

# Correlation of *Lactobacillus rhamnosus* Genotypes and Carbohydrate Utilization Signatures Determined by Phenotype Profiling

Corina Ceapa,<sup>b,c</sup> Jolanda Lambert,<sup>b</sup> Kees van Limpt,<sup>b</sup> Michiel Wels,<sup>d</sup> Tamara Smokvina,<sup>a</sup> Jan Knol,<sup>b,c</sup> Michiel Kleerebezem<sup>d,e</sup>

Danone Nutricia Research, Palaiseau, France<sup>a</sup>; Danone Nutricia Research, Utrecht, The Netherlands<sup>b</sup>; Laboratory of Microbiology, Wageningen University, Wageningen, The Netherlands<sup>c</sup>; NIZO Food Research, Ede, The Netherlands<sup>d</sup>; Host-Microbe Interactomics, Wageningen University, Wageningen, The Netherlands<sup>e</sup>

*Lactobacillus rhamnosus* is a bacterial species commonly colonizing the gastrointestinal (GI) tract of humans and also frequently used in food products. While some strains have been studied extensively, physiological variability among isolates of the species found in healthy humans or their diet is largely unexplored. The aim of this study was to characterize the diversity of carbohydrate utilization capabilities of human isolates and food-derived strains of *L. rhamnosus* in relation to their niche of isolation and genotype. We investigated the genotypic and phenotypic diversity of 25 out of 65 *L. rhamnosus* strains from various niches, mainly human feces and fermented dairy products. Genetic fingerprinting of the strains by amplified fragment length polymorphism (AFLP) identified 11 distinct subgroups at 70% similarity and suggested niche enrichment within particular genetic clades. High-resolution carbohydrate utilization profiling (OmniLog) identified 14 carbon sources that could be used by all of the strains tested for growth, while the utilization of 58 carbon sources differed significantly between strains, enabling the stratification of *L. rhamnosus* strains into three metabolic clusters that partially correlate with the genotypic clades but appear uncorrelated with the strain's origin of isolation. Draft genome sequences of 8 strains were generated and employed in a gene-trait matching (GTM) analysis together with the publicly available genomes of *L. rhamnosus* GG (ATCC 53103) and HN001 for several carbohydrates that were distinct for the different metabolic clusters: L-rhamnose, cellobiose, L-sorbose, and  $\alpha$ -methyl-D-glucoside. From the analysis, candidate genes were identified that correlate with L-sorbose and  $\alpha$ -methyl-D-glucoside utilization, and the proposed function of these genes could be confirmed by heterologous expression in a strain lacking the genes. This study expands our insight into the phenotypic and genotypic diversity of the species *L. rhamnosus* and explores the relationships between specific carbohydrate utilization capacities and genotype and/or niche adaptation of this species.

Strains of a specific bacterial species can display a remarkable degree of phenotypic and genotypic diversity, allowing them to survive in a variety of habitats and/or under a variety of stress conditions. A microorganism's ability to adapt to environmental changes relies on its capacity to acquire and use the available nutrient resources and to counteract and overcome externally exerted physicochemical challenges. The processes of genome evolution, gene acquisition, and gene loss occur at a relatively long time scale and play a prominent role in long-term environmental adaptation of bacteria. The evolution of gene content and its chromosomal organization is stimulated by differences in environmentally selective conditions, such as nutrient availability, antimicrobial activity, or diverse stress conditions exerted by nonoptimal temperature, pH, or osmotic pressure (1). The plasticity in the genetic repertoire is essential for adaptation to specific environmental habitats and, therefore, reflects niche-specific adaptation.

To study the diversity of bacterial species, high-throughput methods for genotypic and phenotypic analysis are increasingly used. These methods, which include amplified fragment length polymorphism (AFLP), restriction fragment length polymorphism (RFLP), multilocus sequence typing (MLST), OmniLog (Biolog) phenotyping (2) infrared spectroscopy, cell mass spectrometry, and more recently genome sequencing, are recognized not only for their high-throughput nature but also for their level of reliability and standardization (3). The development of efficient microbial genomics tools provides novel avenues to effectively evaluate strain diversity and allows for the identification of novel gene functions.

Because of their industrial relevance in a variety of food fer-

mentations as well as their potential interaction with human and animal hosts, lactic acid bacteria (LAB) are an important group of microorganisms. LAB belong to the low-G+C-content Gram-positive bacteria that share the capacity to ferment different carbohydrates into lactic acid. Testifying to the role of phenotyping in industrial fermentation, dairy strains of the paradigm LAB species *Lactococcus lactis* were found to be diverse in their metabolic capacity, which is reflected in their flavor-forming properties (57, 58). These properties are of significant relevance to their application in food fermentations, such as in cheese production. Genotype-phenotype correlation studies contributed to the discovery of new industrial properties for several LAB species, including *Lactobacillus plantarum* (4), *Lactobacillus casei* (5), and *Streptococcus thermophilus* (6).

In human-associated niches, LAB can contribute to the meta-

Received 15 March 2015 Accepted 28 May 2015

Accepted manuscript posted online 5 June 2015

Citation Ceapa C, Lambert J, van Limpt K, Wels M, Smokvina T, Knol J, Kleerebezem M. 2015. Correlation of *Lactobacillus rhamnosus* genotypes and carbohydrate utilization signatures determined by phenotype profiling. *Appl Environ Microbiol* 81:5458–5470. doi:10.1128/AEM.00851-15.

Editor: C. A. Elkins

Address correspondence to Corina Ceapa, corinaceapa@uchicago.edu.

Supplemental material for this article may be found at <http://dx.doi.org/10.1128/AEM.00851-15>.

Copyright © 2015, American Society for Microbiology. All Rights Reserved. doi:10.1128/AEM.00851-15

bolic capacities of the resident microbial ecosystem (7). Moreover, they can interact with the host's mucosal tissues and the immune system (8). Genotypic and phenotypic high-throughput analyses targeted several *Lactobacillus* species (4, 5). Combining phenotypic profiling and strain-specific genetic information has proven to be an effective method for the assignment of so-far-unknown functions to specific genetic loci that are important for industrial traits or the interaction with the host (4). For instance, screening of 14 *L. plantarum* strains for their capacity to adhere to mannose, and correlating this analysis with their genotypes, led to the identification of the gene encoding the mannose-specific adhesin (Msa) in this species (9). Adhesion of *L. plantarum* to mannose residues is thought to be relevant for their capacity to adhere to mucosal epithelial cells that commonly display mannose conjugation moieties on their surface (10), which was proposed to provide a competitive exclusion mechanism that could prevent the mannose-specific recognition of mucosal tissue by FimH-expressing pathogenic *Escherichia coli* cells, thereby preventing their pathogenic potential. Although the protective role of *L. plantarum* in competitive exclusion has yet to be proven, experiments that employed an Msa-deficient mutant and its Msa-expressing parental strain showed that only the wild-type strain effectively induced the expression of the antimicrobial pancreatitis-associated protein (PAP) gene in the intestinal mucosa (11). These results illustrated a possible (dual) role for mannose-specific adhesion in the induction of the host's innate immunity responses, illustrating the importance of identifying these genotype-phenotype relationships in relevant strains.

*Lactobacillus rhamnosus* is a LAB species that colonizes diverse environmental habitats, including dairy and plant materials, as well as the mammalian gastrointestinal tract. *L. rhamnosus* is a species of interest for industry, especially for its potential health-promoting and industrial properties (e.g., cheese ripening and lactate production). The proposed health-promoting properties of specific strains of *L. rhamnosus* have led to their application in products that are marketed as probiotics. While the most extensively studied probiotic is strain *L. rhamnosus* GG (ATCC 53103), other strains of the species, such as HN001 (12) and GR-1 (13), have also been studied for their probiotic potential. To date, it remains unclear to what extent *L. rhamnosus* strains share these properties or if they are specific for a particular strain, highlighting the necessity of determining the diversity within the strains of the species and the identification of potentially strain-specific genes and functions that are responsible for the observed health-promoting effects.

Several genome sequences of *L. rhamnosus* probiotic strains have been determined to date, including *L. rhamnosus* GG (ATCC 53103) (14) and HN001 (NCBI BioProject identifier [ID] 29219). Genomes of strains isolated from industrial fermentations have also been sequenced, including the cheese production isolate *L. rhamnosus* LC705 (15) and the beer spoilage isolate *L. rhamnosus* ATCC 8530 (16), and the genomes sequenced also include an environmental soil isolate, *L. rhamnosus* CASL (17).

The *L. rhamnosus* genomes are predicted to carry a large number of carbohydrate transport and utilization genes that display substantial variations among strains (18). As some niches display unique carbohydrate compositions, variability in carbohydrate utilization capacity is likely to reflect an important aspect of niche-specific adaptation. To better understand the diversity and niche adaptation of strains belonging to the species *L. rhamnosus*, we

analyzed the carbon utilization capacities of 25 strains that were originally isolated from different niches, using the OmniLog Phenotype MicroArray platform. We correlated the carbohydrate utilization profiles with genotyping data obtained for these strains (AFLP) and low-pass genomic sequences, which enabled the identification of candidate genes that could be responsible for the transport and metabolism of specific carbon sources, which were validated by genetic complementation of genes encoding proteins/enzymes for L-sorbose and  $\alpha$ -methyl-glucoside utilization.

## MATERIALS AND METHODS

**Bacterial strains and growth conditions.** For the purpose of this study, 65 *Lactobacillus rhamnosus* strains from various niches of isolation were obtained from Danone Nutricia Research (Palaiseau, France, and Utrecht, The Netherlands) (Table 1). As a reference, several publicly available *L. rhamnosus* strains were included (strains ATCC 53103 [GG] and HN001), and six representative strains of the "closely related" species *L. casei* were added as an outgroup in the genetic fingerprinting by AFLP (Fig. 1). The strains were routinely cultured in de Man-Rogosa-Sharp (MRS) broth or on MRS agar plates (Oxio), under anaerobic conditions at 37°C. Strains were stored at -80°C in MRS medium containing 20% glycerol. Where appropriate, media were supplemented with 10  $\mu\text{g} \cdot \text{ml}^{-1}$  erythromycin.

**Genomic DNA isolation.** Total DNA was extracted from 10 ml of cultures harvested in the mid-log phase (optical density at 600 nm [OD<sub>600</sub>] of 0.5 to 1) using a previously described method (19). In short, cells were lysed by a freezing-thawing step followed by incubation with lysis buffer, TES [*N*-tris(hydroxymethyl)methyl-2-aminoethanesulfonic acid] containing 1,330 U/ml mutanolysin and 40 mg/ml lysozyme, for 1 h at 37°C and bead beating in 20% SDS in TE buffer (10 mM Tris HCl, 1 mM EDTA, pH 8.0). The DNA was recovered by phenol chloroform extraction followed by isopropanol precipitation, washing in 70% ice cold ethanol, drying, and dissolution in 100  $\mu\text{l}$  TE. DNA yield and purity were assessed by measurement of absorbance at 260 nm and at 280 nm (19).

***L. rhamnosus* genomic fingerprinting by AFLP.** Amplified fragment length polymorphism (AFLP) genotyping (20) was used to classify the 65 isolates of *L. rhamnosus* included in this study. To this end, total DNA of the *L. rhamnosus* strains was isolated using an above-described procedure (19) and was digested with EcoRI and TaqI. Specific adapters (Table 2) were ligated to the digested DNA and were used to selectively amplify EcoRI-TaqI fragments using primers extended with G and C selective nucleotides in the combinations E01/T11 and E01/T13 (Table 2). Amplification products were separated according to their length using an ABI Past 3130XL genetic analyzer (Applied Biosystems, Foster City, CA, USA), and fragments containing the fluorescently labeled primer (6-carboxy-fluorescein [FAM]) were visualized using gel-view representations. Band position and intensity were recorded for each strain, creating a strain-specific AFLP profile. The resulting AFLP patterns were band-normalized and subjected to a band pattern recognition procedure using the Gene Mapper 4 software (Applied Biosystems, Foster City, CA, USA). Bands occurring at the same gel position, with an accepted variation of 0.15%, were considered shared between different profiles. Normalized patterns that encompassed fragments of 40 to 580 bp were imported into BioNumerics 4.61 (Applied Maths, Kortrijk, Belgium), and similarities between AFLP-fingerprint profiles were calculated using Dice correlation and UPGMA (unweighted pair group method using average linkages) clustering for the construction of AFLP-based strain dendrograms. Replicate AFLP analyses of strain Lr90 (cluster 5) were used to determine the discriminative threshold of the AFLP analyses at 70%. The type strain ATCC 7469 was obtained from both the Danone and Numico culture collections and appears twice in the analysis.

**Low-pass genome sequencing.** Low-pass genomic sequences of 8 *L. rhamnosus* strains were determined (Table 1). Strains from different AFLP clusters intended to cover the diversity of the species were selected. Total DNA of these strains was prepared as described above (19). Draft genome sequences were obtained (GATC Biotech, Germany) by Roche 454 FLX

TABLE 1 *L. rhamnosus* strains used in this study and availability of genome, coverage and number of contigs, origin, and phenotype data availability<sup>a</sup>

Strain ID	Origin	Biolog	No. of genes	No. of contigs	Coverage (fold)	BioProject ID	BioSample ID
Lr003	Probiotic product	No					
Lr004	Unknown	No					
Lr009	Vegetable drink	Yes					
Lr010	Cheese	No					
Lr026	Unknown	Yes					
<b>Lr032</b>	<b>Human baby feces</b>	<b>Yes</b>	<b>3,357</b>	<b>1,104</b>	<b>6.5</b>	SUB583571	SAMN03196659
Lr035	Human feces	No					
Lr037	Fermented milk	Yes					
Lr040	Unknown	Yes					
<b>Lr044</b>	<b>Cheese</b>	<b>Yes</b>	<b>3,344</b>	<b>2,050</b>	<b>10.5</b>	SUB583571	SAMN03196658
Lr047	Cheese	Yes					
<b>Lr053 (NCIMB 8608)</b>	<b>Unknown</b>	<b>Yes</b>	<b>3,260</b>	<b>1,209</b>	<b>11.2</b>	SUB583571	SAMN03196655
<b>Lr055 (HN001)</b>	<b>Cheese</b>	<b>Yes</b>	<b>2,811</b>	<b>94</b>	<b>30</b>	PRJNA55109	SAMN02469790
Lr064	Unknown	Yes					
<b>Lr071</b>	<b>Human feces</b>	<b>Yes</b>	<b>3,289</b>	<b>690</b>	<b>7.9</b>	SUB583571	SAMN03196657
Lr072	Human feces	No					
<b>Lr073</b>	<b>Human feces</b>	<b>Yes</b>	<b>3,280</b>	<b>808</b>	<b>7.7</b>	SUB583571	SAMN03196656
Lr074	Human feces	Yes					
Lr075	Human feces	Yes					
Lr076	Human feces	No					
Lr077	Human feces	No					
Lr078	Human feces	No					
Lr079	Human feces	No					
Lr080	Human feces	No					
Lr081	Human feces	No					
Lr082	Human feces	No					
Lr083	Human feces	No					
Lr084	Human feces	No					
Lr085	Human feces	No					
Lr086	Human feces	No					
Lr088	Human feces	No					
Lr089	Human feces	No					
Lr090	Human feces	No					
Lr091	Human feces	No					
Lr092	Human feces	No					
Lr093	Human feces	No					
Lr094	Human feces	No					
Lr095	Human feces	No					
Lr096	Human feces	No					
Lr097	Human feces	No					
Lr098	Human feces	No					
Lr099	Human feces	No					
Lr100	Human feces	No					
Lr101	Human feces	No					
Lr102	Human feces	No					
Lr103	Human feces	No					
Lr104	Human feces	No					
Lr105	Human feces	No					
Lr106	Human feces	No					
Lr107	Dairy	No					
<b>Lr108</b>	<b>Human baby feces</b>	<b>Yes</b>	<b>2,649</b>	<b>161</b>	<b>12.5</b>	SUB583571	SAMN03196652
Lr109	Human baby feces	No					
Lr110	Goat feces	Yes					
Lr111	Goat feces	No					
Lr132	Unknown	No					
Lr133	Human vagina	No					
Lr134	Saliva	Yes					
Lr135	Vegetable probiotic drink	Yes					
Lr136	Soy sauce	Yes					
Lr137	Cheese	Yes					
<b>Lr138</b>	<b>Cheese</b>	<b>Yes</b>	<b>3,085</b>	<b>585</b>	<b>8.5</b>	SUB583571	SAMN03196654
Lr139	Cheese	Yes					
<b>Lr140</b>	<b>Fermented milk</b>	<b>Yes</b>	<b>3,401</b>	<b>703</b>	<b>8.2</b>	SUB583571	SAMN03196660
Lr141	Fermented milk	Yes					
<b>Lr142 (LGG)</b>	<b>Healthy human intestine</b>	<b>Yes</b>	<b>2,985</b>	<b>1</b>	<b>30</b>	PRJNA59313	SAMEA2272375

<sup>a</sup> Boldface indicates strains whose genome was included in the GTM analysis.

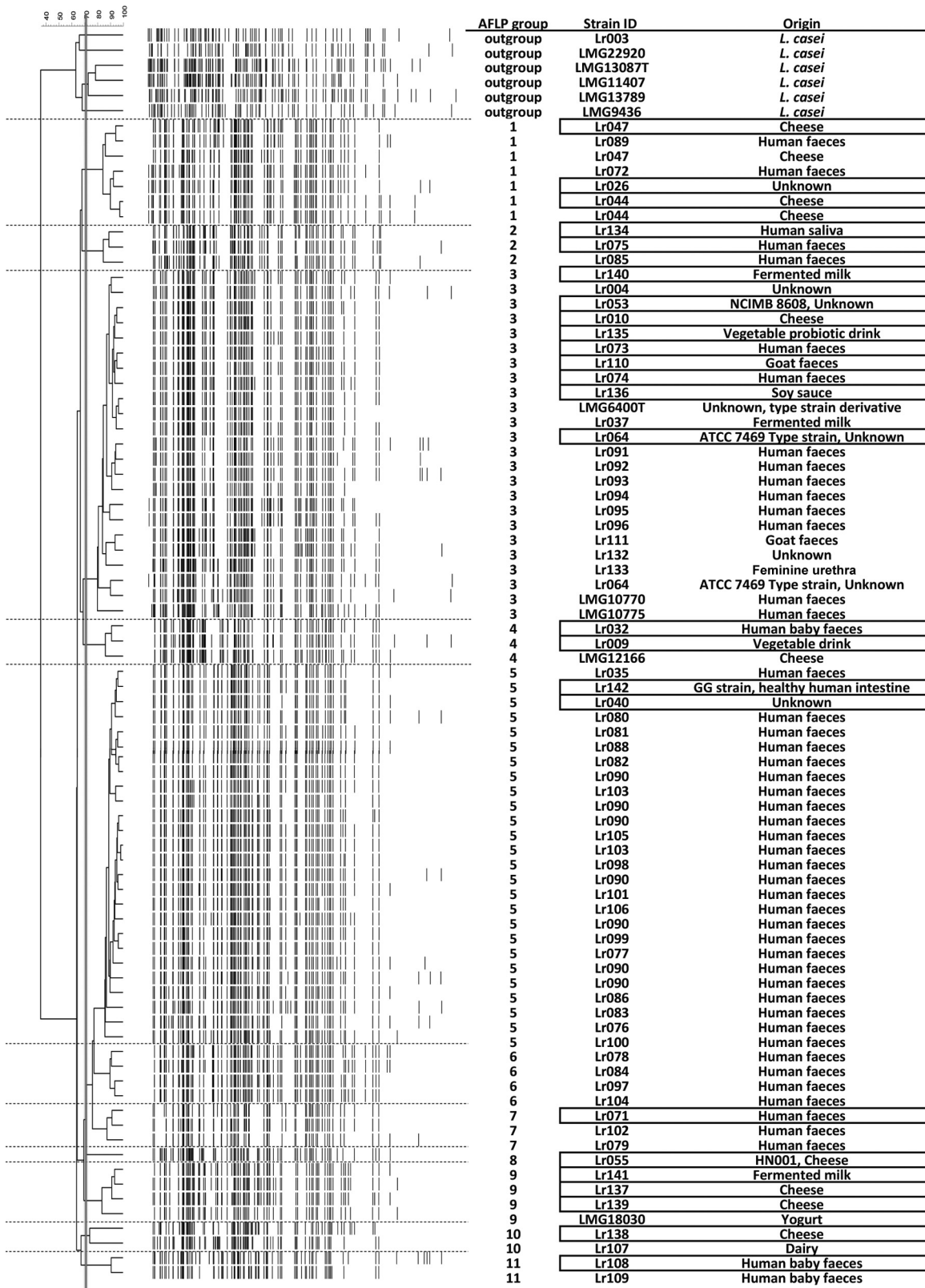


FIG 1 Dendrogram based on the analysis of AFLP patterns of the primer combination E01-T13 with visualization of the banding patterns. The strain origin and clustering (at a similarity level of 70%) are represented in the columns. The 25 strains selected for OmniLog growth experiments are highlighted with boxes.

TABLE 2 Plasmids and primers used in this study

Plasmid or primer	Relevant feature(s) or sequence (5' to 3')	Reference
<b>Plasmids</b>		
pIL253	Ery <sup>r</sup> ; cloning vector; medium copy number in Gram-positive bacteria	1
pSOR253	Ery <sup>r</sup> ; pIL253 derivative containing the L-sorbose candidate operon	This study
pAMG253	Ery <sup>r</sup> ; pIL253 derivative containing the $\alpha$ -methyl-D-glucoside candidate operon	This study
<b>Primers</b>		
Primer E01	GACTGCGTACCAATTCA	
Primer T11	GTTTCTTATGAGTCCTGACCGAA	
Primer T13	GTTTCTTATGAGTCCTGACCGAG	
Adapter (5' to 3') EcoRI	CTCGTAGACTGCGTACC	
Adapter (3' to 5') EcoRI	CTGACGCATGGTTAA	
Adapter (5' to 3') TaqI	GACGATGAGTCCTGAC	
Adapter (3' to 5') TaqI	TACTCAGGACTGGC	
$\alpha$ MetDglu_F	CCGGATCGGGAAACTGAT	
$\alpha$ MetDglu_R	CGAACACAATCTTGCCGTTT	
LSor_F	CAATTCTTCTCTGCATCGC	
LSor_R	AGTATCCATATTGGCATCCG	

Titanium sequencing with average read lengths of 450 bp (Table 1 shows sequencing statistics per strain), while the genome sequences of the reference strains *L. rhamnosus* ATCC 53103 and HN001 were downloaded from the public NCBI Genomes database. The two public domain genome sequences (14) were reannotated in the same way as the newly sequenced genomes and were included in the strain-specific orthologous gene matrix construction (see below). Raw sequence data were assembled into contigs using standard settings of Newbler 2.6 software. Genomic data were subjected to a complete *de novo* RAST pipeline for open reading frame (ORF) prediction using limited-overlap (maximum of 100 bp) allowance for detection of predicted ORFs and, in case of a larger overlap, discarding the smaller of the two ORFs. Gene function annotation for the identified ORFs was performed with the web-based automatic annotator RAST using standard settings (21). Identification of orthologous groups (OGs) of genes shared between the novel low-pass draft genomes and the newly annotated publicly available genomes was performed using locally installed OrthoMCL version 5 (22), containing 150 bacterial reference genomes from the NCBI Genomes database (23). As a result, a gene matrix of all OGs and their presence/absence profile was created. The matrix was used as a basis for the gene-trait matching (GTM) approach.

**PM tests.** Phenotype microarray (PM) analyses were performed to determine the phenotype diversity among 25 selected *L. rhamnosus* strains that represent the different AFLP clusters (Fig. 1) and were isolated from diverse ecological niches (Table 1). Most AFLP clusters are represented, except cluster 6, which appears highly similar to clusters 7 and 8. The carbohydrate utilization profile of these 25 strains was analyzed using the Phenotype MicroArray (OmniLog) 96-well plates PM1 and PM2A (Biolog Inc., Hayward, CA, USA), which determine utilization and growth on 192 different carbon sources, using the provider's protocols. In short, bacteria were precultured in MRS medium and transferred to PM plates in duplicate. Plates were incubated at 37°C in an OmniLog reader (Biolog) for 48 h, and quantitative responses were recorded automatically every 15 min by a charge-coupled device (CCD) camera. Readouts were stored using OmniLog file management software (version 12.0; Technopath). The complete list of compounds assayed by PM1 and PM2A can be obtained at [http://www.biolog.com/pdf/pm\\_lit/PM1-PM10.pdf](http://www.biolog.com/pdf/pm_lit/PM1-PM10.pdf). PM technology employs tetrazolium violet reduction as a reporter of active metabolism (24), where the reduction of the dye causes the formation of a purple color that is recorded by a charge-coupled device camera in time, thus providing quantitative and kinetic information on metabolic activity. Dye reduction is directly correlated with the quantity of NADH produced by the bacteria when degrading a single carbon source. To assess reproducibility, three independent experiments were performed using

strain *L. rhamnosus* GG (ATCC 53103), revealing an identical carbohydrate growth pattern and only very minor variation in the growth kinetics (data not shown).

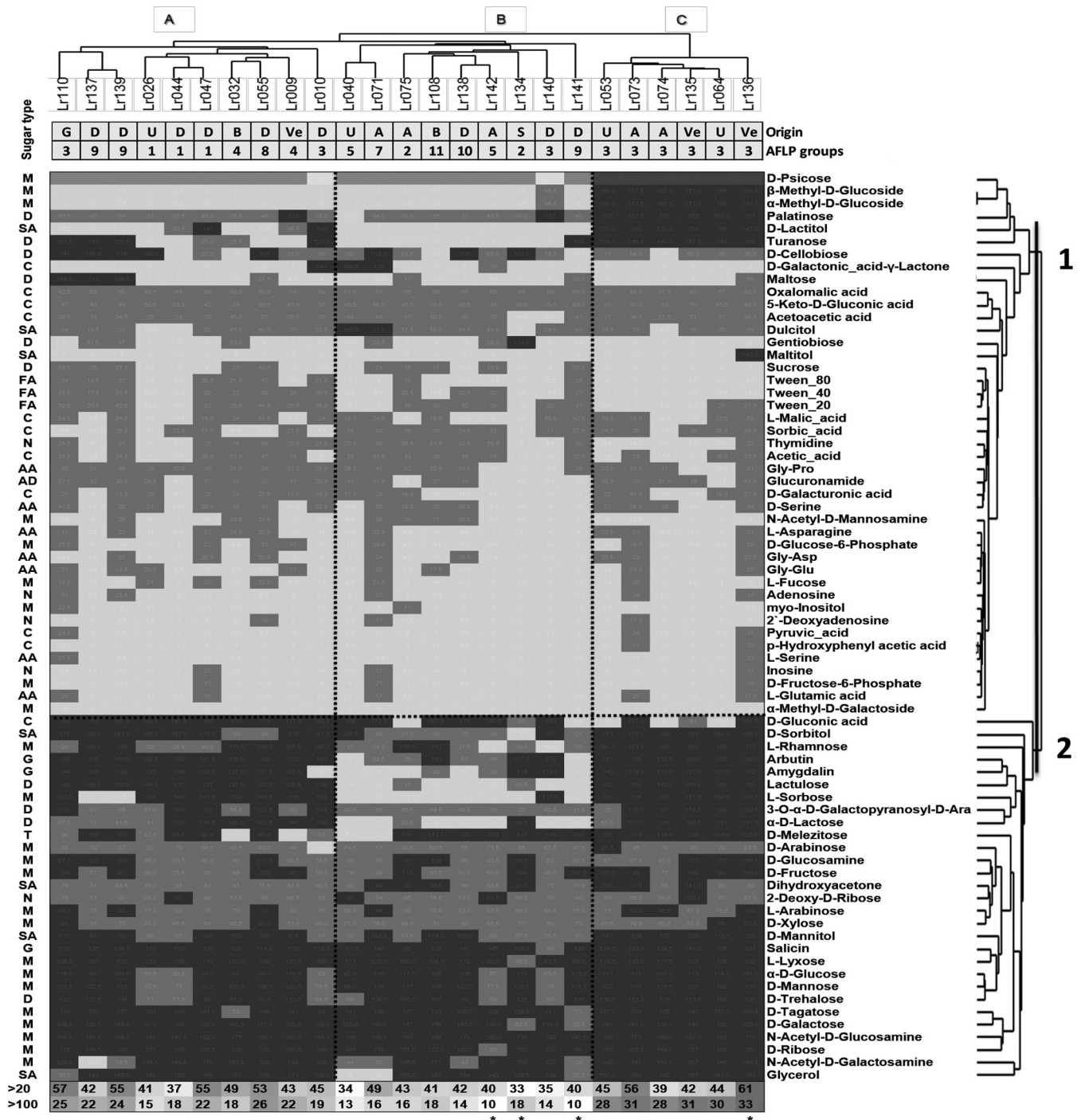
Phenotype microarray results are expressed as average height of reaction (or kinetic curve) as a proxy for overall cell growth over a period of 48 h, after subtraction of the background represented by culture medium, with values ranging from 0 to 198. All the raw PM data are presented in Fig. S3 in the supplemental material.

**Correlation analyses within the AFLP, origin, and OmniLog data sets.** Raw data were processed using PAST software (25). To detect potential correlations among AFLP, OmniLog groups, and niche of isolation, these variables were incorporated in multivariate matrices, and underlying structures within and between the groups of strains were explored using clustering analyses in the PAST software suite (25). Clustering was performed based on Euclidian distances calculated with raw data using UPGMA.

For the determination of relationships between nominal variables (origin of isolation information and presence in an AFLP clade), a contingency table was created (26), containing frequency distribution of the strains (see Table S2 in the supplemental material). Strains of unknown origin were treated as a separate group. The significance of association between the nominal variables was quantified using two methods. The Monte Carlo randomization test relies on repeated random sampling to estimate how likely it is for a certain event to happen randomly and gives a *P* value that is significant (the event is not random) at values lower than 0.05 (27). Cramer's *V* is a chi-square-based measure of variable independence (28), giving a value between 0 (statistical independence) and 1 (associated variables). Both calculations were performed in the PAST software suite (25).

Sørensen-Dice similarity indices provide a way to test statistically whether there is a significant similarity between two or more groups of numerical sampling units. In our case, the sampling units are represented by the growth profiles for carbohydrates for a single strain. Similarity values higher than 0.95 representing a *P* value of <0.05 are significant. Values of the Sørensen-Dice similarity matrix for each of the strain OmniLog endpoint measurement profiles against profiles of all other strains are given in Table S1 in the supplemental material. The similarity data are the basis for the strain clustering in Fig. 2.

**Identification of candidate genes involved in the metabolism of OmniLog carbon sources by gene-trait matching.** Candidate genes potentially involved in carbon source utilization were identified by *in silico* GTM using genomic information from 10 *L. rhamnosus* strains and the OmniLog data sets (Table 1; strains in boldface). We selected the thresh-



**FIG 2** Heat map and clustering of *L. rhamnosus* strains and their substrates based on growth endpoint data (Phenotype MicroArray Biolog data) for carbohydrates in clusters 1 and 2. The shading of the heat map refers to the growth level: black, >100 OU; dark gray, 100 to 20 OU; light gray, <20 OU. The two-way clustering of the strains and substrates was created in PAST using the Euclidian distance matrix based on UPGMA clustering. Origins of isolates and AFLP grouping of the strains are displayed above the strain's dendrogram. The origin abbreviations are as follows: A, adult feces; B, baby feces; G, goat feces; D, dairy; S, saliva; Va, vagina; Ve, vegetable; U, unknown. The type of carbohydrate is displayed to the left of the heat map: monosaccharide (M), disaccharide (D), trisaccharide (T), glycoside (G), fatty acid (FA), nucleoside (N), sugar alcohol (SA), carboxylic acid (C), amino acids (AA), and amide (AD). The bottom rows display the numbers of carbohydrates above 20 OU and above 100 OU for each *L. rhamnosus* strain, and the highest and lowest values are marked with an asterisk.

old 100 OmniLog units (OU) to differentiate between growth and no growth, in agreement with previous reports (29). Using this threshold, the OmniLog raw data matrix was transformed into a positive/negative growth matrix. For data sets with a maximum growth value lower than

100, the threshold was set at half the maximum growth. To assess the significant cooccurrence of *L. rhamnosus* OGs with each of the growth data sets, a mathematical equation was used that focuses on the number of occurrences where the gene presence correlates with the phenotype

(either positively or negatively) but also takes into account the number of occurrences in which such correlation is absent. The higher the outcome of the equation, the higher the probability that the gene correlates with the utilization of a certain carbohydrate. The equation used was  $S = (\text{Pos} + 1) \times (\text{Neg} + 1) / (\text{Mis} + 1)$ . In the above equation, the abbreviations are as follows: *S* represents a final score used for gene classification, *Pos* represents the number of strains where the OG is present and that display the phenotype, *Neg* represents the number of strains where the OG is absent and that do not display the phenotype, and *Mis* represents the number of strains in which the presence and absence of the OG and phenotype are inconsistent (mismatching). While the highest score recognizes a positive gene-phenotype correlation, the lowest score indicates a negative correlation. The equation is dependent on the size and distribution of the phenotype and therefore was used for balanced phenotypes with higher than 25% representation of either the negative or positive groups.

A list of the highest-scoring candidate genes and their RAST-based annotation is provided in the supplemental material (for *L*-sorbose, see Table S3 in the supplemental material; for  $\alpha$ -methyl-D-glucoside, see Table S4 in the supplemental material; for *L*-rhamnose, see Table S5 in the supplemental material; for cellobiose, see Table S6 in the supplemental material). The differential use of cellobiose by strains of *L. rhamnosus* (Fig. 2) was considered a positive control of the analysis, since the results included predicted phosphotransferase system (PTS) genes for cellobiose utilization, which were experimentally proven to be used in cellobiose utilization (14, 30). The genes identified by the equation as the most likely candidates for the utilization of particular carbohydrate sources were further analyzed manually, where preference was given to genes that are located in operons of which multiple genes were identified to have the same correlation score.

**Construction of gene expression mutants.** Gene expression mutants were constructed using the medium-copy-number expression vector pIL253 (31). Both operons targeted by genetic engineering are absent from the genome of *L. rhamnosus* ATCC 53103, which could therefore be employed as the heterologous expression host. The operons that were identified to be potentially involved in utilization of *L*-sorbose (5 kb) and  $\alpha$ -methyl-D-glucoside (6 kb) were amplified by PCR using genomic DNA isolated from strain Lr136 as a template, in combination with the primer pairs designed to amplify each of the target loci (Table 2), using KOD long-range polymerase (Novagen, Darmstadt, Germany), according to the instructions of the manufacturer. The PCR products were purified from an agarose gel and cloned in *Sca*I-digested pIL253 vector. Ligation mixtures were transformed directly into *L. rhamnosus* ATCC 53103 by electroporation (32), and transformants were selected on MRS medium containing  $10 \mu\text{g} \cdot \text{ml}^{-1}$  erythromycin. Two single colonies were selected based on stable antibiotic resistance (erythromycin) and designated M12 and M13, harboring the plasmids comprising the cloned operons predicted to be involved in *L*-sorbose and  $\alpha$ -methyl-D-glucoside, respectively. The growth phenotype of the M12 and M13 expression derivatives of *L. rhamnosus* ATCC 53103 was evaluated on in-house-prepared MRS medium without a carbon source supplemented with *L*-sorbose or  $\alpha$ -methyl-D-glucoside as sole carbon sources, respectively, using the parental *L. rhamnosus* ATCC 53103 as a negative control.

**Nucleotide sequence accession numbers.** The genome sequences for the eight newly sequenced *Lactobacillus rhamnosus* strains were deposited in the GenBank database with the following accession numbers: [JUIH00000000](#) (*Lactobacillus rhamnosus* Lr108), [JUII00000000](#) (*Lactobacillus rhamnosus* Lr138), [JUIJ00000000](#) (*Lactobacillus rhamnosus* Lr053), [JUIK00000000](#) (*Lactobacillus rhamnosus* Lr073), [JUIL00000000](#) (*Lactobacillus rhamnosus* Lr071), [JUIM00000000](#) (*Lactobacillus rhamnosus* Lr044), [JUIO00000000](#) (*Lactobacillus rhamnosus* Lr032), and [JUIP00000000](#) (*Lactobacillus rhamnosus* Lr140) (Table 1). The versions described in this paper have the numbers “XXXXX01000000,” where XXXX represents the first four letters of each accession number.

## RESULTS

**Strain collection, niches of strain isolation, and genotype fingerprinting.** Sixty-five *L. rhamnosus* strains included in the study were obtained from the Danone Nutricia Research Culture Collection (Palaiseau, France, and Utrecht, The Netherlands) and were originally isolated from various environments (Fig. 1). For some strains, including the type strain (ATCC 7469), no information is available in the public domain with respect to their origin of isolation. For the others, most isolates are of human origin (feces [baby and adult], oral cavity, and reproductive system) or from dairy (cheese or fermented milk) products, but there were also some strains derived from other isolation sources, i.e., two strains from animal (goat) feces and two from fermented plant material (soy sauce and fermented vegetable drink). In addition, reference strains of the species *L. rhamnosus* (strains ATCC 53103 and HN001 and the type strain ATCC 7469) and six representatives of the closely related species *L. casei* (Fig. 1) were included in the AFLP analysis.

AFLP provides a high-throughput method for high-resolution genomic fingerprinting (20) that has been employed frequently for the classification of strains of various species, including lactobacilli (2, 33). AFLP classification of 65 *L. rhamnosus* strains enabled the distinction of 11 genotypic groups at a similarity cutoff of 70%, with all *L. casei* strains constituting an outgroup, since they belong to a closely related but genetically distinct species. The overall similarity level of the species by AFLP profiling was estimated at 60% (this work and reference 34), with an intragroup diversity of up to 20%. These data illustrate the remarkable diversity of the *L. rhamnosus* species, as AFLP diversity for some more-specialized species can be lower than this threshold: *Lactobacillus reuteri*, 8% (35); *Lactobacillus delbrueckii*, 15% (36); *Lactobacillus acidophilus*, 22% (37). At the same time, the AFLP classification is similar to other species that reside in various environmental niches such as *L. plantarum* (38) that also encompass highly diverse clusters of strains. Notably, the two primer pairs used in AFLP profiling generated corresponding results in terms of clade classification of strains (Fig. 1; see also Fig. S1 in the supplemental material), as both analyses identified subgroups that contained the same strains, demonstrating the accuracy of the genetic classification.

The AFLP clades encompassed variable numbers of strains. Among the 11 clades, eight were comprised of only a few strains (between one and five strains were clustered within clades 2, 4, 6, 7, 8, 9, 10, and 11). Clade 1 was a medium-sized group containing 7 strains, while most strains (75%) were classified in clades 3 and 5, which contain 23 and 26 *L. rhamnosus* strains, respectively. The strains with publicly available genomic sequences were classified into distinct AFLP clusters: type strain ATCC 7469 belonged to clade 3, strain ATCC 53103 (Lr064) belonged to clade 5, and HN001 was the only strain in clade 8.

**Establishing correlations between origin and AFLP grouping.** As genetic classification for several LAB species indicated that strains with similar origins sometimes cluster together (35, 39, 40), we were interested to see if cataloguing *L. rhamnosus* strains on the basis of the genetic fingerprinting by AFLP could significantly reflect their origin of isolation. Most strains (83%) were classified into one of the large AFLP clusters, 3 and 5 (see Fig. S2 in the supplemental material). The strains in the large cluster 3 and smaller clusters 2 and 4 have a highly variable origin of isolation, supporting the idea that there is a subgroup of *L. rhamnosus*

strains that frequently migrate between different environments (18). Other AFLP clusters appeared to display niche enrichment, illustrated by the observation that the large AFLP cluster 5, as well as the smaller clusters 6 and 7, contained mainly human strains, including 23 fecal strains (56% of all fecal isolates) and only a single strain of unknown origin, while the two infant strains cluster together in a separate group (cluster 11) that appeared genetically more similar to a dairy isolate (cluster 10) than to adult feces isolates. Fermented product isolates were predominately classified into four genotype clusters: cluster 1, which also contained two fecal strains, and clusters 8, 9, and 10, which exclusively captured strains of dairy origin. Fecal and dairy strains could not be clearly separated by the clustering.

As for the statistical significance of this enrichment, the Monte Carlo simulation analysis showed that the origin distribution is not random ( $P = 0.0126$ ) (see Table S2 in the supplemental material). The Cramer test value ( $P = 0.53$ ) supported a tendency toward niche enrichment within AFLP grouping. We found a stronger correlation between adult fecal isolates and AFLP clades 3, 5, 6, and 7 and between dairy strains and clade 9 (see Table S2 in the supplemental material).

**Strain-specific carbon source utilization profiling.** A variety of carbon sources that have been associated with various habitats were included in the OmniLog growth experiments. Cluster analysis identified 3 main clusters of carbon sources based on the variation of their utilization by *L. rhamnosus* strains (see Fig. S3 in the supplemental material). Cluster 1 grouped 34 carbon sources with a lower average growth of  $23 \pm 19$  OUs and contains many amino acids, carboxylic acids, fatty acids, and nucleosides but also several monosaccharides and disaccharides. Some substrates were used by all strains in this group with an intermediate efficiency (20 to 100 OU; oxomalic acid and 5-keto-D-gluconic acid) and a low efficiency ( $<20$  OU; 2-dioxiadensine and  $\alpha$ -methyl-D-galactoside). Carbohydrate cluster 2 contained 38 carbon sources with a relatively high average level of growth ( $114 \pm 28$  OU) per strain and growth values of up to 196 OU for some strains. The majority of monosaccharides, all glycosides (i.e., sugar bound to another functional group via a glycosidic bond; salicin, arbutin, and amygdalin), several disaccharides, sugar alcohols, and a single trisaccharide (melezitose) were grouped within cluster 2. Thirteen cluster 2 carbohydrate substrates were utilized by all strains, i.e., D-ribose, N-acetyl-D-glucosamine, D-galactose, D-tagatose, D-trehalose, D-mannose,  $\alpha$ -D-glucose, L-lyxose, salicin, D-mannitol, L-arabinose, 2-deoxy-D-ribose, and 3-O- $\alpha$ -D-galactopyranosyl-D-arabinose. Carbohydrate cluster 3 contained 120 carbon sources that were concluded not to be utilized by any of the strains (raw values maximally reached 30 OU) and were therefore excluded for strain classifications (see below).

From the carbon sources that were utilized differentially among the 25 isolates, several can be utilized by the majority (21 of the 25 strains) of the strains tested, e.g., D-melezitose, D-gluconic acid, D-cellobiose, dulcitol, arbutin, glycerol, acetoacetic acid, L-rhamnose, and D-sorbitol. In contrast, several carbohydrates could be utilized by only a few strains (fewer than four), e.g., D-galactonic acid- $\gamma$ -lactone, 2-deoxy-D-ribose, dihydroxyacetone, L-arabinose, and maltose. Notably, dulcitol, maltitol, and gentiobiose appeared to be utilized by only a single strain.

The *L. rhamnosus* strains could be categorized into three metabolic groups (MGs) based on their utilization of the 72 carbon sources (clusters 1 and 2) and using a similarity coefficient cutoff

(Sørensen-Dice) of 95%. The three MGs (designated MG-A, -B, and -C [Fig. 2]) contained 10, 9, and 6 strains, respectively. UPGMA clustering allowed the identification of specific carbohydrates that were most stringently separating the MGs (Fig. 2). The utilization of D-lactitol, D-psicose,  $\alpha$ -methyl-D-glucoside,  $\beta$ -methyl-D-glucoside, turanose, and palatinose correlated well with the majority of strains clustered in MG-C, whereas only a few strains from MG-A and MG-B could utilize these carbohydrates. Many strains of group MG-B were not able to utilize L-sorbose, D-sorbitol, 3-O- $\alpha$ -DL-galactopyranosyl-D-arabinose, lactulose, amygdalin, arbutin, and L-rhamnose, whereas most strains in MG-A and MG-C could utilize these carbon sources.

There appeared to be no obvious correlation of the MG classification of the strains with their niche of isolation, illustrated by the fact that each of the MGs encompassed isolates from all niches (Fig. 2). However, strains classified in the same AFLP cluster tended to be classified in the same MG (Fig. 2) and strains of AFLP groups 1, 3, and 4 shared very similar carbohydrate utilization profiles between them. For example, AFLP group 1 strains formed a tight subgroup within MG-A. Similarly, a subgroup of AFLP cluster 3 (6 of the 9 members) captured all MG-C clustered strains, with the individual strains sharing 95.4% of their carbohydrate utilization profile. Similarly, MG-B captured all the members of AFLP clusters 2, 5, 7, and 10, whereas MG-A contained all strains of AFLP clusters 1, 4, and 8.

Unlike strains from other groups, MG-A strains could efficiently use plant-derived maltose, arbutin, and artificially created gluconic acid and lactulose but could not use the natural sugar alcohols dulcitol, gentiobiose, and maltitol, and only some strains utilized lactose. MG-B was characterized by a decreased ability to grow in lactose (dairy-derived carbohydrate) or sorbitol, sorbose, rhamnose, arbutin, lactulose, ribose, salicin, and lyxose, many of which are plant material-associated carbohydrates. MG-C strains were capable of growth on lactose, N-acetyl-D-galactosamine, and mannose, as well as palatinose, psicose,  $\alpha$ - and  $\beta$ -methyl-D-glucosides, turanose, maltitol, arbutin, mannitol, trehalose, and tagatose but not cellobiose, D-galactonic acid- $\gamma$ -lactone, and gluconic acid. In particular, strains belonging to MG-C utilized a multitude of carbohydrates with high efficiency, whereas strains of MG-A could utilize fewer carbohydrates, many of them at an average efficiency. MG-B appeared to be highly restricted in the number of carbohydrates that it could grow on, with relatively low efficiency (Fig. 2).

**Identification of genes responsible for specific carbohydrate utilization.** The availability of combined genomic and phenotypic information provided the opportunity to identify genes responsible for specific carbohydrate utilization. The genomic data of two publicly available genome sequences of *L. rhamnosus* strains ATCC 53103 (a single circular genome sequence) and HN001 (draft genome assembled to 94 contigs), in combination with eight draft genome sequences of strains included in this study (average coverage, 11.4-fold; assembled into 161 to 2,050 contigs) (Table 1), were employed in a GTM approach. The strains selected for low-pass genomic sequencing aimed to retain diversity in both AFLP grouping (which constituted the primary criterion) and origin of isolation. Both the public genomic sequences and the novel draft genomes were *de novo* annotated using automatic open reading frame detection and annotation of protein functions. Orthologous gene detection for the 10 genome data sets generated 6,476 OGs, of which



**TABLE 3** Genes selected for expression by complementation after the GTM analysis for  $\alpha$ -methyl-D-glucoside and L-sorbose

Contig length (bp)	Potential carbohydrate	No. of strains in which present ( $n = 10$ )	Annotation	Present in other public genomes	Complementation strain
4.451	L-Sorbose	6	Phosphopentomutase (EC 5.4.2.7)	No	GG
4.451	L-Sorbose	6	Benzoate MFS <sup>a</sup> transporter BenK	No	GG
4.451	L-Sorbose	6	Uridine phosphorylase (EC 2.4.2.3)	No	GG
4.451	L-Sorbose	6	Sugar diacid utilization regulator	No	GG
4.451	L-Sorbose	5	Benzoate MFS transporter BenK	No	GG
6.145	$\alpha$ -Methyl-D-glucoside	4	Transporter	LC705_02793	GG
6.145	$\alpha$ -Methyl-D-glucoside	4	Hypothetical protein	LC705_02792	GG
6.145	$\alpha$ -Methyl-D-glucoside	4	ATPase	LC705_02791	GG

<sup>a</sup> MFS, major facilitator superfamily.

1,793 OGs (~27%) were shared among all strains, whereas 1,982 OGs (~30%) were present in only a single strain.

*L. rhamnosus* candidate genes that may be involved in the metabolism of specific carbohydrates were identified *in silico* by correlating OmniLog carbohydrate utilization data with the strain-specific genomic data, using the OG matrix constructed by GTM. To this end, carbohydrates were selected that displayed a balanced distribution of utilizing (>100 OU) and nonutilizing (<100 OU) strains among the 10 strains to maximize the likelihood of identifying credible candidate genes associated with the phenotype. The carbon sources that fulfilled this requirement were cellobiose, D-gluconic acid, D-melezitose,  $\alpha$ -methyl-D-glucoside, 3-O- $\beta$ -D-galactopyranosyl-D-arabinose, L-sorbose, dulcitol, and D-galactonic acid- $\gamma$ -lactone. These carbohydrates also encompass those that enabled discrimination among MG-A, -B, and -C, e.g., L-sorbose and D-gluconic acid are typically not utilized by strains of MG-B, whereas  $\alpha$ -methyl-D-glucoside is exclusively utilized by strains belonging to MG-C. Notably, GTM that employed the strain-specific cellobiose utilization capacity led to the identification of a PTS operon that was annotated to be involved in cellobiose utilization (see Table S6 in the supplemental material), supporting the reliability of the approach. Nevertheless, this analysis did not identify any of the genes encoding the other 3 PTSs that are annotated as cellobiose PTS, which may be related to the apparent redundancy of this transport function, which disables the unambiguous and consistent identification of a single locus. The differential utilization of L-sorbose was used in a GTM approach and correlated strongly with a genomic region of 6 kb encompassing genes encoding a putative PTS transporter (see Table S3 in the supplemental material). Similarly,  $\alpha$ -methyl-D-glucoside utilization correlated strongly with two candidate genetic regions of 5 and 19 kb in length, respectively, encoding a transporter and ATPase, and two PTSs and a carbohydrate hydrolase, respectively (Table 3; see Table S4 in the supplemental material). To validate the postulated roles of the identified genomic regions in the utilization of these carbohydrates, the identified genetic regions were amplified from Lr136 and cloned into pIL253, resulting in pSOR253 (containing the 6-kb locus linked to sorbose utilization) and pAMG253 (containing the 5-kb locus linked to  $\alpha$ -methyl-D-glucoside utilization), respectively.

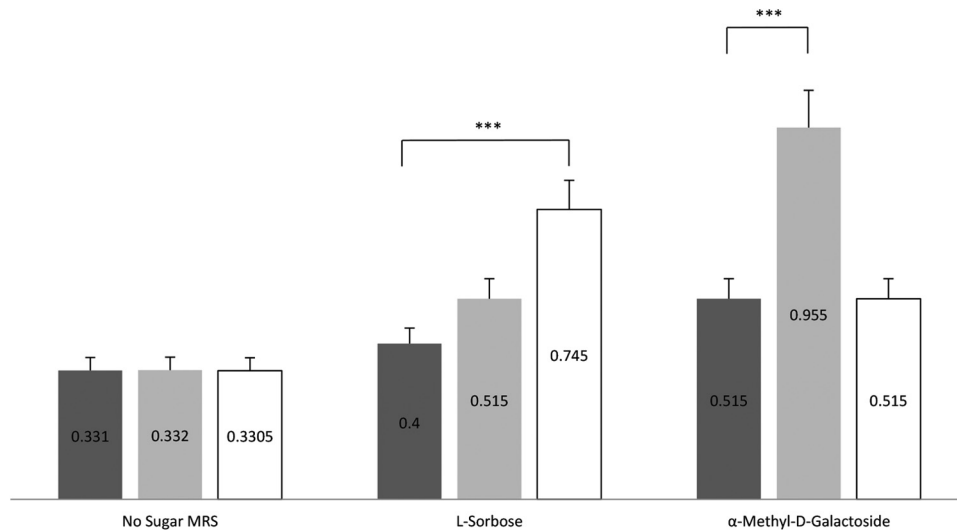
These constructs enabled the heterologous expression of these genes in *L. rhamnosus* ATCC 53103, a strain that lacks these OGs (Table 1). The resulting carbohydrate utilization by the pSOR253- and pAMG253-harboring derivatives of *L. rhamnosus* ATCC 53103 was evaluated, using in-house-prepared MRS medium supplemented with the relevant carbon source, revealing that in con-

trast to the strain transformed with the empty pIL253 vector, the pSOR253- and pAMG253-harboring derivatives were able to grow on sorbose and  $\alpha$ -methyl-D-glucoside, respectively (Fig. 3). These results confirm the role of these genes in the postulated phenotypes and illustrate the value of the GTM approach to identify gene functions.

## DISCUSSION

With respect to carbohydrate substrate utilization and metabolism, strains isolated from diverse niches can vary greatly, probably reflecting adaptation to the niche-specific conditions. The importance of understanding at a molecular level the functional diversity within individual bacterial species that are considered for probiotic applications has only recently been appreciated (41). Bearing in mind their potential use in food and/or probiotic applications, we studied the differences within strains of the species *Lactobacillus rhamnosus* using both food and commensal isolates and employing two high-throughput methods for their classification, AFLP-based genotyping and OmniLog profiling for metabolic phenotyping. The outcome of this study is in agreement with previous studies showing a wide phenotypic diversity among strains of *L. rhamnosus* and the identification of several genetically distinct groups (18, 34, 42).

Carbon source utilization profiling is an important tool for the analysis of diversity and phenotyping of bacterial genera and species, as illustrated by the routine use of the API strips or OmniLog plates. For the species *L. rhamnosus*, 53 isolates were analyzed previously using an API50-based classification approach, revealing that strain *L. rhamnosus* GG (ATCC 53103) could be distinguished from cheese isolates on the basis of its carbohydrate utilization capacity (42), a conclusion that we confirmed in the present study. In addition, the present study supports the high versatility and adaptability to multiple niches in the species *Lactobacillus rhamnosus*. There are substantial differences in carbohydrate availability in the niches in which *L. rhamnosus* can be found. The dairy environment is rich in lactose and contains also some free oligosaccharides (composed of mannose, fucose, and sialic acids), while plant-associated environments can be rich in sucrose, trehalose, maltose, cellobiose, raffinose, starch, inulin, and fructosans. Notably, the intestinal tract of mammals can contain, next to a variety of diet-derived carbohydrates, also substantial quantities of host-derived carbohydrates such as fucose, hexosamines, mannose, and galactose (43). The metabolic grouping of the *L. rhamnosus* strains presented here classifies strains belonging to MG-A as generalists, albeit that these strains also display some specific limitations such as the use of glucosides, turanose, and



**FIG 3** Growth of the wild-type GG strain (dark gray) and expression strains M12 (transformed with pSOR253 containing the L-sorbose operon) (white) and M17 (transformed with pAMG253 containing the  $\alpha$ -methyl-D-glucoside operon) (light gray) on L-sorbose and  $\alpha$ -methyl-D-glucoside in batch cultures. The numbers represent OD values per milliliter for the cultures in sugar-free MRS medium supplemented with either L-sorbose or  $\alpha$ -methyl-D-glucoside. \*\*\*, a *P* value of  $<0.001$  using Student's *t* test.

psicose. These plant derivatives are encountered only in particular niches, which might explain this adaptation. The inability to utilize lactose among strains clustered in MG-B indicates that these strains cannot grow efficiently in dairy environments. However, there were dairy isolates clustered in MG-B that were derived from cheese production, where they are used for their proteolytic capacity, which plays a role during late stages of cheese ripening. During these stages of cheese ripening, the lactose is already depleted from the cheese matrix, suggesting that the MG-B cheese strains could represent non-starter culture isolates that play a role in late-stage cheese ripening only. The strains clustered within MG-C appear to be adapted to use both dairy- and plant-derived carbohydrates.

Interestingly, the dairy-derived *L. rhamnosus* isolates could consistently utilize gluconic acid. This carbohydrate is not a natural component of milk but is extensively used in the dairy industry to stabilize fermented dairy products and to retain calcium (44). Therefore, it is tempting to speculate that the strains that were adapted to this environment for many generations would have acquired the capacity to utilize this carbohydrate and could be an illustration of the adaptive abilities of these bacteria.

In a recent comparative genomics study of 100 *L. rhamnosus* strains, Douillard et al. (18) found mainly two distinctive genophenotypes, of which one appeared to be specialized for stable nutrient-rich niches and the other contained generalists that are adaptable and could potentially reside in multiple niches. The two genophenotypes are detectable in our metabolic classification, but the generalist group is divided into a subgroup with a higher efficiency and larger range of possible carbohydrates (MG-C) and a subgroup that can utilize many carbohydrates for growth, albeit with an intermediate efficiency (MG-A). The more-specialized MG-B strains seem primed toward fewer carbohydrates that they can use at a relatively high efficiency. This is exemplified by the metabolism of currently marketed probiotic strains such as *L. rhamnosus* GG (ATCC 53103), which was originally isolated from the intestine, clustered in MG-B, and appears to be a metabolic

specialist adapted to nutrient-rich environments. It can grow efficiently ( $>100$  OU) on 10 carbohydrates, glucose, galactose, glycerol, salicin, *N*-acetyl-D-glucosamine, *N*-acetyl-D-galactosamine, D-melezitose, tagatose, L-lyxose, and D-gluconic acid, while its overall carbohydrate capacity encompasses 40 substrates. It is long since *L. rhamnosus* GG was transferred from the complex and highly variable intestinal niche to the relatively constant and nutrient-rich industrial production environment, which may be selective for metabolic simplification (45). This tendency was supported for *L. rhamnosus* GG, for which PCR analyses of six commercial probiotic products confirmed that four products contain a derivative of the original strain that lacked major DNA segments (46). This simplifying tendency may have contributed to the inability of this strain to utilize a relatively low number of carbohydrates with good efficiency. In contrast, the probiotic *L. rhamnosus* HN001, which was isolated from cheese and clustered in MG-A, displays efficient growth on 26 carbon sources and is capable of utilizing 53 different carbon sources, overall. Therefore, this isolate displays efficient growth on a wide range of carbohydrates, which is in disagreement with the notion that dairy isolates of lactobacilli would display more-specialized carbohydrate utilization patterns than those of intestinal isolates (6, 47). This may also imply that the HN001 strain was relatively recently introduced into the cheese matrix and has not yet lost its metabolic flexibility as a consequence of the consistent exposure to the constant dairy environment.

These, and several other observations, suggest that the origin of isolation is only remotely informative about a strain's metabolism, which may especially be valid for species that frequently migrate between various niches, which has been proposed for various lactobacilli, including *L. plantarum* (48), *L. casei* (47), *Lactobacillus sakei* (49), and *L. rhamnosus* (50).

Phenotyping was complemented in our study by genetic fingerprinting by AFLP, which created a separate, genetic classification of the strains. Some AFLP clusters were apparently enriched with strains from particular environmental niches. Isolates from

fecal material collectively comprise 56% of all strains and belong to AFLP clusters 5, 6, and 7. Dairy isolates, amounting to 20% of all strains, appeared to divide into two AFLP clusters, possibly reflecting two main lineages of strains that evolved separately in the dairy environment. All other clusters had a mixed composition in terms of niche of isolation.

Similar observations concerning the correlation of genetic fingerprinting methods and the strain's niche of isolation were made for several other lactic acid bacterial species (51–53). Nevertheless, in some cases, the genetic fingerprints obtained appeared to correlate relatively well with the niche-specific fitness. For example, the species *L. reuteri* was divided into ecotypes that display high host specificity and can colonize either pigs, rodents, or humans, which underlines the idea that some microbes from the vertebrate gut are not promiscuous but have diversified into host-adapted lineages, probably involving a long-term evolutionary process (35). Notably, the ecotype specification of *L. reuteri* isolates displayed an excellent correlation with the genetic stratification of the strains by multilocus sequence typing (MLST) or AFLP fingerprinting. Similarly, random amplified polymorphic DNA-PCR (RAPD-PCR) enabled the separation of a genetic subgroup of strains of the species *L. plantarum* that were exclusively isolated from dairy environments, whereas the other subgroups contained strains of various origins of isolation (38).

In conclusion, it is not trivial to extract predictive information related to the degree of niche-specific adaptations in microbial strains of a *Lactobacillus* species that could potentially transit from one niche to another with relatively high frequency, such as *L. rhamnosus*. Consequently, the niche of isolation is not very informative with respect to the strain's niche-specific fitness, and neither genetic fingerprinting nor high-resolution metabolic profiling provides a highly reliable approach to define a strain's degree of adaptation to any particular niche. To enable the determination of niche-specific fitness among strains of such species, other approaches are needed and may require comparative and strain-specific *in situ* fitness determination in the respective niches that the strains can inhabit.

In the statistical analysis, we concluded that AFLP-based clade distribution of the strains displayed a tendency for enrichment of niche of isolation within certain genetic clusters, whereas the OmniLog classifications into MGs appeared to cocluster with specific AFLP clusters. These observations illustrate how two fundamentally different classification approaches provide a consistent discrimination of subgroups of strains. However, the MG classification of the strains did not correlate with the niche of isolation of the strains, which is somewhat unexpected in view of the correlations detected between AFLP clusters and the niche of isolation. This may be largely due to the relative noisiness of the latter correlation, which is characterized by several confounding strains in each of the AFLP clusters. Moreover, true niche specialization is a complex phenotype, which is unlikely to be represented by single genetic markers and probably involves both multiple discriminating genes that may be genetically unlinked and strain-specific and divergent gene regulatory patterns. Although it is among the highest-resolution methodologies for genetic fingerprinting, AFLP patterns still represent a crude way of genetic typing (2), and whole-genome sequencing and comparative genomics provide a genetic strain typing methodology of substantially higher resolution and may therefore be more appropriate for the niche-fitness correlation analyses.

The value of genome sequencing and genotype-phenotype correlation analyses was illustrated here to identify candidate genes that were associated with discriminative carbohydrate utilization capacities among the *L. rhamnosus* strains included in this study. Although the genome sequence information employed for this purpose was incomplete due to the low-pass quality of the sequence information generated, this analysis still accurately identified the gene clusters involved in L-sorbose and  $\alpha$ -methyl-D-glucoside utilization. Genetic engineering enabled the confirmation of the functions of these genes by heterologous expression in an *L. rhamnosus* strain that lacked these genetic functions and thereby gained the capacity to utilize the corresponding carbohydrate source for growth. The GTM algorithm used is considerably simpler than the advanced modeling method employed in some tools available, such as the PhenoLink module (54), which may enable the identification of candidate genes involved in the utilization of other carbohydrates that did not allow gene identification in the GTM method employed here.

Further gene function identification is required to advance our understanding of bacterial diversity and evolution in relation to niche-specific adaptation functions. Progress in nucleotide sequencing technologies and comparative genomics (39), experimental evolution approaches (55), and competitive intestinal passage models (56) can provide access to the enormous genetic diversity of bacteria at an unprecedented level of resolution. Such approaches can aid in determining and understanding relative niche fitness levels of different strains of a species at a molecular level, providing critical information to truly assess niche adaptation.

The method described in the current work can be useful in the identification of genes linked to other phenotypic characteristics of *Lactobacillus* diversity of interest to science and the consumer, as long as the phenotype is sufficiently varying in the strains studied. Some of these phenotypes for the *Lactobacillus* genus could encompass (i) metabolic traits such as proteinase and peptidase activities, acidification capacity, and vitamin or short-chain fatty acid production and host interaction parameters such as (ii) cytokine production by immune cells, (iii) the production of antibacterial peptides, (iv) bile salt hydrolase activity, or (v) intestinal colonization capacity. When expanding these concepts to bacteria in general, genotype-phenotype association studies can have a much wider applicability, for instance, to identify the genetic basis of pathogenicity traits such as toxin or adhesin production, host cell invasion capacity, and/or disease progression or outcome.

Taken together, our findings indicate that high-resolution phenotyping and genotyping enable the detection of distinct genetic clades as well as metabolic groups among the strains of the species *L. rhamnosus*. Both high-throughput methods reveal similar relationships between the strains, illustrating the idea that high-resolution phenotyping and genotyping can provide a means to stratify strains of a species into genetically and functionally distinct groups. This stratification of strains was employed to guide the selection of strains for genomic sequencing, eventually enabling the identification of function-related genetic markers that, in the example presented, relate to carbohydrate utilization but could also include other relevant phenotypic traits (11). Such identified genetic markers for a specific phenotype could be used to accelerate the selection of strains with specific characteristics that are relevant for their industrial applications.

## ACKNOWLEDGMENTS

Regarding conflicts of interest, C.C., J.L., T.S., K.V.L., and J.K. work for Danone Nutricia Research, part of the Danone Group. Danone is selling products that contain lactobacilli. Danone Nutricia Research financed part of this study.

This work was supported by a Marie Curie International Training Mobility Network grant (Cross-Talk, grant agreement 21553-2) awarded to C.C. and Nutricia Research, Utrecht, The Netherlands.

We thank Marie Christine Degivry for help in managing the data from the Danone Culture Collection.

## REFERENCES

- Barracough T, Balbi K, Ellis R. 2012. Evolving concepts of bacterial species. *Evol Biol* 39:148–157. <http://dx.doi.org/10.1007/s11692-012-9181-8>.
- Di Cagno R, Minervini G, Sgarbi E, Lazzi C, Bernini V, Neviani E, Gobetti M. 2010. Comparison of phenotypic (Biolog System) and genotypic (random amplified polymorphic DNA-polymerase chain reaction, RAPD-PCR, and amplified fragment length polymorphism, AFLP) methods for typing *Lactobacillus plantarum* isolates from raw vegetables and fruits. *Int J Food Microbiol* 143:246–253. <http://dx.doi.org/10.1016/j.ijfoodmicro.2010.08.018>.
- Moraes P, Perin L, Junior AS, Nero LA. 2013. Comparison of phenotypic and molecular tests to identify lactic acid bacteria. *Braz J Microbiol* 44:109–112. <http://dx.doi.org/10.1590/S1517-83822013000100015>.
- Siezen RRJ, Tzeneva VV, Castioni A, Wels M, Phan HTK, Rademaker JLW, Starrenburg MJC, Kleerebezem M, Molenaar D, van Hylckama Vlieg JET. 2010. Phenotypic and genomic diversity of *Lactobacillus plantarum* strains isolated from various environmental niches. *Environ Microbiol* 12:758–773. <http://dx.doi.org/10.1111/j.1462-2920.2009.02119.x>.
- Smokvina T, Wels M, Polka J, Chervaux C, Brisse S, Boekhorst J, van Hylckama Vlieg JET, Siezen RJ. 2013. *Lactobacillus paracasei* comparative genomics: towards species pan-genome definition and exploitation of diversity. *PLoS One* 8:e68731. <http://dx.doi.org/10.1371/journal.pone.0068731>.
- Goh Y, Goin C, O'Flaherty S, Altermann E, Hutkins R. 2011. Specialized adaptation of a lactic acid bacterium to the milk environment: the comparative genomics of *Streptococcus thermophilus* LMD-9. *Microb Cell Fact* 10:S22. <http://dx.doi.org/10.1186/1475-2859-10-S1-S22>.
- Den Besten G, van Eunen K, Groen AK, Venema K, Reijngoud D-J, Bakker BM. 2013. The role of short-chain fatty acids in the interplay between diet, gut microbiota, and host energy metabolism. *J Lipid Res* 54:2325–2340. <http://dx.doi.org/10.1194/jlr.R036012>.
- Tsai Y-T, Cheng P-C, Pan T-M. 2012. The immunomodulatory effects of lactic acid bacteria for improving immune functions and benefits. *Appl Microbiol Biotechnol* 96:853–862. <http://dx.doi.org/10.1007/s00253-012-4407-3>.
- Pretzer G, Snel J, Molenaar D, Wiersma A, Bron PA, Lambert J, de Vos WM, van der Meer R, Smits MA, Kleerebezem M. 2005. Biodiversity-based identification and functional characterization of the mannose-specific adhesin of *Lactobacillus plantarum*. *J Bacteriol* 187:6128–6136. <http://dx.doi.org/10.1128/JB.187.17.6128-6136.2005>.
- Adlerberth I, Ahrné S, Johansson ML, Molin G, Hanson LÅ, Wold AE. 1996. A mannose-specific adherence mechanism in *Lactobacillus plantarum* conferring binding to the human colonic cell line HT-29. *Appl Environ Microbiol* 62:2244–2251.
- Gross G, van der Meulen J, Snel J, van der Meer R, Kleerebezem M, Niewold TA, Hulst MM, Smits MA. 2008. Mannose-specific interaction of *Lactobacillus plantarum* with porcine jejunal epithelium. *FEMS Immunol Med Microbiol* 54:215–223. <http://dx.doi.org/10.1111/j.1574-695X.2008.00466.x>.
- Wickens K, Black P, Stanley TV, Mitchell E, Barthow C, Fitzharris P, Purdie G, Crane J. 2012. A protective effect of *Lactobacillus rhamnosus* HN001 against eczema in the first 2 years of life persists to age 4 years. *Clin Exp Allergy* 42:1071–1079. <http://dx.doi.org/10.1111/j.1365-2222.2012.03975.x>.
- Hummelen R, Changalucha J, Butamanya NL, Cook A, Habbema JDF, Reid G. 2010. *Lactobacillus rhamnosus* GR-1 and *L. reuteri* RC-14 to prevent or cure bacterial vaginosis among women with HIV. *Int J Gynecol Obstet* 111:245–248. <http://dx.doi.org/10.1016/j.ijgo.2010.07.008>.
- Morita H, Toh H, Oshima K, Murakami M, Taylor TD, Igimi S, Hattori M. 2009. Complete genome sequence of the probiotic *Lactobacillus rhamnosus* ATCC 53103. *J Bacteriol* 191:7630–7631. <http://dx.doi.org/10.1128/JB.01287-09>.
- Kankainen M, Paulin L, Tynkkynen S, von Ossowski I, Reunanen J, Partanen P, Satokari R, Vesterlund S, Hendrickx APA, Lebeer S, De Keersmaecker SCJ, Vanderleyden J, Hämäläinen T, Laukkanen S, Salovuori N, Ritari J, Alatalo E, Korpela R, Mattila-Sandholm T, Lassig A, Hatakka K, Kinnunen KT, Karjalainen H, Saxelin M, Laakso K, Surakka A, Palva A, Salusjärvi T, Auvinen P, de Vos WM. 2009. Comparative genomic analysis of *Lactobacillus rhamnosus* GG reveals pili containing a human-mucus binding protein. *Proc Natl Acad Sci U S A* 106:17193–17198. <http://dx.doi.org/10.1073/pnas.0908876106>.
- Pittet V, Ewen E, Bushell BR, Ziola B. 2012. Genome sequence of *Lactobacillus rhamnosus* ATCC 8530. *J Bacteriol* 194:726. <http://dx.doi.org/10.1128/JB.06430-11>.
- Yu B, Su F, Wang L, Zhao B, Qin J, Ma C, Xu P, Ma Y. 2011. Genome sequence of *Lactobacillus rhamnosus* strain CASL, an efficient L-lactic acid producer from cheap substrate cassava. *J Bacteriol* 193:7013–7014. <http://dx.doi.org/10.1128/JB.06285-11>.
- Douillard FP, Ribbera A, Kant R, Pietilä TE, Järvinen HM, Messing M, Randazzo CL, Paulin L, Laine J, Ritari J, Caggia C, Läheteinen T, Brouns SJJ, Satokari R, von Ossowski I, Reunanen J, Palva A, de Vos WM. 2013. Comparative genomic and functional analysis of 100 *Lactobacillus rhamnosus* strains and their comparison with strain GG. *PLoS Genet* 9:e1003683. <http://dx.doi.org/10.1371/journal.pgen.1003683>.
- Gevers D, Huys G, Swings J. 2001. Applicability of rep-PCR fingerprinting for identification of *Lactobacillus* species. *FEMS Microbiol Lett* 205:31–36. <http://dx.doi.org/10.1111/j.1574-6968.2001.tb10921.x>.
- Vos P, Hogers R, Bleeker M, Reijans M, van de Lee T, Hornes M, Frijters A, Pot J, Peleman J, Kuiper M. 1995. AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Res* 23:4407–4414. <http://dx.doi.org/10.1093/nar/23.21.4407>.
- Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formisano K, Gerdes S, Glass EM, Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL, Overbeek RA, McNeil LK, Paarmann D, Paczian T, Parrello B, Pusch GD, Reich C, Stevens R, Vassieva O, Vonstein V, Wilke A, Zagnitko O. 2008. The RAST server: rapid annotations using subsystems technology. *BMC Genomics* 9:75. <http://dx.doi.org/10.1186/1471-2164-9-75>.
- Chen F, Mackey AJ, Stoekert CJ, Roos DS. 2006. OrthoMCL-DB: querying a comprehensive multi-species collection of ortholog groups. *Nucleic Acids Res* 34:D363–D368. <http://dx.doi.org/10.1093/nar/gkj123>.
- Pruitt KD, Tatusova T, Maglott DR. 2007. NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res* 35:D61–D65. <http://dx.doi.org/10.1093/nar/gkl842>.
- Bochner BR, Gadzinski P, Panomitros E. 2001. Phenotype microarrays for high-throughput phenotypic testing and assay of gene function. *Genome Res* 11:1246–1255. <http://dx.doi.org/10.1101/gr.186501>.
- Hammer Ø, Harper DAT, Ryan PD. 2001. PAST: paleontological statistics software package for education and data analysis. *Palaeontol Electron* 4:1–9.
- Enke H. 1979. The information in contingency tables. *Biom J* 21:94–95. <http://dx.doi.org/10.1002/bimj.4710210114>.
- Brown RP. 1997. Testing for heterogeneity among phenotypic correlations: a comparison of methods using Monte Carlo simulations. *Genetica* 101:67–74. <http://dx.doi.org/10.1023/A:1018305905597>.
- McHugh ML. 2013. The chi-square test of independence. *Biochem Med (Zagreb)* 23:143–149.
- Tohsato Y, Mori H. 2008. Phenotype profiling of single gene deletion mutants of *E. coli* using Biolog technology. *Genome Inform* 21:42–52.
- Toh H, Oshima K, Nakano A, Takahata M, Murakami M, Takaki T, Nishiyama H, Igimi S, Hattori M, Morita H. 2013. Genomic adaptation of the *Lactobacillus casei* group. *PLoS One* 8:e75073. <http://dx.doi.org/10.1371/journal.pone.0075073>.
- Domingues S, Aires AC, Mohedano ML, López P, Arraiano CM. 2013. A new tool for cloning and gene expression in *Streptococcus pneumoniae*. *Plasmid* 70:247–253. <http://dx.doi.org/10.1016/j.plasmid.2013.05.005>.
- Kim YH, Han KS, Oh S, You S, Kim SH. 2005. Optimization of technical conditions for the transformation of *Lactobacillus acidophilus* strains by electroporation. *J Appl Microbiol* 99:167–174. <http://dx.doi.org/10.1111/j.1365-2672.2005.02563.x>.

33. McLeod A, Nyquist OL, Snipen L, Naterstad K, Axelsson L. 2008. Diversity of *Lactobacillus sakei* strains investigated by phenotypic and genotypic methods. *Syst Appl Microbiol* 31:393–403. <http://dx.doi.org/10.1016/j.syapm.2008.06.002>.
34. Vancanneyt M, Huys G, Lefebvre K, Vankerckhoven V, Goossens H, Swings J. 2006. Intraspecific genotypic characterization of *Lactobacillus rhamnosus* strains intended for probiotic use and isolates of human origin. *Appl Environ Microbiol* 72:5376–5383. <http://dx.doi.org/10.1128/AEM.00091-06>.
35. Oh PL, Benson AK, Peterson DA, Patil PB, Moriyama EN, Roos S, Walter J. 2010. Diversification of the gut symbiont *Lactobacillus reuteri* as a result of host-driven evolution. *ISME J* 4:377–387. <http://dx.doi.org/10.1038/ismej.2009.123>.
36. Tanigawa K, Watanabe K. 2011. Multilocus sequence typing reveals a novel subspeciation of *Lactobacillus delbrueckii*. *Microbiology* 157:727–738. <http://dx.doi.org/10.1099/mic.0.043240-0>.
37. Gancheva A, Pot B, Vanhonacker K, Hoste B, Kersters K. 1999. A polyphasic approach towards the identification of strains belonging to *Lactobacillus acidophilus* and related species. *Syst Appl Microbiol* 22:573–585. [http://dx.doi.org/10.1016/S0723-2020\(99\)80011-3](http://dx.doi.org/10.1016/S0723-2020(99)80011-3).
38. Torriani S, Clementi F, Vancanneyt M, Hoste B, Dellaglio F, Kersters K. 2001. Differentiation of *Lactobacillus plantarum*, *L. pentosus* and *L. paraplantarum* species by RAPD-PCR and AFLP. *Syst Appl Microbiol* 24:554–560. <http://dx.doi.org/10.1078/0723-2020-00071>.
39. Siezen RJ, Bayjanov JR, Felis GE, van der Sijde MR, Starrenburg M, Molenaar D, Wels M, van Hijum SA, van Hylckama Vlieg JET. 2011. Genome-scale diversity and niche adaptation analysis of *Lactococcus lactis* by comparative genome hybridization using multi-strain arrays. *Microb Biotechnol* 4:383–402. <http://dx.doi.org/10.1111/j.1751-7915.2011.00247.x>.
40. Kütahya OE, Starrenburg MJC, Rademaker JLW, Klaassen CHW, van Hylckama Vlieg JET, Smid EJ, Kleerebezem M. 2011. High-resolution amplified fragment length polymorphism typing of *Lactococcus lactis* strains enables identification of genetic markers for subspecies-related phenotypes. *Appl Environ Microbiol* 77:5192–5198. <http://dx.doi.org/10.1128/AEM.00518-11>.
41. Kleerebezem M, Vaughan EE. 2009. Probiotic and gut lactobacilli and bifidobacteria: molecular approaches to study diversity and activity. *Annu Rev Microbiol* 63:269–290. <http://dx.doi.org/10.1146/annurev.micro.091208.073341>.
42. Succi M, Tremonte P, Reale A, Sorrentino E, Grazia L, Pacifico S, Coppola R. 2005. Bile salt and acid tolerance of *Lactobacillus rhamnosus* strains isolated from Parmigiano Reggiano cheese. *FEMS Microbiol Lett* 244:129–137. <http://dx.doi.org/10.1016/j.femsle.2005.01.037>.
43. Flint HJ, Scott KP, Duncan SH, Louis P, Forano E. 2012. Microbial degradation of complex carbohydrates in the gut. *Gut Microbes* 3:289–306. <http://dx.doi.org/10.4161/gmic.19897>.
44. Ramachandran S, Fontanille P, Pandey A, Larroche C. 2006. Gluconic acid: properties, applications and microbial production. *Food Technol Biotechnol* 44:185–195.
45. Cai H, Thompson R, Budinich MF, Broadbent JR, Steele JL. 2009. Genome sequence and comparative genome analysis of *Lactobacillus casei*: insights into their niche-associated evolution. *Genome Biol Evol* 1:239–257. <http://dx.doi.org/10.1093/gbe/evp019>.
46. Sybesma W, Molenaar D, van IJcken W, Venema K, Kort R. 2013. Genome instability in *Lactobacillus rhamnosus* GG. *Appl Environ Microbiol* 79:2233–2239. <http://dx.doi.org/10.1128/AEM.03566-12>.
47. Broadbent JR, Neeno-Eckwall EC, Stahl B, Tandee K, Cai H, Morovic W, Horvath P, Heidenreich J, Perna NT, Barrangou R, Steele JL. 2012. Analysis of the *Lactobacillus casei* supragenome and its influence in species evolution and lifestyle adaptation. *BMC Genomics* 13:533. <http://dx.doi.org/10.1186/1471-2164-13-533>.
48. Siezen RJ, van Hylckama Vlieg JE. 2011. Genomic diversity and versatility of *Lactobacillus plantarum*, a natural metabolic engineer. *Microb Cell Fact* 10(Suppl 1):S3. <http://dx.doi.org/10.1186/1475-2859-10-S1-S3>.
49. Chiamonte F, Anglade P, Baraige F, Gratadoux J-J, Langella P, Champomier-Vergès M-C, Zagorec M. 2010. Analysis of *Lactobacillus sakei* mutants selected after adaptation to the gastrointestinal tracts of axenic mice. *Appl Environ Microbiol* 76:2932–2939. <http://dx.doi.org/10.1128/AEM.02451-09>.
50. Douillard FP, Ribbera A, Järvinen HM, Kant R, Pietilä TE, Randazzo C, Paulin L, Laine K, Caggia C, Von Ossowski I, Satokari R, Salminen S, Palva A, Vos De WM, Laine PK, von Ossowski I, Reunanen J, de Vos WM. 2013. Comparative genomic and functional analysis of *Lactobacillus casei* and *Lactobacillus rhamnosus* strains marketed as probiotics. *Appl Environ Microbiol* 79:1923–1933. <http://dx.doi.org/10.1128/AEM.03467-12>.
51. Bayjanov JR, Starrenburg MJC, van der Sijde MR, Siezen RJ, van Hijum SA. 2013. Genotype-phenotype matching analysis of 38 *Lactococcus lactis* strains using random forest methods. *BMC Microbiol* 13:68. <http://dx.doi.org/10.1186/1471-2180-13-68>.
52. Khatri B, Fielder M, Jones G, Newell W, Abu-Oun M, Wheeler PR. 2013. High throughput phenotypic analysis of *Mycobacterium tuberculosis* and *Mycobacterium bovis* strains' metabolism using biologic phenotype microarrays. *PLoS One* 8:e52673. <http://dx.doi.org/10.1371/journal.pone.0052673>.
53. Bachmann H, Starrenburg MJC, Dijkstra A, Molenaar D, Kleerebezem M, Rademaker JLW, Van Hylckama Vlieg JET. 2009. Regulatory phenotyping reveals important diversity within the species *Lactococcus lactis*. *Appl Environ Microbiol* 75:5687–5694. <http://dx.doi.org/10.1128/AEM.00919-09>.
54. Bayjanov JR, Molenaar D, Tzeneva V, Siezen RJ, van Hijum SA. 2012. PhenoLink—a web-tool for linking phenotype to -omics data for bacteria: application to gene-trait matching for *Lactobacillus plantarum* strains. *BMC Genomics* 13:170. <http://dx.doi.org/10.1186/1471-2164-13-170>.
55. Bachmann H, Starrenburg MJC, Molenaar D, Kleerebezem M, Van Hylckama Vlieg JET. 2012. Microbial domestication signatures of *Lactococcus lactis* can be reproduced by experimental evolution. *Genome Res* 22:115–124. <http://dx.doi.org/10.1101/gr.121285.111>.
56. Van Bokhorst-van de Veen H, van Swam I, Wels M, Bron PA, Kleerebezem M. 2012. Congruent strain specific intestinal persistence of *Lactobacillus plantarum* in an intestine-mimicking in vitro system and in human volunteers. *PLoS One* 7:e44588. <http://dx.doi.org/10.1371/journal.pone.0044588>.
57. Brandsma JB, van de Kraats I, Abee T, Zwietering MH, Meijer WC. 2012. Arginine metabolism in sugar deprived *Lactococcus lactis* enhances survival and cellular activity, while supporting flavour production. *Food Microbiol* 29:27–32. <http://dx.doi.org/10.1016/j.fm.2011.08.012>.
58. Dhaisne A, Guellerin M, Laroute V, Laguerre S, Coccagn-Bousquet M, Le Bourgeois P, Loubiere P. 2013. Genotypic and phenotypic analysis of dairy *Lactococcus lactis* biodiversity in milk: volatile organic compounds as discriminating markers. *Appl Environ Microbiol* 79:4643–4652. <http://dx.doi.org/10.1128/AEM.01018-13>.