

## Evolutionary Genomics of Lactic Acid Bacteria<sup>∇</sup>

Kira S. Makarova\* and Eugene V. Koonin

National Center for Biotechnology Information, National Library of Medicine,  
National Institutes of Health, Bethesda, Maryland 20894

The lactic acid bacteria (LAB) might be the most numerous group of bacteria linked to humans. They are naturally associated with mucosal surfaces, particularly the gastrointestinal tract, and are also indigenous to food-related habitats, including plant (fruits, vegetables, and cereal grains), wine, milk, and meat environments (60, 61). The LAB include both important pathogens, e.g., several *Streptococcus* species, and extremely valuable nonpathogenic species that are used for industrial fermentation of dairy products, meats, and vegetables, and they are also critical for the production of wine, coffee, silage, cocoa, and sourdough (13, 60, 61). In addition, the LAB are a priceless source of antimicrobial agents, the bacteriocins (reference 10 and references therein).

The term LAB mainly refers to the defining feature of the basal metabolism of these bacteria, the fermentation of hexose sugars yielding, primarily, lactic acid. Various aspects of LAB biology and application are thoroughly covered in several books by Wood and Holzapfel and Wood and Warner (60, 61) and numerous reviews, including those in a recent specialized issue of *FEMS Microbiology Reviews* (12, 14, 18, 20, 23, 24, 31, 39, 42, 49, 58). The definition of LAB is biological rather than taxonomical, i.e., the LAB do not comprise a monophyletic group of bacteria. Most of the LAB belong to the order *Lactobacillales*, a group of nonsporulating, gram-positive bacteria, but a few LAB species belong to the *Actinobacteria* (60).

The early sequencing of LAB genomes involved mostly bacteria of the genus *Streptococcus*, which encompasses most of the pathogenic LABs (50, 60). Currently, 19 complete genomes of streptococci are available, covering different strains of five species. A program aimed at extensive sequencing of the genomes of nonpathogenic LAB was announced in 2002 by the Lactic Acid Bacteria Genome Sequencing Consortium (19), but the actual breakthrough occurred only in the last 2 years (2005 and 2006). At the time of writing (August 2006), 18 complete genome sequences of the nonpathogenic LAB representing 14 species from the order *Lactobacillales* were available (Table 1). The *Lactobacillales* have relatively small genomes for nonobligatory bacterial parasites or symbionts (characteristic genome size, ~2 megabases, with ~2,000 genes), with the number of genes in different species spanning the range from ~1,600 to ~3,000. This variation in the number of genes suggests that the evolution of LAB involved active processes of gene loss, duplication, and acquisition. The current collection of LAB genomes is a unique data set that includes multiple related genomes with a gradient of diver-

gence in sequences and genome organizations. This set of related genomes is amenable to detailed reconstruction of genome evolution, which is not yet attainable with other groups of bacteria.

This review is largely based on recent work on comparative genomic analysis of 12 nonpathogenic LAB from the order *Lactobacillales* (Table 1) (28). We briefly discuss the customized comparative-genomic framework that was developed for the LAB and its application to evolutionary reconstruction and functional genomics.

### CLUSTERS OF ORTHOLOGOUS GENES OF *LACTOBACILLALES*: A FRAMEWORK FOR COMPARATIVE GENOMICS AND A TOOL FOR GENOME ANNOTATION

Robust identification of sets of orthologs (genes derived from the same gene in the last common ancestor of the compared species) is a prerequisite for informative evolutionary-genomic analysis of any group of organisms (25). Construction of orthologous gene clusters for a compact taxon, such as the *Lactobacillales*, results in much finer granularity than is attainable for broader groups of organisms (e.g., all bacteria), with a greater fraction of clusters containing a single member from all or most of the analyzed genomes. *Lactobacillales*-specific clusters of orthologous protein coding genes (LaCOGs) in 12 sequenced *Lactobacillales* genomes were built using the computational procedures that were previously employed for the construction of clusters of orthologous groups of proteins (COGs) from the sequenced prokaryotic and eukaryotic genomes (52, 53). As with the original COG database, manual curation was undertaken to refine the results of the automatic procedure, including tree reconstruction for LaCOGs that had multiple representatives for several genomes, searching for potential missing LaCOG members in untranslated regions of the LAB genomes, merging two or more LaCOGs when they appeared to have evolved from the same ancestral gene, using genomic context for LaCOG validation and refinement, and several additional analyses (37, 52).

Figure 1 shows the projection of the LaCOGs on the set of COGs that includes 63 prokaryotic genomes (52). The finer granularity of ortholog identification in the LaCOGs is apparent from the fact that 42% (1,359) of the LaCOGs correspond to 390 COGs. Not unexpectedly, many LaCOGs in this group represent duplications of widely conserved bacterial genes in the common ancestor of *Bacilli* or at a later stage of evolution. For example, there are three LaCOGs corresponding to the single COG that includes 23S RNA-specific pseudouridylylase synthases from other bacteria (COG0564) and two LaCOGs for SpoU-like rRNA methylases (COG0566), TrmA-like tRNA (uracil-5-)-methyltransferases (COG2265), SrmB-like

\* Corresponding author. Mailing address: National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD 20894. Phone: (301) 435-5912. Fax: (301) 435-7794. E-mail: makarova@ncbi.nlm.nih.gov.

<sup>∇</sup> Published ahead of print on 3 November 2006.

TABLE 1. General features of the sequenced genomes of *Lactobacillales*

Species/strain	Genome length (bp)	No. of plasmids (no. of genes)	No. of proteins	Reference	Association
<i>Lactobacillus plantarum</i> WCFS1 <sup>a</sup>	3,308,274	3 (50)	3,009	22	Plant/meat material
<i>Lactobacillus johnsonii</i> NCC 533 <sup>a</sup>	1,992,676	0	1,821	43	GIT <sup>b</sup>
<i>Lactobacillus acidophilus</i> NCFM	1,993,564	0	1,864	2	GIT
<i>Lactobacillus sakei</i> subsp. <i>sakei</i> 23K	1,884,661	0	1,879	8	Meat material
<i>Lactobacillus salivarius</i> subsp. <i>salivarius</i> UCC118	1,827,111	3 (320)	1,717	9	GIT
<i>Lactobacillus delbrueckii</i> subsp. <i>bulgaricus</i> ATCC 11842	1,864,998	0	1,562	57	Yogurt, cheese
<i>Lactobacillus delbrueckii</i> subsp. <i>bulgaricus</i> <sup>a</sup>	1,856,951	0	1,725	28	
<i>Lactobacillus gasserii</i> <sup>a</sup>	1,894,360	0	1,763	28	GIT
<i>Lactobacillus brevis</i> <sup>a</sup>	2,340,228	2 (37)	2,221	28	Plant/meat material
<i>Pediococcus pentosaceus</i> <sup>a</sup>	1,832,387	0	1,757	28	Plant/meat material
<i>Oenococcus oeni</i> <sup>a</sup>	1,780,517	0	1,701	28	Wine
<i>Leuconostoc mesenteroides</i> <sup>a</sup>	2,075,763	1 (34)	2,009	28	Plant material
<i>Lactobacillus casei</i> <sup>a</sup>	2,924,325	1 (20)	2,776	28	Dairy products
<i>Lactococcus lactis</i> subsp. <i>lactis</i> <sup>a</sup>	2,365,589	0	2,321	7	Dairy products, bacteriocins
<i>Lactococcus lactis</i> subsp. <i>cremoris</i> <sup>a</sup>	2,641,635	5 (129)	2,509	28	
<i>Streptococcus thermophilus</i> CNRZ1066	1,796,226	0	1,915	5	Yogurt, cheese
<i>Streptococcus thermophilus</i> LMG 18311	1,796,846	0	1,889	5	
<i>Streptococcus thermophilus</i> LMD-9 <sup>a</sup>	1,864,178	2 (6)	1,718	28	

<sup>a</sup> Species that were analyzed as part of a recent comparative-genomic project (28).

<sup>b</sup> GIT, gastrointestinal tract.

RNA helicases (COG0513), FtsW-like bacterial cell division membrane protein (COG0772), and many other COGs. Several paralogous genes (i.e., genes that evolved by duplication) in this group appear to be characteristic of *Lactobacillales* and deserve a more detailed discussion. One notable case involves two paralogous enolases (two LaCOGs), a nearly ubiquitous glycolytic enzyme that is present in a single copy in other bacteria. Phylogenetic analysis showed that one of the enolases in *Lactobacillales* is the ancestral version of that in gram-positive bacteria, whereas the other had apparently been acquired by the ancestor of the *Lactobacillales* from a different bacterial lineage, most likely *Actinobacteria* (28). Thus, the enolases of LAB actually appear to be pseudoparalogs, i.e., genes that evolved via acquisition of one of the related genes by horizontal transfer from a distant lineage rather than via duplication within the given lineage (25). It has been shown that both enolases of *Lactococcus lactis* subsp. *lactis* have enzymatic activity (15); however, their specific physiological functions remain unknown.

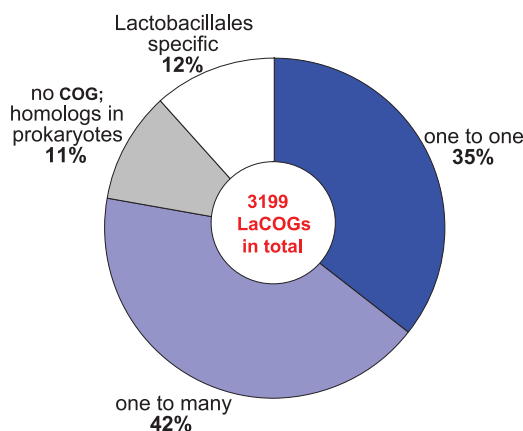


FIG. 1. Projection of LaCOGs onto COGs, showing the finer granularity of ortholog identification among *Lactobacillales*.

Another case analysis revealed an even more complex evolutionary history within the family of enzymes involved in peptidoglycan biosynthesis. The gene in question is present in a single copy in the majority of bacteria and encodes the enzyme UDP-*N*-acetylmuramyl tripeptide synthase (MurE), which catalyzes the addition of mesodiaminopimelate to the peptidoglycan biosynthesis intermediate UDP-*N*-acetylmuramoyl-L-alanyl-glutamate. The phylogenetic tree suggests that two paralogs (LaCOG00743 and LaCOG01200) of this enzyme from another group of bacteria (perhaps, again, actinobacteria) have replaced the ancestral gene in all *Lactobacillales*, with the exception of *Lactobacillus plantarum*, with subsequent functional diversification (Fig. 2). The proteins of LaCOG01200 apparently changed the substrate specificity to lysine and continue to function in the peptidoglycan biosynthesis pathway (4). The function of LaCOG00743 remains unknown, but examination of the genomic context yields some clues. The gene is located in a conserved operon with the gene for  $\alpha$ -carboxyl amidase, which acts on the second amino acid (D-Glu) of the stem peptide precursors; in addition, the LaCOG00743 protein lacks the catalytic residues. Thus, it can be inferred that this protein is the substrate recognition subunit of the amidase.

Another group of interest consists of 707 (22%) LaCOGs that have no counterparts in the general COG set. Of these, 338 (~11%) were shared with one or more non-*Lactobacillales* bacterial genomes among those reported recently and not yet included in the COGs, whereas 369 (~11%) appeared to be specific to *Lactobacillales* and could be considered the genomic signature of the group.

Considerable effort has been invested in assigning reliable functional annotations to LaCOGs to reflect recent experimental findings on LAB genes and to predict functions of uncharacterized proteins on the basis of sequence similarity to proteins with known functions and domain architectures. Thanks to the broad representation of various lineages of *Lactobacillales* in the set of genomes used for LaCOG construction, LaCOGs provide excellent coverage of new genomes

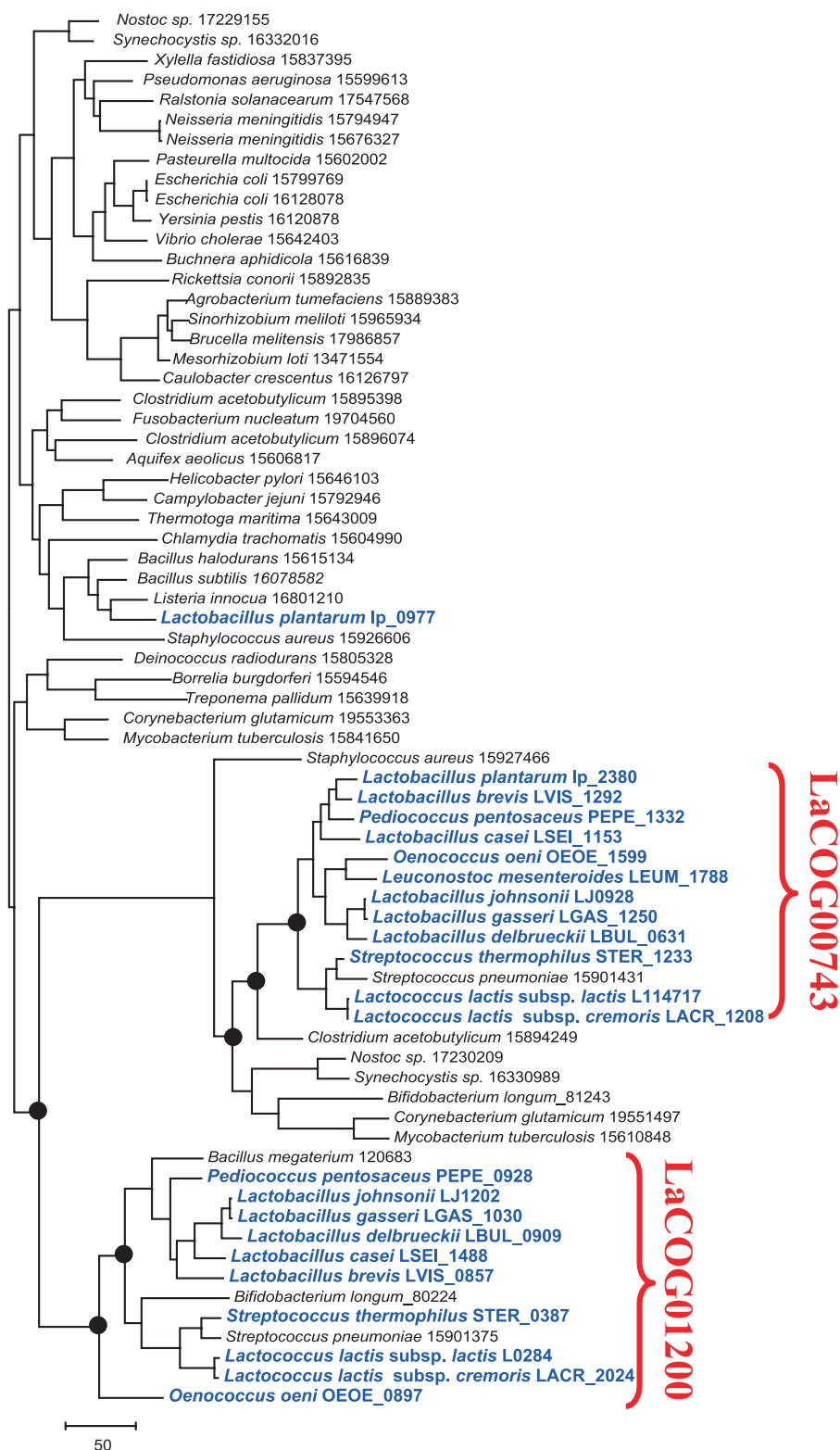


FIG. 2. Phylogenetic analysis of the MurE family of UDP-N-acetylmuramyl tripeptide synthases. The maximum-likelihood unrooted tree was built using the MOLPHY program (1). Each terminal node is labeled with the numeric GenBank identifier number (where available) and the respective species name. The *Lactobacillales* species are shown in blue (temporary numeric identifiers were assigned for new genomes). The circles at some of the nodes in the relevant subtree indicate bootstrap support >70% (not shown for the rest of the tree).

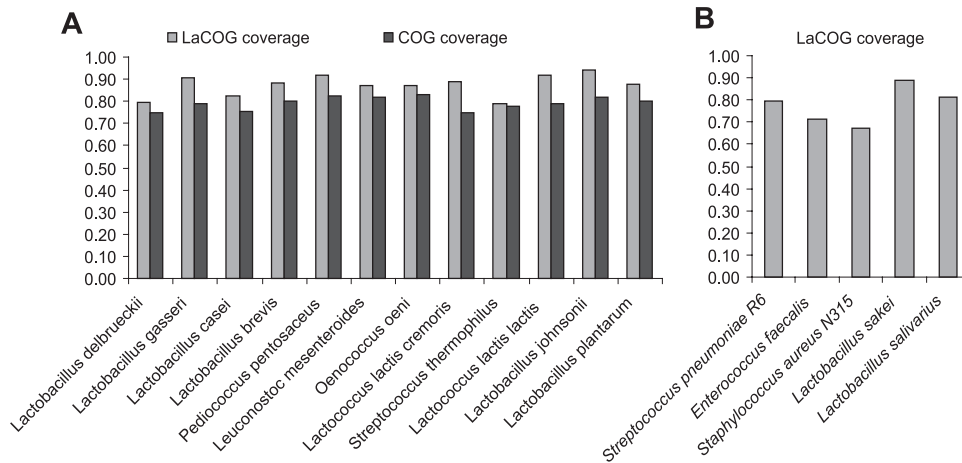


FIG. 3. Coverage of *Lactobacillales* by LaCOGs and COGs. (A) Species that were used for LaCOG construction. (B) Species that were not included in the original LaCOGs.

from this taxon, up to 90% of the genes (Fig. 3). Thus, LaCOGs are expected to become an essential, evolving resource for annotation of new genomes from the order *Lactobacillales*. The complete annotated list of LaCOGs is available at <ftp://ftp.ncbi.nih.gov/pub/wolf/lacto> (file LaCOGS\_table.xls); other detailed results of this analysis, including reconstructions of gene loss and gain, are available upon request.

#### THE CONSERVED GENE CORE OF *LACTOBACILLALES*

An important issue that can be readily addressed with the help of LaCOGs is the identification of the conserved gene core of *Lactobacillales*, i.e., the set of genes that are present in all sequenced genomes from this taxon and, by inference, are likely to be essential for these bacteria. There are 567 LaCOGs (18%) that are present in all 12 analyzed genomes. Not surprisingly, the functional distribution of these LaCOGs shows that the majority encode components of the information-processing systems (translation, transcription, and replication). However, the core also includes 50 genes for which only a general prediction of biochemical activity is available and 41 genes without known or predicted functions. This observation emphasizes the current lack of understanding of even some of the central cellular functions of relatively simple bacteria. Two core genes have no detectable orthologs outside *Lactobacillales* and thus may be considered unique genomic markers of *Lactobacillales*. One of these unique genes encodes a protein containing a peptidoglycan-binding LysM domain (LaCOG01826) (Fig. 4A). In several genomes, this gene is located next to the genes for ribosomal proteins and cytidylate kinase and might be coregulated with these housekeeping genes. The second genomic marker of *Lactobacillales*, the highly conserved LaCOG01237, contains no characterized domains (Fig. 4B). However, this gene is located in a conserved genomic neighborhood encoding two enzymes implicated in 4-thiouridine modification of tRNA [(5-methylaminomethyl-2-thiouridylate) methyltransferase and a predicted sulfurase] (LaCOG00578 and LaCOG01188), suggesting a role of LaCOG01237 proteins in specific modulation of this essential modification (27).

#### PHYLOGENETIC ANALYSIS AND HETEROGENEITY OF EVOLUTIONARY RATES AMONG *LACTOBACILLALES*

The LAB considered here belong to the phylum *Firmicutes*, class *Bacilli*, and order *Lactobacillales*, a sister taxon to the order *Bacillales*. Prior to the recent, extensive genome sequencing, classification of *Lactobacillales* remained an unresolved issue, particularly because the phenotypic classification, which was traditionally based on the type of fermentation, did not match the rRNA-based phylogeny (56). Whole-genome DNA and DNA-RNA hybridization and GC content analysis led to the delineation of three closely related lineages of *Lactobacillales*: the *Leuconostoc* group (*Leuconostoc mesenteroides* and *Oenococcus oeni*), the *Lactobacillus casei*-*Pediococcus* group (*L. plantarum*, *L. casei*, *Pediococcus pentosaceus*, and *Lactobacillus brevis*), and the *Lactobacillus delbrueckii* group (*L. delbrueckii*, *Lactobacillus gasseri*, and *Lactobacillus johnsonii*); streptococci (*Streptococcus thermophilus*) and lactococci (*L. lactis* subsp. *lactis* and *Lactococcus lactis* subsp. *cremoris*) formed a separate branch (48). With the availability of complete genomes for all major branches of *Lactobacillales*, phylogenetic trees can be built by using concatenated protein

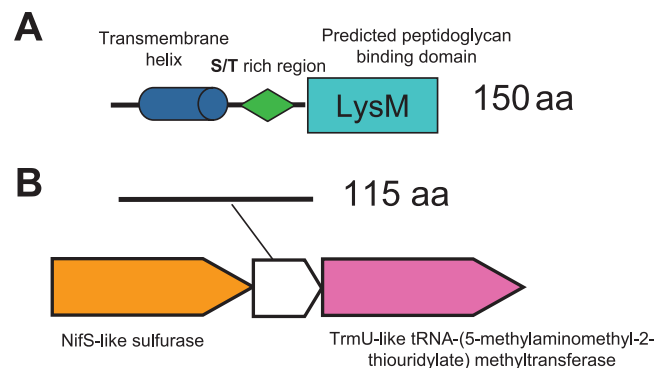


FIG. 4. Two putative genomic markers of *Lactobacillales*. (A) Domain composition of proteins from LaCOG01826. (B) Genome context of genes from LaCOG01826. aa, amino acids.

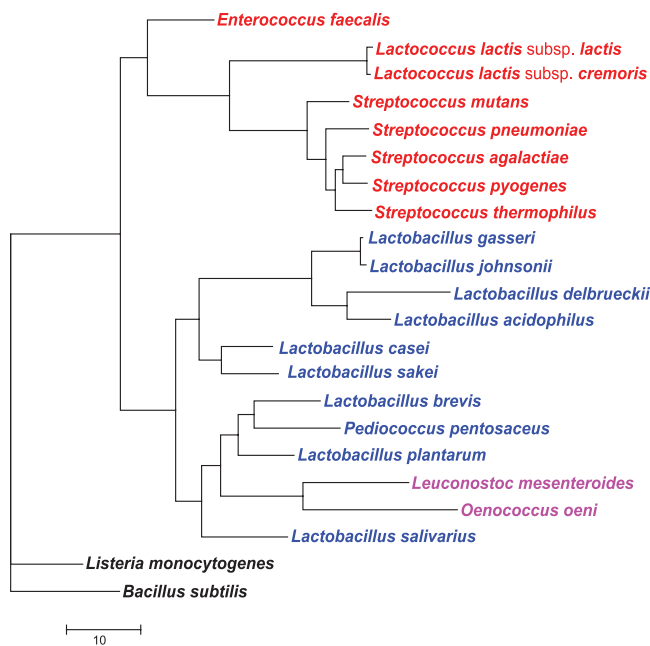


FIG. 5. A phylogenetic tree of *Lactobacillales* constructed on the basis of concatenated alignments of four subunits ( $\alpha$ ,  $\beta$ ,  $\beta'$ , and  $\delta$ ) of the DNA-dependent RNA polymerase. The maximum-likelihood unrooted tree was built using the MOLPHY program (1). All branches are supported with  $>75\%$  bootstrap values. The species are colored according to the current taxonomy: *Lactobacillaceae*, blue; *Leuconostocaceae*, magenta; *Streptococcaceae*, red.

sequences encoded by genes that are unlikely to be transferred horizontally. This approach has been shown to improve the resolution and increase the robustness of phylogenetic analyses (59). A tree of *Lactobacillales* constructed from concatenated sequences of ribosomal proteins and RNA polymerase subunits had the same topology and was supported by high bootstrap values but disagreed in some important respects with the above classification (28). Specifically, the new tree suggests that the *Streptococcus-Lactococcus* branch is basal in the *Lactobacillales* tree and the *Pediococcus* group is a sister to the *Leuconostoc* group within the *Lactobacillus* clade. Thus, the genus *Lactobacillus* appears to be paraphyletic with respect to the *Pediococcus-Leuconostoc* group. *Lactobacillus casei* is confidently placed at the base of the *L. delbrueckii* group. Figure 5 shows the phylogenetic tree of concatenated RNA polymerase subunits for all species of *Lactobacillales* whose genomes are currently available; this tree, made for an expanded species set, was fully compatible with the previous one (28).

A molecular-clock test performed for the phylogenetic tree based on multiple alignment of concatenated ribosomal proteins (51) revealed a high heterogeneity of evolutionary rates among *Lactobacillales*, including confirmation of the previously reported (62) accelerated evolution of the *Leuconostoc* group by a factor of 1.7 to 1.9 relative to the sister *Pediococcus* group (28). Similarly, *O. oeni* was found to evolve substantially faster (by a factor of 1.6) than *Leuconostoc*. This finding is in accord with the experimental observation of an increased mutation rate in *O. oeni* (D. A. Mills, unpublished observation) and the absence in the species of the key enzymes of mismatch

repair, MutL and MutS, which is unique among the *Lactobacillales* (Table 2).

#### PHYLETIC PATTERNS AS A TOOL FOR ANALYSIS OF GENERAL TRENDS OF EVOLUTION FOR DIFFERENT SETS OF SPECIES

Analysis of phyletic (phylogenetic) patterns, i.e., patterns of gene presence/absence in a particular set of genomes, is a valuable approach both for the detection of evolutionary trends and for functional prediction (36, 44). A straightforward examination of frequent phyletic patterns in LaCOGs immediately reveals several trends in the evolution of *Lactobacillales* that mostly reflect gene losses, especially of genes that encode biosynthetic enzymes (Table 2). However, genes shared by distinct sets of bacteria are also of interest. Some of these shared genes apparently reflect recent gene exchanges between distantly related species within the order *Lactobacillales*. Several cases are obvious, e.g., 11 genes that are shared by *L. johnsonii* and *L. lactis* subsp. *cremoris* and are located adjacent to a prophage and therefore in all likelihood have been transferred by the phage vehicle. Another set of genes disseminated via horizontal transfer is the CRISPR-related system (CASS) implicated in the defense against integrative phages and plasmids (6, 33) in *L. delbrueckii* and *L. casei*. In these LAB, the CASS includes a unique gene that encodes a protein with a Cas1 domain fused to a 3'-5' exonuclease domain (29). Other phyletic patterns reflect specific sets of genes shared by related species, often poorly understood in functional terms. Not surprisingly, the second largest gene set (246 genes), after the conserved core, is shared by two most closely related genomes, those of *Lactococcus lactis* subsp. *lactis* and *Lactococcus lactis* subsp. *cremoris*. Many genes in this list ( $>50$ ) apparently belong to prophages that are shared by these two species and have probably integrated into the genome of their common ancestor. Among the 88 genes that are specifically shared by *L. johnsonii* and *L. gasseri*, 48 are uncharacterized; many of them encode secreted and membrane proteins that are likely to be involved in the interaction with mucosal surfaces of the gastrointestinal tract, which these bacteria colonize. Similar trends have been observed for three related species, *L. johnsonii*, *L. gasseri*, and *L. delbrueckii*, that specifically share 39 genes, 27 of which are uncharacterized.

#### GENE LOSS AND GAIN IN THE EVOLUTION OF LACTOBACILLALES

Phyletic patterns of LaCOGs, together with the phylogenetic tree of *Lactobacillales*, can be used for explicit reconstruction of the events that occurred during the evolution of this group after its divergence from the common ancestor with the rest of the *Bacilli*. For the purpose of this reconstruction, we employed a modification of a previously developed method based on the weighted-parsimony approach (32). The results of the reconstruction (28) suggest that the common ancestor of *Lactobacillales* had at least  $\sim 2,100$  to 2,200 genes, having lost 600 to 1,200 genes ( $\sim 25$  to 30%) and gained  $<100$  genes after the divergence from the *Bacilli* ancestor, for which the genome size of  $\sim 2,700$  to 3,700 genes was estimated (Fig. 6). Many of the changes mapped to this stage of evolution seem to be related

TABLE 2. Selected phyletic patterns reflecting the evolutionary trends of different groups of *Lactobacillales*

Pattern <sup>a</sup> : a b c d e f g h i j k l	No. of LaCOGs	Comments and examples
+++++++-----	35	Specific gene loss in <i>O. oeni</i> ; several informational genes lost: MutL, MutS, RadC, Sun
+++++++-----	21	Specific gene loss in <i>L. delbrueckii</i> : several sugar metabolism genes lost: Eno, GalKMT
+++++++-----	17	Specific gene loss in <i>O. oeni</i> and <i>L. mesenteroides</i> ; several RNA modification genes lost: TrmU, TrmE, RsuA, GidA
+++++++-----	12	Specific gene loss in <i>L. johnsonii</i> and <i>L. gasseri</i> ; 8 genes of fatty acid biosynthesis lost
-----	29	Specific loss in <i>S. thermophilus</i> ; 4 genes of 6-phosphogluconate pathway
-----	21	Specific loss <i>Streptococcus-Lactococcus</i> branch; several genes of central metabolism: FumC, GpmA, Gmk
+-----	16	Specific loss in <i>Lactococcus</i> branch; 5 genes for amino acid transport
+++-----	11	A set of bacteria that lost purine biosynthesis genes (9 of 11)
-----	88	Specific signature of gastrointestinal tract-inhabiting bacteria <i>L. johnsonii</i> and <i>L. gasseri</i> ; 48 of these genes are uncharacterized
-----	38	Signature of related bacteria (Fig. 4); 28 of these genes are uncharacterized
+++-----	18	A set of bacteria that encode genes for aromatic acid biosynthesis and histidine biosynthesis (14 of 18)
-----	16	Genes shared between <i>L. delbrueckii</i> and <i>L. casei</i> : 8 belong to a distinct version of the CASS
+++-----	8	All these genes are involved in amino acid biosynthesis, mostly proline and tryptophan

<sup>a</sup> +, gene is present in the species; -, gene is absent. Species are denoted by letters as follows: a, *S. thermophilus*; b, *L. lactis* subsp. *lactis*; c, *L. lactis* subsp. *cremoris*; d, *L. brevis*; e, *L. plantarum*; f, *P. pentosaceus*; g, *L. mesenteroides*; h, *O. oeni*; i, *L. johnsonii*; j, *L. gasseri*; k, *L. delbrueckii* subsp. *bulgaricus*; l, *L. casei*.

to the transition made by the LAB to existence in nutritionally rich medium. Thus, a number of genes for biosynthesis of cofactors, such as heme, molybdenum coenzyme, and pantothenate, were lost, and conversely, some cofactor transporters were acquired, e.g., nicotinamide mononucleotide transporter. Another notable acquisition is a group of diverse peptidases which are obviously an important commodity in the protein-rich environments inhabited by the LAB. The loss of heme/copper-type cytochrome/quinol oxidase-related genes (CyoABCDE) and catalase (KatE), characteristic enzymes of aerobic bacteria, suggest that the ancestor of *Lactobacillales* was a microaerophile or an anaerobe.

Lineage-specific gene loss was extensive in the evolution of all lineages of *Lactobacillales*, but several species stand out as especially notable “losers.” In particular, *S. thermophilus* not only lost numerous genes but also has many fresh pseudogenes,

suggesting an active and ongoing process of genome decay, which has been reported for two different strains of the same species (5). Moreover, substantial gene loss (368 genes, according to the present reconstruction) also occurred at the base of the *Streptococcus-Lactococcus* branch, including several genes involved in cell division that are conserved in most bacteria, such as CrcB, MreB, MreC, and MinD. This is reminiscent of the trends of gene loss that are observed in other symbiotic and pathogenic bacteria (21, 35). The lineages of *Lactobacillales* that are particularly prone to gene loss are *P. pentosaceus*; the *Leuconostoc* and *Oenococcus* branch, with considerable additional loss in each species; and the *L. delbrueckii* group (*L. delbrueckii*, *L. gasseri*, and *L. johnsonii*), with further genome reduction in *L. gasseri* and *L. johnsonii* (Fig. 6). In the species with larger genomes, such as *L. plantarum* and *L. casei*, the loss of ancestral genes was counterbalanced by the emergence of

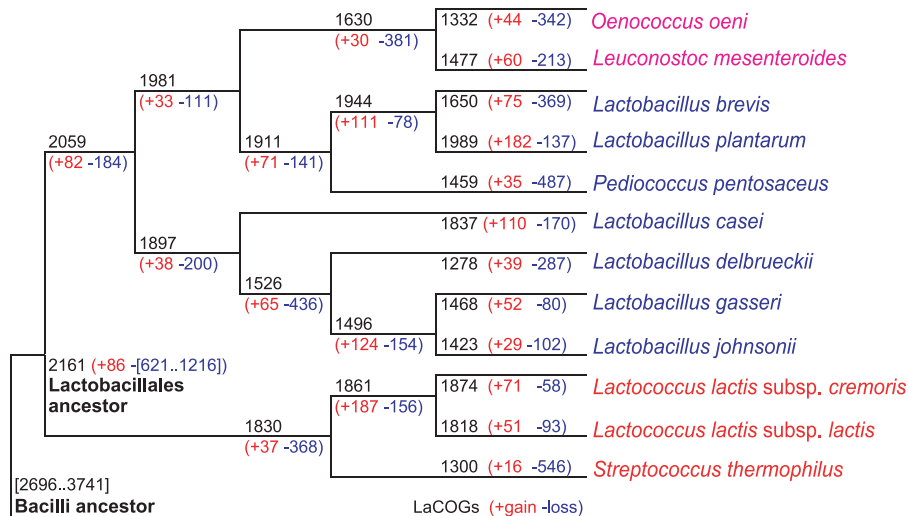


FIG. 6. Reconstruction of gene content evolution in *Lactobacillales*. The tree is a subset of that shown in Fig. 5, rooted by using *Bacillus subtilis* as the outgroup. For each species and each internal node of the tree, the inferred number of LaCOGs present and the numbers of LaCOGs lost (blue) and gained (red) along the branch leading to the given node (species) are indicated. Modified from reference 28 with permission of the publisher.

many new genes via duplication and horizontal gene transfer (HGT) (Fig. 6).

### HORIZONTAL GENE TRANSFER IN *LACTOBACILLALES*

Horizontal gene transfer via bacteriophage-mediated or conjugative pathways has been extensively documented in *Lactobacillales* and appears to be important for niche-specific adaptation in the lactococci (61). Signs of HGT are particularly notable in *L. lactis* subsp. *cremoris* SK11, which harbors a conjugative plasmid (pLAC3) and several additional plasmids carrying genes related to growth in milk (47).

Horizontal gene transfer definitely played an important role in shaping the common ancestor of *Lactobacillales*. As many as 84 genes that were inferred to have been acquired by the *Lactobacillales* ancestor (Fig. 6) most probably were horizontally transferred from different sources (only 2 of the 86 acquired genes currently do not have orthologs outside *Lactobacillales*). In some cases, the ancestor acquired an additional, pseudoparalogous copy of a gene by HGT (e.g., the aforementioned enolases), whereas on other occasions, xenologous displacement (acquisition of genes via HGT, followed by the loss of the ancestral orthologous gene) (26) apparently took place, as in the case of the two forms of the MurE-like UDP-*N*-acetylmuramyl tripeptide synthase.

Recently, a simple approach for the detection of violations of the molecular clock in individual COGs has been developed (41). The approach is based on comparing the evolutionary distances within a set of orthologs to a standard intergenomic distance measured for genes known to be less prone than others to HGT and deviations in evolutionary rates (e.g., ribosomal proteins and subunits of RNA polymerase). Most often, statistically significant deviations from the molecular clock are best explained by HGT (41). This test was applied to the LaCOGs, and significant violations of the molecular clock were detected in at least 25% of the LaCOGs, suggesting a high level of HGT and/or major local accelerations of evolution. Molecular-clock violations are particularly common in certain functional groups of genes, such as those encoding enzymes of sugar metabolism, including key enzymes, such as phosphoketolase, transketolase, and various components of phosphotransferase systems.

Furthermore, most of the unique genes that are present in the individual genomes (not covered by LaCOGs) and that have homologs outside *Lactobacillales* are probable products of recent HGT. Among such examples are two copies of the Mn-containing catalase (COG3546) in *P. pentosaceus* that is shared with a different strain of *L. plantarum* (3) and that were probably transferred to the ancestral lineage of the *Pediococcus* group from a source belonging to *Bacilli* (but not *Lactobacillales*), with subsequent loss in other members of the *Pediococcus* group; a distinct form of the predicted polymerase of CASS (COG1353) that is most often present in thermophiles and was detected in all strains of *S. thermophilus* (29); the urease complex that is present in *S. thermophilus* but absent in other *Lactobacillales* and that has been shown to influence the rate of milk acidification (34); and the propanediol utilization operon, which is present in *L. brevis* and so far has been found only in a few species from different lineages, including *Lactobacillus collinoides*, a heterofermentative LAB contained

in cider, where the enzymes encoded in this operon are thought to be involved in glycerol degradation (45).

### LINEAGE-SPECIFIC EXPANSION OF PARALOGOUS GENE FAMILIES

Like all other bacterial lineages (16), *Lactobacillales* have a substantial number of expanded gene families that evolved either by lineage-specific gene duplication or by acquisition of pseudoparalogous genes via HGT (26). A closer examination of these families indicates that adaptation to growth in nutrient-rich environments was the major driving force behind the fixation of duplications and acquisitions during the evolution of the *Lactobacillales*.

Many genes encoding proteins involved in sugar metabolism and transport were duplicated or acquired early in the evolution of the *Lactobacillales*, including those encoding enolase, several phosphotransferase systems,  $\beta$ -galactosidase, GpmB-family sugar phosphatases, galactose mutarotase, and others. In addition, expansion of peptidases and amino acid transporters seems to have occurred in several lineages of *Lactobacillales*. Interestingly, several expanded families include proteins involved in antibiotic resistance in other bacteria, such as  $\beta$ -lactamases and penicillin V acylases, despite the fact that most LAB species are sensitive to common antibiotics and, after centuries of consumption by humans, have been accordingly “generally recognized as safe” (17, 54). Conceivably, the homologs of antibiotic resistance genes are involved in normal cell wall biosynthesis in the *Lactobacillales*. In the same vein, expansion of a distinct family of tyrosine/serine phosphatases, which are often located in the same operon with a serine/threonine protein kinase fused to  $\beta$ -lactam-binding (PASTA) domains, is likely to be important for regulation of cell wall biosynthesis (63). Furthermore, *Lactobacillales* encode a paralog of class II lysyl-tRNA synthetase that is fused to a membrane-associated domain (COG2898) implicated in oxacillin-like antibiotic resistance (40) and is probably involved in cell wall biosynthesis.

### SIMILAR EVOLUTIONARY TRENDS IN *LACTOBACILLALES* AND *BIFIDOBACTERIUM LONGUM*

*Bifidobacterium longum* is a LAB that belongs to a different major bacterial branch, the actinobacteria. The complete genome sequence of *B. longum* has been reported (46). This bacterium also inhabits the gastrointestinal tract and so shares the environment with several *Lactobacillales* (Table 1). Identification of common genes between *B. longum* and *Lactobacillales* is of interest because it has the potential to delineate a “genomic cognate” of the LAB phenotype. However, only seven genes that are present in *B. longum* but not in the genomes of non-LAB actinobacteria are shared specifically with *Lactobacillales*. Only one of these genes, which encodes a functionally uncharacterized membrane protein, is present in seven genomes of *Lactobacillales* (LaCOG00453), whereas the rest are present only in *B. longum* and two *Lactobacillales* species. By contrast, common trends of gene loss among LAB with different taxonomic affinities are obvious (Table 3). The majority of these common losses map to the last common

TABLE 3. Common trends of gene loss and gain of *Bifidobacterium longum* and *Lactobacillales*

Functional group of genes	Examples of genes/systems lost (gained <sup>a</sup> ) in <i>B. longum</i>	Also lost (gained <sup>a</sup> ) in:
Sugar and energy metabolism	Heme/copper-type cytochrome/quinol oxidases ( <i>cyoABCDE</i> )	LCA <sup>b</sup> of <i>Lactobacillales</i>
Amino acid metabolism	Glycine cleavage system ( <i>gcvTRP</i> )	LCA of <i>Lactobacillales</i>
Cofactor biosynthesis	Most of heme biosynthesis ( <i>hemABCDEFGHIY</i> )	LCA of <i>Lactobacillales</i>
Fatty acid biosynthesis and metabolism	Most of the genes in the pathways ( <i>accABCD, fabHB</i> )	Common ancestor of <i>L. johnsonii</i> and <i>L. gasseri</i>
Transporters	Multisubunit Na <sup>+</sup> /H <sup>+</sup> antiporter ( <i>mnhABCDEFG</i> ) Na <sup>+</sup> /alanine symporter ( <i>alsT</i> )	LCA of <i>Lactobacillales</i>
Oxidative stress	Catalase ( <i>katE</i> )	LCA of <i>Lactobacillales</i>
Amino acid metabolism	<b>Dipeptidase (<i>pepD</i>)</b>	<b>LCA of <i>Lactobacillales</i></b>

<sup>a</sup> Gains are shown in boldface.

<sup>b</sup> LCA, last common ancestor.

ancestor of *Lactobacillales*. These observations indicate that convergent evolution of the LAB phenotype in different bacterial lineages was accompanied by similar processes of extensive loss of genes, primarily those encoding a variety of metabolic capabilities, whereas acquisition of new genes was much less extensive and involved different gene sets. At least in the case of *B. longum* and the *Lactobacillales*, there is no evidence of substantial HGT between taxonomically distant LAB.

#### USING THE GENE CONTEXT APPROACH FOR DETECTION OF NEW BACTERIOCINS

*Lactobacillales* are known for producing specific antimicrobial peptides, the bacteriocins (38, 55). Several proteins responsible for the modification and export of bacteriocins and regulation of bacteriocin biosynthesis are often encoded in the same operon with the bacteriocins themselves (38, 55). Since bacteriocins are small proteins with highly diverged sequences,

they are often difficult to identify by amino acid sequence conservation. A comparative-genomic approach is likely to be effective for a more complete characterization of the bacteriocin repertoire of *Lactobacillales*. Indeed, in seven *Lactobacillales* genomes, we identified clustered genes for putative bacteriocins and associated proteins. Several of these newly detected candidate bacteriocins clearly belong to two families characterized previously: homologs of pediocin from *P. pentosaceus*, homologs of which are also present in *L. mesenteroides* and *L. casei*, and homologs of divercin V41 (30), which are present in *P. pentosaceus* and *L. johnsonii*. In addition, numerous small open reading frames located in the immediate vicinity of the genes for bacteriocins and associated proteins might encode novel bacteriocins, despite the lack of sequence similarity to known ones (Fig. 7). Recently, a similar strategy for genome mining in search of new bacteriocins has been implemented in a specialized online server (11).

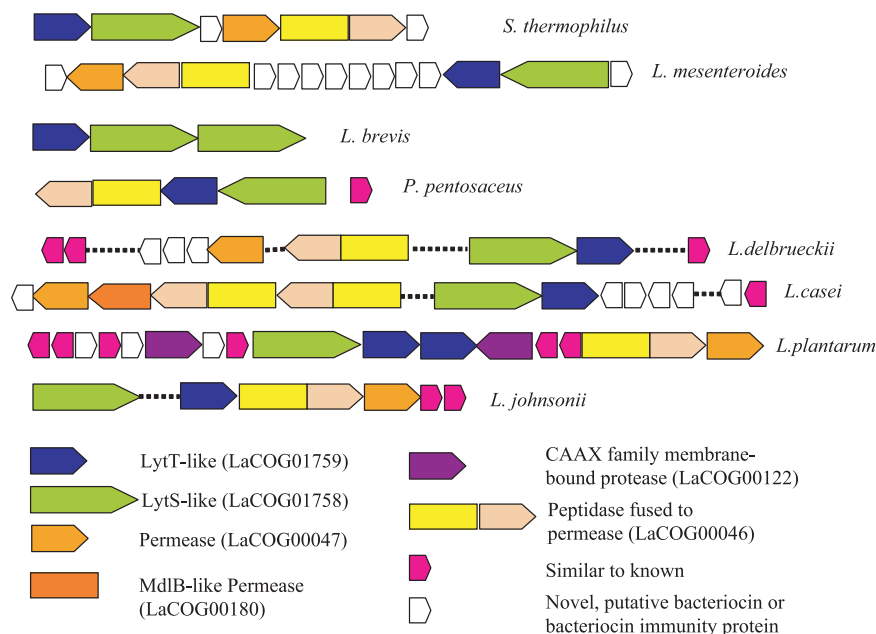


FIG. 7. Gene clusters in *Lactobacillales* encoding known and predicted bacteriocins and bacteriocin export systems. Modified from reference 28 with permission of the publisher.



## CONCLUSIONS

The sequencing of multiple complete genomes has created unprecedented opportunities for evolutionary genomics of LAB, particularly those of the *Lactobacillales* lineage. The clusters of orthologous genes from *Lactobacillales* (LaCOGs) provide a convenient, flexible framework for both functional annotation of new genomes of *Lactobacillales* and evolutionary reconstruction. Loss of ancestral genes, primarily those for various metabolic enzymes, comes across as the central theme in the evolution of *Lactobacillales*, with a clear connection with the adaptation of these bacteria to their nutrient-rich habitats. Substantial gene loss had already occurred at an early stage of the evolution of *Lactobacillales*, after their divergence from the common ancestor with the rest of the *Bacilli* but before the radiation of the extant species. Additional differential gene loss took place during the subsequent evolution of each lineage of *Lactobacillales*. The trends of gene loss are very similar in taxonomically distant LAB, particularly between *Lactobacillales* and *B. longum*, an actinomycete, conceivably due to similar environmental pressures. However, the repertoires of genes that were acquired via HGT are quite different, indicating that the LAB phenotype evolved convergently in diverse lineages of bacteria as opposed to horizontal transfer of a unique suite of genes. Some of the genes acquired by *Lactobacillales* are clearly adaptations to existence in the nutrient-rich habitats of these bacteria. Comparative-genomic analysis substantially facilitates functional annotation of LAB genomes. In particular, this analysis helps to predict new bacteriocins, antimicrobial peptides that are typically produced by LAB, apparently reflecting their long-term existence in complex microbial communities.

## ACKNOWLEDGMENTS

We thank the members of the Lactic Acid Bacteria Genome Sequencing Consortium and, personally, David Mills and Sergay Kozyavkin for extensive collaboration and numerous helpful discussions in the course of the comparative-genomic analysis of LAB.

## REFERENCES

- Adachi, J., and M. Hasegawa. 1992. MOLPHY: programs for molecular phylogenetics. Computer Science Monographs 27. Institute of Statistical Mathematics, Tokyo, Japan.
- Altermann, E., W. M. Russell, M. A. Azcarate-Peril, R. Barrangou, B. L. Buck, O. McAuliffe, N. Souther, A. Dobson, T. Duong, M. Callanan, S. Lick, A. Hamrick, R. Cano, and T. R. Klaenhammer. 2005. Complete genome sequence of the probiotic lactic acid bacterium *Lactobacillus acidophilus* NCFM. Proc. Natl. Acad. Sci. USA 102:3906–3912.
- Barynin, V. V., M. M. Whittaker, S. V. Antonyuk, V. S. Lamzin, P. M. Harrison, P. J. Artymiuk, and J. W. Whittaker. 2001. Crystal structure of manganese catalase from *Lactobacillus plantarum*. Structure 9:725–738.
- Blewett, A. M., A. J. Lloyd, A. Echalié, V. Fulop, C. G. Dowson, T. D. Bugg, and D. I. Roper. 2004. Expression, purification, crystallization and preliminary characterization of uridine 5'-diphospho-*N*-acetylmuramoyl L-alanyl-D-glutamate:lysine ligase (MurE) from *Streptococcus pneumoniae* 110K/70. Acta Crystallogr. D 60:359–361.
- Bolotin, A., B. Quinquis, P. Renault, A. Sorokin, S. D. Ehrlich, S. Kulakauskas, A. Lapidus, E. Goltsman, M. Mazur, G. D. Pusch, M. Fonstein, R. Overbeek, N. Kyrpides, B. Purnelle, D. Prozzi, K. Ngui, D. Masuy, F. Hancy, S. Burtéau, M. Boutry, J. Delcour, A. Goffeau, and P. Hols. 2004. Complete sequence and comparative genome analysis of the dairy bacterium *Streptococcus thermophilus*. Nat. Biotechnol. 22:1554–1558.
- Bolotin, A., B. Quinquis, A. Sorokin, and S. D. Ehrlich. 2005. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. Microbiology 151:2551–2561.
- Bolotin, A., P. Wincker, S. Mauger, O. Jaillon, K. Malarme, J. Weissenbach, S. D. Ehrlich, and A. Sorokin. 2001. The complete genome sequence of the lactic acid bacterium *Lactococcus lactis* ssp. *lactis* IL1403. Genome Res. 11:731–753.
- Chaillou, S., M. C. Champomier-Verges, M. Cornet, A. M. Crutz-Le Coq, A. M. Duzé, V. Martin, S. Beauvais, E. Darbon-Rongere, R. Bossy, V. Loux, and M. Zagorec. 2005. The complete genome sequence of the meat-borne lactic acid bacterium *Lactobacillus sakei* 23K. Nat. Biotechnol. 23:1527–1533.
- Claesson, M. J., Y. Li, S. Leahy, C. Canchaya, J. P. van Pijkeren, A. M. Cerdeno-Tarraga, J. Parkhill, S. Flynn, G. C. O'Sullivan, J. K. Collins, D. Higgins, F. Shanahan, G. F. Fitzgerald, D. van Sinderen, and P. W. O'Toole. 2006. Multireplicon genome architecture of *Lactobacillus salivarius*. Proc. Natl. Acad. Sci. USA 103:6718–6723.
- Cotter, P. D., C. Hill, and R. P. Ross. 2005. Bacteriocins: developing innate immunity for food. Nat. Rev. Microbiol. 3:777–788.
- de Jong, A., S. A. van Hijum, J. J. Bijlsma, J. Kok, and O. P. Kuipers. 2006. BAGEL: a web-based bacteriocin genome mining tool. Nucleic Acids Res. 34:W273–W279.
- de Vos, W. M., M. Kleerebezem, and O. P. Kuipers. 2005. Lactic acid bacteria—genetics, metabolism and application. FEMS Microbiol. Rev. 29:391.
- Dunny, G. M., and P. P. Cleary. 1991. Genetics and molecular biology of streptococci, lactococci, and enterococci. American Society for Microbiology, Washington, DC.
- Hols, P., F. Hancy, L. Fontaine, B. Grossiord, D. Prozzi, N. Leblond-Bourget, B. Decaris, A. Bolotin, C. Delorme, S. Dusko Ehrlich, E. Guedon, V. Monnet, P. Renault, and M. Kleerebezem. 2005. New insights in the molecular biology and physiology of *Streptococcus thermophilus* revealed by comparative genomics. FEMS Microbiol. Rev. 29:435–463.
- Jamet, E., S. D. Ehrlich, F. Duperray, and P. Renault. 2001. Étude des gènes dupliqués de la glycolyse chez *Lactococcus lactis* il1403. Lait 81:115–129.
- Jordan, I. K., K. S. Makarova, J. L. Spouge, Y. I. Wolf, and E. V. Koonin. 2001. Lineage-specific gene expansions in bacterial and archaeal genomes. Genome Res. 11:555–565.
- Katla, A. K., H. Kruse, G. Johnsen, and H. Herikstad. 2001. Antimicrobial susceptibility of starter culture bacteria used in Norwegian dairy products. Int. J. Food Microbiol. 67:147–152.
- Kilstrup, M., K. Hammer, P. Ruhdal Jensen, and J. Martinussen. 2005. Nucleotide metabolism and its control in lactic acid bacteria. FEMS Microbiol. Rev. 29:555–590.
- Klaenhammer, T., E. Altermann, F. Arigoni, A. Bolotin, F. Breidt, J. Broadbent, R. Cano, S. Chaillou, J. Deutscher, M. Gasson, M. van de Guchte, J. Guzzo, A. Hartke, T. Hawkins, P. Hols, R. Hutkins, M. Kleerebezem, J. Kok, O. Kuipers, M. Lubbers, E. Maguin, L. McKay, D. Mills, A. Nauta, R. Overbeek, H. Pel, D. Pridmore, M. Saier, D. van Sinderen, A. Sorokin, J. Steele, D. O'Sullivan, W. de Vos, B. Weimer, M. Zagorec, and R. Siezen. 2002. Discovering lactic acid bacteria by genomics. Antonie Leeuwenhoek 82:29–58.
- Klaenhammer, T. R., R. Barrangou, B. L. Buck, M. A. Azcarate-Peril, and E. Altermann. 2005. Genomic features of lactic acid bacteria effecting bioprocessing and health. FEMS Microbiol. Rev. 29:393–409.
- Klasson, L., and S. G. Andersson. 2004. Evolution of minimal-gene-sets in host-dependent bacteria. Trends Microbiol. 12:37–43.
- Kleerebezem, M., J. Boekhorst, R. van Kranenburg, D. Molenaar, O. P. Kuipers, R. Leer, R. Turchini, S. A. Peters, H. M. Sandbrink, M. W. Fiers, W. Stiekema, R. M. Lankhorst, P. A. Bron, S. M. Hoffer, M. N. Groot, R. Kerkhoven, M. de Vries, B. Ursing, W. M. de Vos, and R. J. Siezen. 2003. Complete genome sequence of *Lactobacillus plantarum* WCFS1. Proc. Natl. Acad. Sci. USA 100:1990–1995.
- Klijn, A., A. Mercenier, and F. Arigoni. 2005. Lessons from the genomes of bifidobacteria. FEMS Microbiol. Rev. 29:491–509.
- Kok, J., G. Buist, A. L. Zomer, S. A. van Hijum, and O. P. Kuipers. 2005. Comparative and functional genomics of lactococci. FEMS Microbiol. Rev. 29:411–433.
- Koonin, E. V. 2005. Orthologs, paralogs, and evolutionary genomics. Annu. Rev. Genet. 39:309–338.
- Koonin, E. V., K. S. Makarova, and L. Aravind. 2001. Horizontal gene transfer in prokaryotes—quantification and classification. Annu. Rev. Microbiol. 55:709–742.
- Leipuviene, R., Q. Qian, and G. R. Bjork. 2004. Formation of thiolated nucleosides present in tRNA from *Salmonella enterica* serovar Typhimurium occurs in two principally distinct pathways. J. Bacteriol. 186:758–766.
- Makarova, K., A. Slesarev, Y. Wolf, A. Sorokin, E. Koonin, A. Pavlov, N. Pavlova, V. Karamychev, N. Polouchin, V. Shakhova, I. Grigoriev, Y. Lou, D. Rohksar, S. Lucas, K. Huang, D. M. Goodstein, T. Hawkins, V. Plengvidhya, D. Welker, J. Hughes, Y. Goh, A. Benson, K. Baldwin, J.-H. Lee, I. Diaz-Muniz, B. Dosti, V. Smeianov, W. Wechter, R. Barabote, G. Lorca, E. Altermann, R. Barrangou, B. Ganesan, Y. Xie, H. Rawsthorne, D. Tamir, C. Parker, F. Breidt, J. Broadbent, R. Hutkins, D. O'Sullivan, J. Steele, G. Unlu, M. Saier, T. Klaenhammer, P. Richardson, S. Kozyavkin, B. Weimer, and D. Mills. 2006. Comparative genomics of the lactic acid bacteria. Proc. Natl. Acad. Sci. USA 103:15611–15616.
- Makarova, K. S., N. V. Grishin, S. A. Shabalina, Y. I. Wolf, and E. V. Koonin. 2006. A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional

- analogies with eukaryotic RNAi, and hypothetical mechanisms of action. *Biol. Direct.* **1**:7.
30. Metivier, A., M. F. Pilet, X. Dousset, O. Sorokine, P. Anglade, M. Zagorec, J. C. Piard, D. Marion, Y. Cenatiempo, and C. Fremaux. 1998. Divercin V41, a new bacteriocin with two disulphide bonds produced by *Carnobacterium divergens* V41: primary structure and genomic organization. *Microbiology* **144**:2837–2844.
  31. Mills, D. A., H. Rawsthorne, C. Parker, D. Tamir, and K. Makarova. 2005. Genomic analysis of *Oenococcus oeni* PSU-1 and its relevance to winemaking. *FEMS Microbiol. Rev.* **29**:465–475.
  32. Mirkin, B. G., T. I. Fenner, M. Y. Galperin, and E. V. Koonin. 2003. Algorithms for computing parsimonious evolutionary scenarios for genome evolution, the last universal common ancestor and dominance of horizontal gene transfer in the evolution of prokaryotes. *BMC Evol. Biol.* **3**:2.
  33. Mojica, F. J., C. Diez-Villasenor, J. Garcia-Martinez, and E. Soria. 2005. Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J. Mol. Evol.* **60**:174–182.
  34. Mora, D., E. Maguin, M. Masiero, C. Parini, G. Ricci, P. L. Manachini, and D. Daffonchio. 2004. Characterization of urease genes cluster of *Streptococcus thermophilus*. *J. Appl. Microbiol.* **96**:209–219.
  35. Moran, N. A. 2003. Tracing the evolution of gene loss in obligate bacterial symbionts. *Curr. Opin. Microbiol.* **6**:512–518.
  36. Natale, D. A., M. Y. Galperin, R. L. Tatusov, and E. V. Koonin. 2000. Using the COG database to improve gene recognition in complete genomes. *Genetica* **108**:9–17.
  37. Natale, D. A., U. T. Shankavaram, M. Y. Galperin, Y. I. Wolf, L. Aravind, and E. V. Koonin. 2000. Towards understanding the first genome sequence of a crenarchaeon by genome annotation using clusters of orthologous groups of proteins (COGs). *Genome Biol.* **1**:RESEARCH0009.
  38. Nes, I. F., and O. Johnsborg. 2004. Exploration of antimicrobial potential in LAB by genomics. *Curr. Opin. Biotechnol.* **15**:100–104.
  39. Neves, A. R., W. A. Pool, J. Kok, O. P. Kuipers, and H. Santos. 2005. Overview on sugar metabolism and its control in *Lactococcus lactis*—the input from in vivo NMR. *FEMS Microbiol. Rev.* **29**:531–554.
  40. Nishi, H., H. Komatsuzawa, T. Fujiwara, N. McCallum, and M. Sugai. 2004. Reduced content of lysyl-phosphatidylglycerol in the cytoplasmic membrane affects susceptibility to moenomycin, as well as vancomycin, gentamicin, and antimicrobial peptides, in *Staphylococcus aureus*. *Antimicrob. Agents Chemother.* **48**:4800–4807.
  41. Novichkov, P. S., M. V. Omelchenko, M. S. Gelfand, A. A. Mironov, Y. I. Wolf, and E. V. Koonin. 2004. Genome-wide molecular clock and horizontal gene transfer in bacterial evolution. *J. Bacteriol.* **186**:6575–6585.
  42. Pedersen, M. B., S. L. Iversen, K. I. Sorensen, and E. Johansen. 2005. The long and winding road from the research laboratory to industrial applications of lactic acid bacteria. *FEMS Microbiol. Rev.* **29**:611–624.
  43. Pridmore, R. D., B. Berger, F. Desiere, D. Vilanova, C. Barretto, A. C. Pittet, M. C. Zwahlen, M. Rouvet, E. Altermann, R. Barrangou, B. Mollet, A. Mercenier, T. Klaenhammer, F. Arigoni, and M. A. Schell. 2004. The genome sequence of the probiotic intestinal bacterium *Lactobacillus johnsonii* NCC 533. *Proc. Natl. Acad. Sci. USA* **101**:2512–2517.
  44. Reichard, K., and M. Kaufmann. 2003. EPPS: mining the COG database by an extended phylogenetic patterns search. *Bioinformatics* **19**:784–785.
  45. Sauvageot, N., C. Muller, A. Hartke, Y. Auffray, and J. M. Laplace. 2002. Characterisation of the diol dehydratase *pdu* operon of *Lactobacillus collinoides*. *FEMS Microbiol. Lett.* **209**:69–74.
  46. Schell, M. A., M. Karmirantzou, B. Snel, D. Vilanova, B. Berger, G. Pessi, M. C. Zwahlen, F. Desiere, P. Bork, M. Delley, R. D. Pridmore, and F. Arigoni. 2002. The genome sequence of *Bifidobacterium longum* reflects its adaptation to the human gastrointestinal tract. *Proc. Natl. Acad. Sci. USA* **99**:14422–14427.
  47. Siezen, R. J., B. Renckens, I. van Swam, S. Peters, R. van Kranenburg, M. Kleerebezem, and W. M. de Vos. 2005. Complete sequences of four plasmids of *Lactococcus lactis* subsp. *cremoris* SK11 reveal extensive adaptation to the dairy environment. *Appl. Environ. Microbiol.* **71**:8371–8382.
  48. Siezen, R. J., F. H. van Enckevort, M. Kleerebezem, and B. Teusink. 2004. Genome data mining of lactic acid bacteria: the impact of bioinformatics. *Curr. Opin. Biotechnol.* **15**:105–115.
  49. Smit, G., B. A. Smit, and W. J. Engels. 2005. Flavour formation by lactic acid bacteria and biochemical flavour profiling of cheese products. *FEMS Microbiol. Rev.* **29**:591–610.
  50. Stiles, M. E., and W. H. Holzapfel. 1997. Lactic acid bacteria of foods and their current taxonomy. *Int. J. Food Microbiol.* **36**:1–29.
  51. Takezaki, N., A. Rzhetsky, and M. Nei. 1995. Phylogenetic test of the molecular clock and linearized trees. *Mol. Biol. Evol.* **12**:823–833.
  52. Tatusov, R. L., N. D. Fedorova, J. D. Jackson, A. R. Jacobs, B. Kiryutin, E. V. Koonin, D. M. Krylov, R. Mazumder, S. L. Mekhedov, A. N. Nikolskaya, B. S. Rao, S. Smirnov, A. V. Sverdlov, S. Vasudevan, Y. I. Wolf, J. J. Yin, and D. A. Natale. 2003. The COG database: an updated version includes eukaryotes. *BMC Bioinformatics* **4**:41.
  53. Tatusov, R. L., E. V. Koonin, and D. J. Lipman. 1997. A genomic perspective on protein families. *Science* **278**:631–637.
  54. Teuber, M., L. Meile, and F. Schwarz. 1999. Acquired antibiotic resistance in lactic acid bacteria from food. *Antonie Leeuwenhoek* **76**:115–137.
  55. Twomey, D., R. P. Ross, M. Ryan, B. Meaney, and C. Hill. 2002. Lantibiotics produced by lactic acid bacteria: structure, function and applications. *Antonie Leeuwenhoek* **82**:165–185.
  56. Vandamme, P., B. Pot, M. Gillis, P. de Vos, K. Kersters, and J. Swings. 1996. Polyphasic taxonomy, a consensus approach to bacterial systematics. *Microbiol. Rev.* **60**:407–438.
  57. van de Guchte, M., S. Pinaud, C. Grimaldi, V. Barbe, K. Bryson, P. Nicolas, C. Robert, S. Oztas, S. Mangenot, A. Couloux, V. Loux, R. Dervyn, R. Bossy, A. Bolotin, J. M. Batto, T. Walunas, J. F. Gibrat, P. Bessieres, J. Weissenbach, S. D. Ehrlich, and E. Maguin. 2006. The complete genome sequence of *Lactobacillus bulgaricus* reveals extensive and ongoing reductive evolution. *Proc. Natl. Acad. Sci. USA* **103**:9274–9279.
  58. Vaughan, E. E., H. G. Heilig, K. Ben-Amor, and W. M. de Vos. 2005. Diversity, vitality and activities of intestinal lactic acid bacteria and bifidobacteria assessed by molecular approaches. *FEMS Microbiol. Rev.* **29**:477–490.
  59. Wolf, Y. I., I. B. Rogozin, N. V. Grishin, R. L. Tatusov, and E. V. Koonin. 2001. Genome trees constructed using five different approaches suggest new major bacterial clades. *BMC Evol Biol.* **1**:8.
  60. Wood, B. J. B., and W. H. Holzapfel. 1995. The genera of lactic acid bacteria, 1st ed. Blackie Academic and Professional, Glasgow, United Kingdom.
  61. Wood, B. J. B., and P. J. Warner. 2003. Genetics of lactic acid bacteria. Kluwer Academic/Plenum Publishers, New York, NY.
  62. Yang, D., and C. R. Woese. 1989. A phylogenetic analysis of lactobacilli, *Pediococcus pentosaceus* and *Leuconostoc mesenteroides*. *Syst. Appl. Microbiol.* **12**:145–149.
  63. Yeats, C., R. D. Finn, and A. Bateman. 2002. The PASTA domain: a beta-lactam-binding domain. *Trends Biochem. Sci.* **27**:438.