein 21 |
| pi122 | 460 | prophage pi1 protein 22 |
| pi123 | 460 | prophage pi1 protein 23 |
| pi124 | 460 | prophage pi1 protein 24 |
| pi125 | 461 | prophage pi1 protein 25 |
| pi126 | 461 | prophage pi1 protein 26 |
| pi127 | 462 | prophage pi1 protein 27 |
| pi128 | 462 | prophage pi1 protein 28 |
| pi129 | 462 | prophage pi1 protein 29 |
| pi130 | 462 | prophage pi1 protein 30 |
| pi131 | 463 | prophage pi1 protein 31 |
| pi132 | 464 | prophage pi1 protein 32 |
| pi133 | 464 | prophage pi1 protein 33, terminase small subunit |
| pi134 | 465 | prophage pi1 protein 34, terminase large subunit |
| pi135 | 466 | prophage pi1 protein 35 |
| pi136 | 467 | prophage pi1 protein 36, prophage protease |
| pi137 | 468 | prophage pi1 protein 37, capsid protein |
| pi138 | 469 | prophage pi1 protein 38 |
| pi139 | 470 | prophage pi1 protein 39 |
| pi140 | 470 | prophage pi1 protein 40, tail component |
| pi141 | 471 | prophage pi1 protein 41, tail component |
| pi142 | 471 | prophage pi1 protein 42, small major structural protein |
| pi143 | 472 | prophage pi1 protein 43 |
| pi144 | 474 | prophage pi1 protein 44, tail component |
| pi145 | 477 | prophage pi1 protein 45, tail component |
| pi146 | 480 | prophage pi1 protein 46, tail component |
| pi147 | 482 | prophage pi1 protein 47 |
| pi148 | 482 | prophage pi1 protein 48, holin |
| pi149 | 483 | prophage pi1 protein 49, muramidase |
| pi150 | 1037 | prophage pi2 protein 01, integrase |
| pi151 | 1038 | prophage pi2 protein 02 |
| pi152 | 1039 | prophage pi2 protein 03 |
| pi153 | 1039 | prophage pi2 protein 04, hypothetical protein |
| pi154 | 1040 | prophage pi2 protein 05 |
| pi155 | 1040 | prophage pi2 protein 06 |
| pi156 | 1041 | prophage pi2 protein 07 |
| pi157 | 1041 | prophage pi2 protein 08 |
| pi158 | 1042 | prophage pi2 protein 09 |
| pi159 | 1042 | prophage pi2 protein 10 |
| pi160 | 1043 | prophage pi2 protein 11, topoisomerase binding protein |
| pi161 | 1043 | prophage pi2 protein 12, single strand binding protein |
| pi162 | 1044 | prophage pi2 protein 13, replisome |
| pi163 | 1045 | prophage pi2 protein 14 |
| pi164 | 1045 | prophage pi2 protein 15 |
| pi165 | 1046 | prophage pi2 protein 16 |
| pi166 | 1046 | prophage pi2 protein 17 |
| pi167 | 1047 | prophage pi2 protein 18 |
| pi168 | 1048 | prophage pi2 protein 19 |
| pi169 | 1048 | prophage pi2 protein 20, hypothetical protein |
| pi170 | 1049 | prophage pi2 protein 21, deoxyuridine 5'-triphosphate nucleotidohydrolase |
| pi171 | 1049 | prophage pi2 protein 22 |

| pi223 | 1049 | prophage pi2 protein 23 |
| pi224 | 1050 | prophage pi2 protein 24 |
| pi225 | 1051 | prophage pi2 protein 25 |
| pi226 | 1051 | prophage pi2 protein 26 |
| pi227 | 1052 | prophage pi2 protein 27 |
| pi228 | 1052 | prophage pi2 protein 28 |
| pi229 | 1053 | prophage pi2 protein 29 |
| pi230 | 1054 | prophage pi2 protein 30, terminase |
| pi231 | 1055 | prophage pi2 protein 31 |
| pi232 | 1056 | prophage pi2 protein 32 |
| pi233 | 1057 | prophage pi2 protein 33, capsid protein |
| pi234 | 1058 | prophage pi2 protein 34 |
| pi235 | 1059 | prophage pi2 protein 35 |
| pi236 | 1059 | prophage pi2 protein 36 |
| pi237 | 1059 | prophage pi2 protein 37 |
| pi238 | 1060 | prophage pi2 protein 38 |
| pi239 | 1060 | prophage pi2 protein 39 |
| pi240 | 1061 | prophage pi2 protein 40 |
| pi241 | 1061 | prophage pi2 protein 41 |
| pi242 | 1062 | prophage pi2 protein 42 |
| pi243 | 1064 | prophage pi2 protein 43 |
| pi244 | 1065 | prophage pi2 protein 44 |
| pi245 | 1068 | prophage pi2 protein 45 |
| pi246 | 1069 | prophage pi2 protein 46 |
| pi247 | 1069 | prophage pi2 protein 47 |
| pi248 | 1070 | prophage pi2 protein 48 |
| pi249 | 1070 | prophage pi2 protein 49 |
| pi250 | 1070 | prophage pi2 protein 50 |
| pi251 | 1071 | prophage pi2 protein 51, holin |
| pi252 | 1071 | prophage pi2 protein 52, muramidase |
| pi301 | 1414 | prophage pi3 protein 01 |
| pi302 | 1415 | prophage pi3 protein 02 |
| pi303 | 1415 | prophage pi3 protein 03 |
| pi304 | 1416 | prophage pi3 protein 04 |
| pi305 | 1416 | prophage pi3 protein 05, muramidase |
| pi306 | 1417 | prophage pi3 protein 06, holin |
| pi307 | 1418 | prophage pi3 protein 07 |
| pi308 | 1419 | prophage pi3 protein 08 |
| pi309 | 1420 | prophage pi3 protein 09 |
| pi310 | 1421 | prophage pi3 protein 10 |
| pi311 | 1422 | prophage pi3 protein 11 |
| pi312 | 1424 | prophage pi3 protein 12 |
| pi313 | 1425 | prophage pi3 protein 13, tail component |
| pi314 | 1428 | prophage pi3 protein 14 |
| pi315 | 1431 | prophage pi3 protein 15 |
| pi316 | 1431 | prophage pi3 protein 16, tail component |
| pi317 | 1432 | prophage pi3 protein 17, major tail protein |
| pi318 | 1433 | prophage pi3 protein 18, tail component |
| pi319 | 1433 | prophage pi3 protein 19, tail component |
| pi320 | 1433 | prophage pi3 protein 20, head-tail joining protein |
| pi321 | 1434 | prophage pi3 protein 21 |
| pi322 | 1435 | prophage pi3 protein 22, major head protein precursor |
| pi323 | 1436 | prophage pi3 protein 23, ATP dependent Clp protease |
| pi324 | 1436 | prophage pi3 protein 24 |
| pi325 | 1437 | prophage pi3 protein 25, head-tail joining protein |
| pi326 | 1438 | prophage pi3 protein 26, terminase large subunit |
| pi327 | 1439 | prophage pi3 protein 27, terminase small subunit |
| pi328 | 1440 | prophage pi3 protein 28 |
| pi329 | 1440 | prophage pi3 protein 29 |
| pi330 | 1441 | prophage pi3 protein 30 |
| pi331 | 1441 | prophage pi3 protein 31 |
| pi332 | 1442 | prophage pi3 protein 32 |
| Protein | Accession |
|---------|-----------|
| 1443    | prophage | protein 33 |
| 1443    | prophage | protein 34 |
| 1443    | prophage | protein 35, deoxyuridine 5'-triphosphate nucleotidohydrolase |
| 1444    | prophage | protein 36 |
| 1444    | prophage | protein 37 |
| 1445    | prophage | protein 38 |
| 1445    | prophage | protein 39 |
| 1446    | prophage | protein 40 |
| 1446    | prophage | protein 41 |
| 1446    | prophage | protein 42 |
| 1447    | prophage | protein 43 |
| 1447    | prophage | protein 44 |
| 1447    | prophage | protein 45 |
| 1448    | prophage | protein 46, DNA replication protein |
| 1449    | prophage | protein 47, repressor |
| 1450    | prophage | protein 48, single strand binding helix destabilising protein |
| 1450    | prophage | protein 49 |
| 1451    | prophage | protein 50 |
| 1451    | prophage | protein 51 |
| 1452    | prophage | protein 52 |
| 1452    | prophage | protein 53 |
| 1452    | prophage | protein 54 |
| 1453    | prophage | protein 55, antirepressor |
| 1453    | prophage | protein 56, cro-like repressor |
| 1454    | prophage | protein 57, cI-like repressor |
| 1455    | prophage | protein 58 |
| 1455    | prophage | protein 59 |
| 1456    | prophage | protein 60, integrase |
| 1720    | prophage | integration protein |
| 36      | prophage | protein 01, hypothetical regulator |
| 36      | prophage | protein 02 |
| 37      | prophage | protein 03, terminase subunit |
| 37      | prophage | protein 04 |
| 38      | prophage | protein 05, DNA primase |
| 40      | prophage | protein 06 |
| 40      | prophage | protein 07 |
| 41      | prophage | protein 08 |
| 41      | prophage | protein 09 |
| 41      | prophage | protein 10 |
| 42      | prophage | protein 11, transcriptional regulator |
| 42      | prophage | protein 12 |
| 42      | prophage | protein 13 |
| 43      | prophage | protein 14 |
| 44      | prophage | protein 15, transcriptional regulator |
| 44      | prophage | protein 16 |
| 45      | prophage | protein 17 |
| 45      | prophage | protein 18 |
| 45      | prophage | protein 19 |
| 46      | prophage | protein 20 |
| 47      | prophage | protein 21 |
| 48      | prophage | protein 22 |
| 49      | prophage | protein 23, integrase |
| 50      | prophage | protein 24, integrase |
| 50      | prophage | protein 25 |
| 50      | prophage | protein 26, transcriptional regulator |
| 50      | prophage | protein 27 |
| 50      | prophage | protein 28, excisionase |
| 50      | prophage | protein 29 |
| 51      | prophage | protein 30 |
| 51      | prophage | protein 31 |
| 51      | prophage | protein 32 |
| 52      | prophage | protein 33 |
| 52      | prophage | protein 34 |
| 52      | prophage | protein 35 |
| 52      | prophage | protein 36 |
| 53      | prophage | protein 37 |
| 53      | prophage | protein 38 |
| 54      | prophage | protein 39 |
| 54      | prophage | protein 40 |
| 54      | prophage | protein 41 |
| 54      | prophage | protein 42 |
| 54      | prophage | protein 43 |
| 54      | prophage | protein 44 |
| 55      | prophage | protein 45 |
| 55      | prophage | protein 46 |
| 56      | prophage | protein 47 |
| 56      | prophage | protein 48 |
| 57      | prophage | protein 49 |
| 58      | prophage | protein 50 |
| 59      | prophage | protein 51 |
| 60      | prophage | protein 52 |
| 60      | prophage | protein 53 |
| 60      | prophage | protein 54 |
| 61      | prophage | protein 55 |
| 62      | prophage | protein 56 |
| 63      | prophage | protein 57 |
| 64      | prophage | protein 58 |
| 65      | prophage | protein 59 |
| 66      | prophage | protein 60, integrase |

Transposon related functions

| Accession | Description |
|-----------|-------------|
| tra1077A  | 53 transposase of IS1077A |
| tra1077B  | 140 transposase of IS1077B |
| tra1077C  | 375 transposase of IS1077C |
| tra1077D  | 628 transposase of IS1077D |
| tra1077E  | 838 transposase of IS1077E |
| tra1077F  | 2156 transposase of IS1077F |
| tra1077G  | 2217 transposase of IS1077G |
| tra904A   | 54 transposase of IS904A |
| tra904B   | 138 transposase of IS904B |
| tra904C   | 40 transposase of IS904C |
| tra904D   | 374 transposase of IS904D |
| tra904E   | 627 transposase of IS904E |
| tra904F   | 836 transposase of IS904F |
| tra904G   | 839 transposase of IS904G |
| tra904H   | 2155 transposase of IS904H |
| tra904I   | 2215 transposase of IS904I |
| tra904J   | 1225 transposase of IS904J |
| tra904A   | 92 transposase of IS981A |
| tra904B   | 93 transposase of IS981B |
| tra904C   | 651 transposase of IS981C |
| tra904D   | 729 transposase of IS981D |
| tra904E   | 1217 transposase of IS981E |
| tra904F   | 1222 transposase of IS981F |
| tra904G   | 1276 transposase of IS981G |
| tra904H   | 1586 transposase of IS981H |
| tra904I   | 1748 transposase of IS981I |
| tra904J   | 2103 transposase of IS981J |
| tra904K   | 640 transposase of IS982 |
| tra903A   | 682 transposase of IS983A |
| tra903B   | 707 transposase of IS983B |
| tra903C   | 958 transposase of IS983C |
| tra903D   | 1338 transposase of IS983D |
| tra903E   | 1396 transposase of IS983E |
| tra903F   | 1556 transposase of IS983F |
| tra903G   | 1755 transposase of IS983G |
| tra903H   | 1954 transposase of IS983H |
| tra903I   | 1978 transposase of IS983I |
even distribution of three other IS elements is not caused by a particular property of the \textit{L. lactis} cell. We suggest that this distribution indicates a lateral transfer of a large portion of the genome from a lactococcus donor, carrying one type of IS, to a recipient, carrying the other type. Two lines of evidence lend support to this hypothesis. First, IS1076, which corresponds to the association of IS1077 and IS904 described above, is distributed over the whole genome of the strain \textit{L.lactis} ssp. \textit{cremoris} MG1363 (Le Bourgeois et al. 1995) rather than being restricted to one region of the genome. This transposon has, therefore, no particular hot region for insertion in the lactococcal genome. Second, the restriction map of another strain, \textit{L. lactis} ssp. \textit{lactis} DL11, coincides with that of IL1403 in the area between \textit{rrnF} (550 kb) and \textit{rrnE} (1980 kb), while it is divergent elsewhere (Le Bourgeois et al. 1992). We suggest that DL11 may be close to one of the putative parental strains of IL1403. Investigations of the distribution of IS1077 and IS983 among different lactococci might allow identification of both putative parents of the IL1403 strain.

Three potential prophages, designated \textit{pi}1, \textit{pi}2, and \textit{pi}3, were detected at positions near 460, 1050, and 1460 kb (Fig. 1). They are large (35–44 kb), encode 49–60 proteins, and are related to known temperate phages of \textit{L. lactis}. Another three prophages, designated \textit{ps}1, \textit{ps}2, and \textit{ps}3, are localized near 42, 509, and 2020 kb (Fig. 1). They are small (11–15 kb), encode only 16–23 proteins and might be satellites of the other phages, as they lack most of the genes that code for phage structural elements. A copy of IS983 is present in \textit{ps}3, which might, thus, be defective. The six prophages comprise a total of 175 kb of DNA and 221 protein coding genes. Recently, Chopin et al. (2001) characterized five phages, which can be found in the supernatant of IL1403 after mitomycin C treatment, and demonstrated the correspondence between the phage DNA extracted from the supernatant and the chromosome sequence. Phage \textit{bIL285} from the supernatant corresponds to \textit{pi}2, \textit{bIL286} to \textit{pi}3, \textit{bIL309} to \textit{pi}1, \textit{bIL310} to \textit{ps}1, and \textit{bIL312} to \textit{ps}2. \textit{ps}3, designated also as \textit{bIL311} (Chopin et al. 2001), cannot be induced, probably because of the IS983 element present in its genome. Detecting the circular forms of DNA of these phages allowed precise determination of the integration sites. About 9.2\% of the \textit{L. lactis} genome is thus

\begin{table}
\centering
\caption{(Continued)}
\begin{tabular}{lll}
\hline
\textbf{Gene} & \textbf{Start (kb)} & \textbf{Function} \\
\hline
\textit{tra983J} & 2012 & transposase of IS983J \\
\textit{tra983K} & 2017 & transposase of IS983K \\
\textit{tra983L} & 2084 & transposase of IS983L \\
\textit{tra983M} & 2148 & transposase of IS983M \\
\textit{tra983N} & 2203 & transposase of IS983N \\
\textit{tra983O} & 2268 & transposase of IS983O \\
yafG & 53 & hypothetical protein \\
yafI & 55 & hypothetical protein \\
yajE & 92 & transposase \\
yajG & 94 & transposase \\
ybdk & 138 & hypothetical protein \\
ybdL & 139 & hypothetical protein \\
ybeG & 141 & hypothetical protein \\
ydhD & 373 & hypothetical protein \\
ydhE & 375 & hypothetical protein \\
yfjB & 593 & transposon-related protein \\
ygcD & 628 & hypothetical protein \\
ygeE & 629 & hypothetical protein \\
yjff & 651 & transposase \\
yjhc & 729 & transposase \\
yidF & 837 & hypothetical protein \\
yidG & 838 & hypothetical protein \\
yidH & 839 & hypothetical protein \\
yimA & 1212 & integrase \\
yimB & 1217 & transposase \\
ymcD & 1222 & transposase \\
ymFD & 1259 & integrase-recombinase \\
ymbH & 1276 & transposase \\
ypl & 1587 & transposase \\
yrdA & 1748 & transposase \\
yui & 2104 & transposase \\
yvfC & 2157 & hypothetical protein \\
yvfD & 2156 & hypothetical protein \\
yvfE & 2216 & hypothetical protein \\
yvbc & 2217 & hypothetical protein \\
\textbf{Other} & & \\
\textit{crtK} & 574 & carotenoid biosynthetic protein CrtK \\
yebB & 412 & carotenoid biosynthetic protein \\
\hline
\end{tabular}
\end{table}
formed by IS elements and prophages, suggesting that they may be important for horizontal gene transfer in these bacteria.

**Paralogous Gene Families**

We define here as a paralogous protein family a group of proteins within which each protein shares at least one homologous domain with another protein of the group. By this criterion, there are 370 paralogous families, comprising 1189 gene products, in the *L. lactis* genome. Among the smaller families (<10 members) there are 208 of two members, 80 of three, 36 of four, 13 of five, 13 of six, 8 of seven, 4 of eight, and 2 of nine. The larger families contain 10, 11, 15, 18, 26, and 60 members, the last corresponding to ATP-binding proteins of ABC transporters, as is the case in many bacteria. In the four smallest families, distribution of the number of proteins resembles that of *B. subtilis* (Kunst et al. 1997). It decreases, very approximately, twofold when the family member count increases by one (568: 273:168:100 in *B. subtilis* and 416:240:144:65 in *L. lactis* for doublets, triplets, quadruplets, and quintuplets, respectively).

**Information Processing and Gene Regulation**

Information processing refers to the genes constituting replication, transcription, and translation machinery. In *L. lactis*, it is overall very similar to that of *B. subtilis*, the best characterized AT-rich gram-positive bacterium (Kunst et al. 1997). There are 67 genes involved in DNA metabolism in *L. lactis*. All the genes involved in DNA replication in *B. subtilis* are present in *L. lactis*, including counterparts of *dnaB*, *dnaD*, and *dnaI*, genes essential for initiation of replication in *B. subtilis* and absent in gram-negative bacteria. Two DNA-polymerase III α-chain genes, one corresponding to *polC* and another to *dnaE* of *B. subtilis*, were also detected in *L. lactis*. In contrast, *E. coli* has only the *dnaE* gene.

Transcription machinery in both *L. lactis* and *B. subtilis* comprises some 30 genes other than the α-factors. However, the number of α-factors differs greatly, as there are only three in *L. lactis*, while there are 18 in *B. subtilis*, pointing to a considerable difference in the mode of gene-expression regulation in the two organisms. Translation machinery comprises 119 genes in *L. lactis* and 131 genes in *B. subtilis*. There are no duplicated aminoacyl-tRNA synthetase genes in *L. lactis*, while there are three (for threonine, tyrosine, and histidine) in *B. subtilis*. Posttranslational protein modification genes mostly differ, as there are 27 such genes in *B. subtilis* and only 10 in *L. lactis*. A particular regulation of translation might also operate in *L. lactis*. As discussed more fully below, all the late competence genes of *L. lactis* seem to be controlled by a mechanism relaying on leaderless mRNAs and, thus, on a particular mode of translation. Recent evidence shows that the involvement of translation initiation factor 3, present in all bacteria, in start codon recognition is important for restriction of translation in such systems (Tedin et al. 1999). This provides a link between regulation of translation and competence in *L. lactis*. Such interaction has not been detected previously.

Analysis of homology allowed us to assign regulatory functions to 138 genes, half of which were classified further by their similarity to regulatory proteins of known families. The overall number of regulatory systems is about twofold lower in *L. lactis* than in *B. subtilis*, but the proportion of these genes is similar in the two organisms. Among the interesting differences is a much lower number of the two-component signal transducers in *L. lactis* than in *B. subtilis* (eight instead of 34) and of α-factors (three instead of 18), both of which regulate complex responses to changing environmental conditions.

**Energy Metabolism and Transporters**

The most important industrial applications of *L. lactis* are based on its energy metabolism, which leads mainly to the production of high amounts of lactic acid (homolactic fermentation). Anaerobic glycolysis is the principal energy-generating process in *L. lactis*, and very little of the fermented sugar (∼5%) is used for synthetic reactions (Poolman 1993). All the genes required for the conversion of the glucose to pyruvate are present in the genome. The pyruvate is converted into lactic acid, thus allowing the oxidation of reduced NAD and the lactate dehydrogenase gene *ldh*, essential for this process, was studied intensely (Griffin et al. 1992). Three other genes, highly similar to *ldh* (*ldhB*, *ldhX* and *hcd*) are present in the genome, but their role is not known. The product of the last gene has a high similarity (42% identity) to hydroxyisocaproate dehydrogenase and may, therefore, be involved in the metabolism of branched-chain amino acids. Lactate is transported into the growth medium, causing the efflux of protons and, thus, providing transmembrane potential indispensable for growth and energy recycling (Ten Brink et al. 1985).

Genome analysis indicates that the full citric acid cycle, gluconeogenesis enzymes, and many anaplerotic reactions do not exist in *L. lactis*. Unexpectedly, the functions necessary for aerobic respiration are encoded in the genome. *L. lactis* has *mer* and *cytABCD* operons, encoding proteins required for menaquinone synthesis and cytochrome d biogenesis. It also has three genes involved in the late steps of heme synthesis (*hemH*, *hemK*, and *hemN*, required for oxidation of porphyrinogen and attachment of iron to heme) but not the genes required for the early steps. *L. lactis* may thus be able to carry out oxidative phosphorylation if the topoporphyrinogen is provided. Indeed, improved growth properties in media containing hemin were
observed for certain Streptococci (Sijpesteijn 1970; Mickelson 1972). The genome analysis thus suggests the existence of aerobic respiration in this bacterium, generally considered an exclusively fermentative microorganism.

Use of L. lactis in the food industry also exploits its ability to form fermentation products other than lactate (mixed acid fermentation). The balance of products depends on activities of enzymes that act on the key metabolite generated by glycolysis, the pyruvate. A number of genes encoding such enzymes (pyruvate dehydrogenase, pdhABC; α-acetolactate synthase, als; pyruvate-formate lyase, pf; and lactate dehydrogenase, ldh) have been identified previously in L. lactis and confirmed by genome analysis. We detected a novel gene, posL, encoding pyruvate oxidase, which also acts on pyruvate and might, therefore, play a role in switching between different fermentation modes.

Besides gene activity, the availability of cofactors, such as NADH and FAD, also affects the balance of different fermentation products. Artificial changing of NADH/NAD ratio in L. lactis can redirect carbon flow from lactic acid to acetoin and diacetyl (Lopez de Felipe et al. 1998). There are more than five NADH dehydrogenase genes in the L. lactis genome, which may affect the type of fermentation products. Some NADH dehydrogenases generate hydrogen peroxide, which is toxic for the cells. L. lactis has no gene encoding catalase, which can remove the toxic H₂O₂. However, there is a gene encoding thiol peroxidase (tpx) and two genes (ahpC and ahpF) encoding alkyl hydroperoxide reductases. These proteins could possibly act on H₂O₂. Active sodA, encoding superoxide dismutase, which converts oxygen radicals to H₂O₂, was shown to be important for the oxidative stress response (Sanders et al. 1995). Also, the gshR gene encoding glutathion reductase may be involved in response of L. lactis to the aerobic growth conditions.

The heterofermentative metabolism takes place in L. lactis when pentose-phosphate pathway is active, as in this case, glycolysis generates not only a three-carbon compound that can be converted to lactate but also a two-carbon compound. We detected glucose-6P dehydrogenase (zwf), phosphogluconate dehydrogenase (gdh), and ribuloso-5P epimerase (rpe), which can lead to the formation of xyluloso-5P. Phosphoketolase, encoded by ptk gene, can catalyze formation of glycer- aldehyde-3P and acetyl-P, which enters the fermentation pathways that yield lactate and ethanol, respectively.

Understanding the molecular basis of the switch between different fermentation types is of interest not only for standard uses of L. lactis but also for the metabolic engineering in this organism, aiming to enhance synthesis of certain metabolites to industrially useful levels. We detected a correlation between the presence of the phosphoenolpyruvate dependent transport system (PTS) and the fermentation profile for a given carbon source. PTS systems for fructose, mannose, sucrose or trehalose, mannitol, and cellobiose are present in the genome, and the homolactic fermentation profiles were reported for growth on fructose, mannose, glucose (which uses mannose or mannitol PTS) and sucrose (Cocaing-Bousquet et al. 1996). In contrast, mixed acid or heterofermentation profiles were observed for growth on galactose, xylose, maltose, glucose, ribose, and lactose, which are not imported by a PTS system. When L. lactis cells harbor a plasmid encoding lactose-specific PTS system, lactose fermentation becomes homolactic (Gasson 1983). Our genome analysis thus strengthens the proposal that sugar consumption rate, which is the highest when PTS system is available, determines the ability for efficient homolactic fermentation (Cocaing-Bousquet et al. 1996). The correlation of information derived from genome analysis with experimental data on fermentation product distribution indicates that critical parameters regulating the final product balance may be found by a thorough analysis of the carbon source use and transport systems.

Proteases and Amino Acid Catabolism Genes
Proteases and peptidases provide a selective advantage for bacteria growing in milk, as this medium is rich in caseins and relatively poor in free amino acids. Amino acid catabolism has an impact on fermentation regulation and on the flavor of dairy products.

Genome sequence revealed 19 protease-encoding genes (Table 1). These include the membrane protease HtrA, which is responsible for degradation of the precursors of foreign exported proteins (Pouquet et al. 2000). Some 16 peptidases from LAB were characterized previously, including the products of 13 genes detected in L. lactis (Christensen et al. 1999).

Catabolism of amino acids usually starts by deamination. Arginine catabolic genes, organized in an operon near 2110 kb, encode the enzymes for the deamination pathway as well as the arginine tRNA synthetase, suggesting complex regulation. Another operon for arginine catabolism, near 1755 kb, contains genes arcC3 and otcA. It could have a regulatory function, as it also contains the genes ilvH and yrfE, representing a signal transduction system of a new type. Aspartate aminotransferase (aspC) and asparaginase (ansB) are involved in aspartate and asparagine catabolism. No genes for aspartate decarboxylase or aspartase were detected, although such enzymatic activities were identified in Lactobacillus, another prominent group of LAB (Rollan et al. 1985). Recent studies on catabolism and biosynthesis of glutamate in L. lactis identified the existence of a pathway leading to the production of γ-aminobutyrate (GABA; Sanders et al. 1998). We identified
gadRCB operon for GABA production, gltBD genes for glutamate synthase, and an operon involved in citric acid metabolism: pyCA, gltA, citB, and lcl. Under appropriate physiological conditions, products of some of these genes might carry out glutamate catabolism, rather than biosynthesis. Serine can be directly converted to pyruvate by serine dehydratase encoded by the sdaAB operon.

Genome sequence provides inventory of 12 aminotransferases, of which some can initiate degradation of aromatic, branched-chain, and sulfur-containing amino acids, important for cheese flavor. The specificity of seven aminotransferases (aspC, serC, argD, glmS, hisC, aspB, and arcT) can be predicted from sequence comparisons, whereas those of other five (araT, nifZ, yeiG, bcaT, and ytfE) are less obvious. It was recently shown that araT and bcaT are involved in the degradation of aromatic and branched-chain amino acids, respectively (Yvon et al. 2000). The product of ytfE might be specific for methionine, as the gene is cotranscribed with the relevant biosynthesis genes. Degradation of tryptophane seems to proceed via indole aldehyde because of indole pyruvate decarboxylase gene ipd. It is not clear which pathways L. lactis uses to catabolize phenylalanine and tyrosine. It is possible that phenyl pyruvate and p-OH-phenyl pyruvate are degraded further by decarboxylation. This would depend on the specificity of the phenolic acid decarboxylase encoded by pdc.

Amino Acid, Vitamin, and Nucleotide Biosynthesis
L. lactis requires certain metabolites in the growth medium, although it has a genetic potential to synthesize some of them. Synthetic medium for L. lactis should contain at least six amino acids (isoleucine, valine, leucine, histidine, methionine, and glutamic acid) and seven vitamins (biotin, pyridoxal, folic acid, riboflavin, nicotinamide, thiamine, and pantothenic acid; Jensen and Hammer 1993). L. lactis has the genes to synthesize the 20 standard amino acids and at least four co-factors (folic acid, menaquinone, riboflavin, and thioetherin). One reason for the requirement of the compounds that can potentially be synthesized is that some of the existing genes are not functional, as was reported previously for amino acid biosynthesis genes (Godon et al. 1993). We carefully checked sequencing tracks for the genes that could contain a frameshift mutation and could not rule out the presence of a mutation in 30 of them. This relatively high level of pseudogenes in IL1403 could possibly be, at least in part, caused by the treatments used to care the parental strain of its plasmids (Chopin et al. 1984).

Milk does not contain sufficient levels of purine compounds to support growth of L. lactis and, therefore, de novo biosynthesis is necessary (Dickely et al. 1995). We detected 57 genes involved in this metabolism. Therefore, physiological and genomic evidence shows that L. lactis has sufficient and fairly active capacities for biosynthesis and also for salvage of nucleic acid compounds.

Cell Wall Metabolism
Many L. lactis properties that are important for applications, such as phage sensitivity, stress resistance, autolysis, and mucosal immunostimulation, depend on the structure of the cell wall. There are 29 genes encoding enzymes required for the synthesis of the main cell wall component, peptidoglycan. Among these, three encode amino acid racemases: dal for alanine, murK for glutamate, and racD for aspartate. D-alanine and D-glutamate are the components of linear peptide moieties of peptidoglycan, whereas D-aspartate forms cross-bridges. There are no genes for synthesis of modified peptidoglycan, containing D-lactate or D-serine instead of D-alanine, reported for several other LAB.

Cheese ripening can be accelerated by induction of enzymes that process peptidoglycan. There are six genes related to such processing in L. lactis: dacA and dacB, encoding alanine–alanine carboxypeptidase; and acmA, B, C, and D, encoding four lysozymes. Carboxypeptidases alone cannot cause the cell lysis, as their activity does not destabilize the wall. Modulation of the level of their production can, however, influence the action of lysozymes. acmA, responsible for separation of daughter cells, was used for artificial induction of autolysis (Buist et al. 1997).

Lipoteichoic acid is another main component of the L. lactis cell wall. Neither teichoic nor teichuronic acids were detected in this microorganism (Valyasevi et al. 1990). However, there is a cluster of seven tag genes near 950 kb. Only three genes from teichuronic acid biosynthesis pathway were found: ycbK, ycbF, and ycbH, corresponding to tuaB, tuaC, and tuaG of B. subtilis. dlt operon, encoding D-alanylation of lipoteichoic acid, is of crucial importance for properties of the cell wall and whole-cell physiology. A knockout mutation in dltD causes filamentous growth and UV sensitivity and facilitates penetrability of the cells (Duwat et al. 1997).

Synthesis of extracellular polysaccharides is important for the industrial use of many LAB, as these polymers affect the texture of the fermented products. There are >20 genes involved in the biosynthesis of such molecules in the region near 200 kb. They encode functions providing activated sugars and other components involved in production of surface or extracellular polysaccharide. A plasmid that carries an operon involved in the formation of the repeating unit, linking activated sugar to the lipid carrier, export, and polymerization, was recently identified (Van Kranenburg et al. 1997). Conjunction of plasmid-carried and chromo-
somal functions presumably determine the amounts and the structure of extracellular polysaccharides.

Protein Secretion

*L. lactis* has only eight genes identified as implicated in protein secretion. Contrary to *B. subtilis* and *E. coli*, this bacterium does not have *secDF* genes, known to improve the secretion efficiency (Pogliano and Beckwith 1994; Bolhuis et al. 1998). There is only one membrane protease, HtrA, involved in degradation of hybrid exported proteins (Pouquet et al. 2000). Gene *pmpA* (protein maturation protein) encodes a homolog of PrsA from *B. subtilis* and might be involved in stabilization of secreted proteins by facilitating their folding. *L. lactis* was shown to secrete up to 20 mg/L of foreign protein with optimized gene constructs (Le Loir et al. 1998). This value could possibly be improved by manipulating the gene expression levels and supplying the missing components of the secretion machinery.

Competence to Genetic Transformation

Natural competence to DNA transformation was not demonstrated in *L. lactis*. We detected four operons (*comE, comF, comC, and comG*) containing genes similar to the late competence genes from *B. subtilis* and *S. pneumoniae*. In addition, we found a gene for ComX, which is similar to the *S. pneumoniae* ECF-type α-factor required for transcription of the competence genes (Lee and Morrison 1999). The regions preceding the first ORF of the four operons resemble competence promoters from *S. pneumoniae* and might be transcribed by ComX. There are three common sequences in front of all competence operons, two of which, GGTACATT and TTTTCGTATA, are in the −35 and −10 domains of the promoter, while the third, AGTATG, includes the ATG start codon of the first gene in each operon. The relative position of the three conserved elements indicates that all mRNAs start at the ATG codon of the first gene and are, therefore, leaderless, lacking the canonical ribosome-binding site. Search for the consensus sequence over the whole genome, using PatScan (Dsouza et al. 1997), revealed six such promoters other than those of the late competence operons. The genes downstream of these promoters are *radA, colA, dprA, recQ, ssbA*, and *yqfG*. Only the *radA* gene, encoding a DNA repair protein, has leaderless mRNA. Three of the genes, *colA, dprA, and recQ*, affect DNA transformation in *S. pneumoniae, H. influenzae*, and *B. subtilis*, respectively (Karudaparam et al. 1995; Fernandes et al. 1998; Pestova and Morrison 1998). *ssbA* encodes single-strand DNA-binding protein and could be involved in the processing of transforming DNA, which enter gram-positive bacteria in the single-stranded form. *yqfG* encodes a protein of unknown function. The existence of the competence-related genes in *L. lactis* indicates that this bacterium might be naturally transformable by DNA. There are no genes homologous to those involved in early steps of competence development in *S. pneumoniae*, which indicates that, in *L. lactis*, the regulation cascade upstream of ComX α-factor is very different from that in *Streptococci*.

Another difference between *L. lactis* and *S. pneumoniae* competence systems is that the leaderless mRNAs are present in the former organism only. The translation of such mRNAs requires that they start precisely at the initiation codon of the gene (Kravchenko et al. 1988; Van Etten and Janssen 1998). Synthesis of competence-related proteins would, therefore, not take place on spurious transcription of the cognate genes by leakage from upstream operons. This might tighten the control of the competence development and does limit it to very strict environmental conditions.

Horizontal Gene Transfer between Lactococci and Gram-Negative Enteric Bacteria

We detected a gene of unknown function, designated *ycdB*, which appears to be present in all bacteria and some eukaryotes. The level of identity between the YcdB protein and a homolog from *S. pyogenes* or *S. pneumoniae*, phylogenetically close to *L. lactis*, is ~80%, while the identity with the homologous genes from gram-negative bacteria is ~40%. Very surprisingly, the *E. coli* and *S. typhimurium* genomes encode not only a protein that is 40% identical with YcdB but also a protein that is 94% identical to YcdB. We conclude that this second *ycdB* gene has been transferred from lactococci to enteric bacteria. The divergence of the synonymous nucleotide sites in *L. lactis* IL1403, compared with *Salmonella* and *E. coli*, is ~10%. If the rate of nucleotide changes at such sites is ~1% per million years (Ochman et al. 1999), the genes in *Salmonella/E. coli* and *L. lactis* IL1403 started to diverge 10 million years ago. However, comparison of the *ycdB* genes in different strains of lactococci and in gram-negative enteric bacteria may reveal even more closely related genes and allow us to better assess the time of the gene transfer, the species that may have been involved in the transfer, and the mechanism of the transfer. Nevertheless, anticipating that closer homologs will be found, it is tempting to speculate that the transfer may have taken place in the digestive tract of ruminants, if it involved wild-type lactococci, or of humans, if it involved the domesticated lactococci, massively introduced there by cheese consumption.

Analysis of completely sequenced genomes, available from the NCBI server, revealed that most bacteria have only one homolog to YcdB. Some (*E. coli, S. typhimurium, B. subtilis, E. faecalis*, and *Shewanella putrefaciens*), however, have two, indicating that the family might be undergoing an expansion where, at least for enteric bacteria, a lateral gene transfer from lactococci might be a driving force. As the function of this gene is
unknown, the advantage that the second copy confers is not known. Elucidation of the gene function would help to answer this question.

METHODS

Genome Cloning, Sequencing, and Data Verification

The strain IL1403 is a plasmid-free derivative of the strain IL594, isolated from a cheese starter culture (Chopin et al. 1984). Diagnostic sequencing, involving 10,235 sequencing reactions and yielding a total of 4,687,630 bases, has been described previously (Bolotin et al. 1999). Further sequencing was carried out to assure that each nucleotide in the genome was read at least four times and at least once on each strand. For this purpose, a collection of short insert clones was constructed. A total of 9,888,620 bases, covering 93% of the total genome, were produced by 15,578 more sequencing reactions. To reduce the error rate level to <0.01%, 978 more reactions, with average read length of 632 bases, were carried out using genome-specific primers. The redundancy of the final assembling is 6.44.

Informatics and Gene Nomenclature

Assembling manual corrections of sequencing errors and consensus generation were carried out concurrently with data accumulation, using the XBASE program (Dear and Staden 1991; version 14.0). To predict protein-coding regions, we used a conceptual translation of the whole genome in six possible coding frames. The predicted proteins >60 amino acids were checked for the statistical consistency with the output of the GENMARK program (Borodovsky and McIninch 1993) using parameters for Streptococcal genes. EBI server (http://www2.ebi.ac.uk/genemark) and pyogenes_3.xdr matrix dated November 14, 1996, were used for this analysis. The presence of a putative ribosome-binding site upstream of the 3′ end of the candidate was searched next. As a ribosome binding site, we considered the presence of initiator codon ATG, TTG, or GTG and a short sequence homologous to the 3′ end of 16S rRNA of L. lactis (5′-GGAUCACCCUCUUCUAU3′) upstream of it (Chiaruttini and Milet 1993). Genome notation was done by using several homemade shell or Perl scripts, generating convenient html format tables linked to BLAST (Altschul et al. 1990) output files. NCBI server (http://www.ncbi.nlm.nih.gov/Entrez) was used to generate updated bacterial protein databases. Homology analysis of YcdB with the unpublished genome sequences was carried out by using the relevant NCBI server (http://www.ncbi.nlm.nih.gov/Microb_blast/unfinishedgenome.html). The functional classification of genes was done according to the list of categories presented earlier (Bolotin et al. 1999). Fully automatic computer-generated classification was used as the starting material. Each protein was then analyzed by an expert to improve the category assignment, which is presented in Table 1 and Figure 2. The expert usually used three means to confirm or to alter the automated function assignment and classification: first, phylogenetic or COGnitor (Tatusov et al. 1997) assisted scrutiny of BLAST or FASTA results (performed with different parameters); second, complete knowledge of particular biochemical pathways or biological systems, existing in other than L. lactis IL1403 organisms (such as protein secretion or the competence system). Phage-specific proteins were classified to those because of their clustering in the areas identified as prophages. Also, specialized databases (Quentin et al. 1999) were used by the expert to classify the ABC transporters; third, results of numerous experiments in L. lactis, published previously (148 functional assignments). Although it is never absolutely explicit, the provided classification of gene functions in L. lactis IL1403 is biological, rather than biochemical. L. lactis paralogous gene families were constructed by searching each predicted protein against all predicted proteins, using BLASTP with different parameters. Alignments of proteins in the identified families were then scrutinized to make a decision of how many proteins belong to a family. This decision was based either on the size of homologous domains or on the similarity levels. A protein was always assigned to only one family of paralogs.

We tried to keep the same gene symbols as proposed by the previous authors for ORFs with functions experimentally confirmed in L. lactis (148 genes). A y prefix with the gene symbol consistent with its position on the chromosome (Fig. 2) was kept for unascertained functions (1149 genes). Other gene symbols, consistent with those for homologs found in other bacteria, are proposed here (1017 genes).

Accessibility of Data

The nucleotide sequence of the L. lactis IL1403 genome is available from NCBI with accession no. AE005176. Updated annotations are supported at the Genétique Microbienne (INRA) server at http://spock.jouy.inra.fr. A PatScan of Ross Overbeek (Douza et al. 1997) for pattern searches in DNA sequence and proteins, implemented for IL1403, and peptide spectrum identification tool PeptOko for L. lactis proteome research are also available from this server.

ACKNOWLEDGMENTS

We thank Jacek Bordovski and Saulius Kulakauskas for giving samples of the L. lactis IL1403 strain and Marie-Christine and Alain Chopin, Patrick Duwat, Emmanuel Janet, Alexandra Gruss, Emmanuelle Maguin, Isabelle Poquet, Pierre Renaut, and Catherine Robert for helpful discussions. We thank also the Genome Centers that contributed to the Unfinished Microbial Genome Database available for BLAST search through the NCBI server (http://www.ncbi.nlm.nih.gov/Microb_blast/unfinishedgenome.html).

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 USC section 1734 solely to indicate this fact.

REFERENCES

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. 1990. Basic local alignment search tool. J. Mol. Biol. 215: 403–410.

Bolotin, A., Brochhuizen, C.P., Sorokin, A., van Roosmalen, M.L., Venema, G., Bron, S., Qax, W.J., and van Dijl, J.M. 1998. SecDF of Bacillus subtilis, a molecular Siamese twin required for the efficient secretion of proteins. J. Biol. Chem. 273: 21217–21224.

Bolotin, A., Mauger, S., Malarme, K., Ehrlich, S.D., and Sorokin, A. 1999. Efficiency of secretion of proteins. J. Biol. Chem. 273: 21217–21224.
Bolotin et al.

1999. Low-redundancy sequencing of the entire Lactococcus lactis IL1403 genome. Antonie Leeuwenhoek 76: 27–76.

Borodovsky, M. and McIninch, J. 1993. GENMARK: A parallel gene recognition for both DNA strands. Comput. Chem. 17: 123–133.

Buist, G., Karsens, H., Nauta, A., van Sinderen, D., Venema, G., and Kok, J. 1997. Autolysis of Lactococcus lactis caused by induced overproduction of its major autolysin, Acm. Appl. Environ. Microbiol. 63: 2722–2728.

Chianuttini, C. and Milet, M. 1993. Gene organization, primary structure and RNA processing analysis of a ribosomal RNA operon in Lactococcus lactis. J. Mol. Biol. 230: 57–76.

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

Dickely, F., Nilsson, D., Hansen, E.B., and Johansen, E. 1995. Dear, S. and Staden, R. 1991. A sequence assembly and editing program for efficient management of large projects. Nucleic Acids Res. 19: 3907–3911.

Dickely, F., Nilsson, D., Hansen, E.B., and Johansen, E. 1995. Isolation of Lactococcus lactis nonsense suppressors and construction of a food-grade cloning vector. Mol. Microbiol. 15: 839–847.

Douro, M., Larsen, N., and Overbeek, R. 1997. Searching for patterns in genomic data. Trends Genet. 13: 497–498.

Duwat, P., Cochu, A., Ehrlich, S.D., and Gruss, A. 1997. Characterization of Lactococcus lactis UV-sensitive mutants obtained by ISS1 transposition. J. Bacteriol. 179: 4473–4479.

Fernandez, S., Sorokin, A., and Alonso, J.C. 1998. Genetic recombination in Bacillus subtilis 168: Effects of recU and recS mutations on DNA repair and homologous recombination. Cell 94: 451–462.

Dear, S. and Saden, R. 1991. A sequence assembly and editing program for efficient management of large projects. Nucleic Acids Res. 19: 3907–3911.

Dickely, F., Nilsson, D., Hansen, E.B., and Johansen, E. 1995. Isolation of Lactococcus lactis nonsense suppressors and construction of a food-grade cloning vector. Mol. Microbiol. 15: 839–847.

Douro, M., Larsen, N., and Overbeek, R. 1997. Searching for patterns in genomic data. Trends Genet. 13: 497–498.

Duwat, P., Cochu, A., Ehrlich, S.D., and Gruss, A. 1997. Characterization of Lactococcus lactis UV-sensitive mutants obtained by ISS1 transposition. J. Bacteriol. 179: 4473–4479.

Fernandez, S., Sorokin, A., and Alonso, J.C. 1998. Genetic recombination in Bacillus subtilis 168: Effects of recU and recS mutations on DNA repair and homologous recombination. Cell 94: 451–462.

Dear, S. and Saden, R. 1991. A sequence assembly and editing program for efficient management of large projects. Nucleic Acids Res. 19: 3907–3911.

Douro, M., Larsen, N., and Overbeek, R. 1997. Searching for patterns in genomic data. Trends Genet. 13: 497–498.

Duwat, P., Cochu, A., Ehrlich, S.D., and Gruss, A. 1997. Characterization of Lactococcus lactis UV-sensitive mutants obtained by ISS1 transposition. J. Bacteriol. 179: 4473–4479.

Fernandez, S., Sorokin, A., and Alonso, J.C. 1998. Genetic recombination in Bacillus subtilis 168: Effects of recU and recS mutations on DNA repair and homologous recombination. Cell 94: 451–462.

Dear, S. and Saden, R. 1991. A sequence assembly and editing program for efficient management of large projects. Nucleic Acids Res. 19: 3907–3911.

Dickely, F., Nilsson, D., Hansen, E.B., and Johansen, E. 1995. Isolation of Lactococcus lactis nonsense suppressors and construction of a food-grade cloning vector. Mol. Microbiol. 15: 839–847.

Douro, M., Larsen, N., and Overbeek, R. 1997. Searching for patterns in genomic data. Trends Genet. 13: 497–498.

Duwat, P., Cochu, A., Ehrlich, S.D., and Gruss, A. 1997. Characterization of Lactococcus lactis UV-sensitive mutants obtained by ISS1 transposition. J. Bacteriol. 179: 4473–4479.

Fernandez, S., Sorokin, A., and Alonso, J.C. 1998. Genetic recombination in Bacillus subtilis 168: Effects of recU and recS mutations on DNA repair and homologous recombination. Cell 94: 451–462.

Dear, S. and Saden, R. 1991. A sequence assembly and editing program for efficient management of large projects. Nucleic Acids Res. 19: 3907–3911.

Dickely, F., Nilsson, D., Hansen, E.B., and Johansen, E. 1995. Isolation of Lactococcus lactis nonsense suppressors and construction of a food-grade cloning vector. Mol. Microbiol. 15: 839–847.

Douro, M., Larsen, N., and Overbeek, R. 1997. Searching for patterns in genomic data. Trends Genet. 13: 497–498.
W., Fiers, W., and Remaut, E. 2000. Treatment of murine colitis by Lactococcus lactis secreting interleukin-10. Science 289: 1352–1355.
Tatusov, R.L., Koonin, E.V., and Lipman, D.J. 1997. A genomic perspective on protein families. Science 278: 631–637.
Tedin, K., Moll, I., Grill, S., Resch, A., Graschopf, A., Gualerzi, C.O., and Blasi, U. 1999. Translation initiation factor 3 antagonizes authentic start codon selection on leaderless mRNAs. Mol. Microbiol. 31: 67–77.
Ten Brink, B., Otto, R., Hansen, U.P., and Konings, W.L. 1985. Energy recycling by lactate flux in growing and non-growing cells of Streptococcus cremoris. J. Bacteriol. 162: 383–390.
Tettelin, H., Saunders, N.J., Heidelberg, J., Jeffries, A.C., Nelson, K.E., Eisen, J.A., Ketchum, K.A., Hood, D.W., Peden, J.F., Dodson, R.J., et al. 2000. Complete genome sequence of Natrierea meningitidis serogroup B strain MC58. Science 287: 1809–1815.
Valyasevi, R., Sandine, W.E., and Geller, B.L. 1990. The bacteriophage kh receptor of Lactococcus lactis subsp. cremoris KH is the rhamnose of the extracellular wall polysaccharide. Appl. Environ. Microbiol. 56: 1882–1889.
Van Etten, W.J., and Janssen, G.R. 1998. An AUG initiation codon, not codon-anticodon complementarity, is required for the translation of unleadered mRNA in Escherichia coli. Mol. Microbiol. 27: 987–1001.
Van Kranenburg, R., Marugg, J.D., van Swam, I.I., Willem, N.J., and de Vos, W.M. 1997. Molecular characterization of the plasmid-encoded eps gene cluster essential for exopolysaccharide biosynthesis in Lactococcus lactis. Mol. Microbiol. 24: 387–397.
Yvon, M., Chambellon, E., Bolotin, A., and Roudot-Algaron, F. 2000. Characterisation and role of the branched-chain aminotransferase (BcaT) isolated from Lactococcus lactis subsp. cremoris NCDO763 Appl. Environ. Microbiol. 66: 571–577.

Received October 31, 2000; accepted in revised form February 5, 2001.