

# Functional genomics of *Lactobacillus casei* establishment in the gut

# Hélène Licandro-Seraut<sup>a,b,c</sup>, Hélène Scornec<sup>a</sup>, Thierry Pédron<sup>b,c</sup>, Jean-François Cavin<sup>a</sup>, and Philippe J. Sansonetti<sup>b,c,d,1</sup>

<sup>a</sup>Unité Mixte de Recherche A 02102 Procédés Alimentaires et Microbiologiques, AgroSup Dijon/Université de Bourgogne, 21000 Dijon, France; <sup>b</sup>Unité de Pathogénie Microbienne Moléculaire and <sup>c</sup>Institut National de la Santé et de la Recherche Médicale U786, Institut Pasteur, 75724 Paris Cedex 15, France; and <sup>d</sup>Chaire de Microbiologie et Maladies Infectieuses, Collège de France, 75005 Paris, France

Contributed by Philippe J. Sansonetti, June 25, 2014 (sent for review April 1, 2014)

Although the composition of the gut microbiota and its symbiotic contribution to key host physiological functions are well established, little is known as yet about the bacterial factors that account for this symbiosis. We selected Lactobacillus casei as a model microorganism to proceed to genomewide identification of the functions required for a symbiont to establish colonization in the gut. As a result of our recent development of a transposonmutagenesis tool that overcomes the barrier that had prevented L. casei random mutagenesis, we developed a signature-tagged mutagenesis approach combining whole-genome reverse genetics using a set of tagged transposons and in vivo screening using the rabbit ligated ileal loop model. After sequencing transposon insertion sites in 9,250 random mutants, we assembled a library of 1,110 independent mutants, all disrupted in a different gene, that provides a representative view of the L. casei genome. By determining the relative quantity of each of the 1,110 mutants before and after the in vivo challenge, we identified a core of 47 L. casei genes necessary for its establishment in the gut. They are involved in housekeeping functions, metabolism (sugar, amino acids), cell wall biogenesis, and adaptation to environment. Hence we provide what is, to our knowledge, the first global functional genomics analysis of L. casei symbiosis.

commensalism | Lactic acid bacteria

The pioneering studies that led to the characterization of the gut microbiota were reviewed in 2001 (1). These studies and recent investigations have revealed mutualistic functions (2), including a barrier effect against allogenic microbes (3), fermentation of complex sugars (4, 5), and maturation and homeostasis of the immune system (6). Recent metagenomic studies have revealed an extraordinary diversity of genes constituting the gut microbiome (7), opening the way to correlative studies linking microbiome diversity, homeostasis, and diseases (5, 8, 9).

In parallel, some representative species, i.e., "model symbionts," now are being studied functionally (10). As it was done for pathogens, it is essential to develop the cellular microbiology of symbionts and particularly to identify the genes required for their establishment and persistence in the gut. Transcriptomic profiling identified up-regulated genes linked to metabolic functions, stress responses, and pili synthesis during early colonization (11-13). Comparative genomics among Lactobacilli identified strain-specific candidate genes for extended colonization: In Lactobacillus rhamnosus, persistence was attributed to an spaCBA locus encoding LPXTG-like pilins (14), and in Lactobacillus johnsonii it was attributed to specific glycosyltransferases, a phosphotransfer system, and a protease (15). Otherwise, a functional in vivo screening based on the expression of a genomic library of Bacteroides fragilis identified a locus encoding polysaccharide utilization as essential for stable colonization of murine colonic crypts (16). Alternatively, colonization of germ-free mice with a collection of random mutants of Bacteroides thetaiotaomicron followed by deep sequencing showed that mutants unable to synthesize vitamin B12 were impaired in gut colonization (17).

Lactobacillus spp. pioneer initial gut colonization (18), and they participate in the gut immunological and nutritional symbiosis. Because of our permanent exposure to Lactobacillus spp., and particularly to Lactobacillus casei, as components of dairy products, tools are required to decipher its symbiosis with the gut. Genetic manipulation of lactic bacteria often is problematic because of their natural resistance to numerous antibiotics and the lack of dedicated genetic tools and efficient transformation procedures (19). We recently developed a transposon-mutagenesis tool, the P<sub>junc</sub>-TpaseIS<sub>1223</sub> system (20), which overcomes the barrier to random mutagenesis in Lactobacillus casei. Using this tool, we adapted a signature-tagged mutagenesis (STM) approach (21) to L. casei that combined whole-genome reverse genetics using a set of tagged transposons with an in vivo screening in the rabbit ligated ileal loop model identifying mutants impaired in gut establishment. The term "establishment" qualifies the early steps of colonization explored by this model. After sequencing the 1,110 independent mutants obtained in this study, we identified a core of 47 L. casei genes belonging to five major functional groups that are required for its establishment in the gut.

### Results

**Generation of a Library of** *L. casei* **Tagged Mutants.** To generate a large library of L. *casei* tagged mutants and to proceed to STM, tagged derivatives of the  $P_{junc}$ -TpaseIS<sub>1223</sub> transposable vector were generated using 70 DNA tags previously used for *Salmonella typhimurium* STM (21). For each tag, among the ~5,000 integrants obtained per transformation, clones were selected randomly and assembled in 96-well plates. Thus, a library of 9,250 tagged mutants labeled with 70 different tags was generated. To extend the contribution of STM, we introduced real-time PCR, rather than dot-blot analysis, to allow relative quantification of bacteria in addition to their detection.

### **Significance**

Lactobacillus casei, a food bacterium recognized for its beneficial effects, was selected as a model microorganism to proceed to genomewide identification of the functions required for a symbiont to establish colonization in the gut. We recently have developed a mutagenesis tool that overcomes the barrier that prevented *L. casei* random mutagenesis. After identifying 9,250 mutations, we assembled a library of 1,110 mutants disrupted in different genes and tested them for their ability to colonize an in vivo model, the rabbit ligated ileal loop. With this global functional genomic analysis of *L. casei* symbiosis (the first, to our knowledge), we identified a core of 47 *L. casei* genes necessary for its establishment in the gut.

Author contributions: H.L.-S., T.P., and P.J.S. designed research; H.L.-S., H.S., and T.P. performed research; H.L.-S., H.S., T.P., J.-F.C., and P.J.S. analyzed data; and H.L.-S., T.P., J.-F.C., and P.J.S. wrote the paper.

The authors declare no conflict of interest.

<sup>&</sup>lt;sup>1</sup>To whom correspondence should be addressed. Email: philippe.sansonetti@pasteur.fr.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10. 1073/pnas.1411883111/-/DCSupplemental.

Analysis and Assembly of a Library of *L. casei* Tagged Mutants. Based on an initial screening showing that the intergenic regions contributed very little to gut establishment (SI Text), we focused on mutants in genes. Hence, we sequenced the transposon target in each mutant before in vivo screening (i.e., 9,250 mutants). Among the 8,053 readable sequences (87% of the total), 3,037 (37%) indicated that transposon integration occurred in a gene (2,787 integrations into the chromosome and 250 into the plasmid) (Fig. 1A). In most cases, when several mutations were found in the same gene, the sites of transposon insertion differed (84.5% of insertions in genes occurred in unique sites), and we selected one representative mutant. Mutations in intergenic regions (IGR) tended to occur in palindromic regions surrounding endogenous transposase genes (32.4% of the insertions in IGRs occurred in unique sites). In summary, our library of tagged L. casei tagged mutants 1,096 mutants in distinct chromosomal genes and 14 mutants in distinct plasmid genes. Characterizing and assembling this library considerably reduced the number of mutants to be screened. The mutations appear to be evenly distributed throughout the genome (Fig. 1A). Overall, the proportion of the functional groups in the library is similar to their proportion in the genome (Fig. 1B). However, the proportion of genes encoding translation-related functions [with the COG letter (J)] is lower in the library, likely because of the high number of essential genes related to protein synthesis (22). Because most bacterial genes are organized in operons [1.5 genes per transcription unit predicted for L. casei, according to genome analysis using the Biocyc website (23)], transposon integration can cause a polar effect; thus the location of each gene of interest in a given operon is specified.

Screening of Mutants in the Rabbit Ligated Ileal Loop Model and Identification of Crucial *L. casei* Genes for Establishment in the Gut. The rabbit ligated ileal loop model allowed the screening of a large number of *Shigella* mutants in competitive pools in the gut (24–26), and it explores the ileal conditions in which *L. casei* has been shown to persist in an active physiological state (27).

Twenty-two pools containing 25-70 differently tagged mutants were organized from the library of 1,110 discrete mutants (Fig. 2). The pools then were challenged in rabbit ligated ileal loops  $(5.10^7 \text{ cfu per loop})$  for 16 h (Fig. 2). Each pool was tested twice in different rabbits. The ratio of each mutant to the whole pool was determined using tags for the injected and the recovered pools (Fig. 2). Based on the observed variation of mutant quantities in a pool, we considered a threefold difference in quantity between the injection and recovery pools a relevant threshold (SI Text and Fig. S1). Sixty-nine mutants were selected at this step because their quantity decreased at least threefold in two rabbits. We identified the mutants that were exclusively altered in vivo and did not show a growth defect in vitro by comparing the colony size and the turbidity of cultures at stationary phase in MRS medium of the mutants and the wild-type strain (Table S1). This comparison allowed us to eliminate six mutants (i.e., LSEI 0281, 1278, 1468, 1565, and 1566). To eliminate growth biases linked to pool composition, the resulting 63 mutants were reassembled to create five new pools. Each pool was composed of 12 or 13 mutants being retested and 12 mutants picked randomly from the IGR library. A ratio of one mutant to 70 competitors was maintained. The in vivo challenge and the quantification methods were identical to those used in the first round, except that, to eliminate false positives caused by growth deficiencies on plates, the injected pool was quantified from colonies obtained by plating.

This scheme of selection identified 47 mutants of *L. casei* (4% of the screened genes) that were altered in their capacities of establishment in the gut (Table 1). The mutated genes encompassed various functions, including housekeeping, metabolism, biogenesis of the cell wall, and adaptation to environment. Most



Fig. 1. Analysis of transposon insertions in the genome of the 9250 *L. casei* mutants. (*A*) Genome atlas of transposon insertions. Circle 1: genes on the positive strand in light gray; circle 2: genes on the negative strand in dark gray; circle 3: disrupted genes in blue; circle 4: disrupted IGRs in red; circle 5: GC content; circle 6: GC skew. (*B*) Relative abundance of cluster of orthologous groups (COG) functional categories of genes in the *L. casei* genome and in the disrupted genes of the STM library. Hits found in COG represent 68.3% of the genome and 69% of the disrupted ORFs of the STM library.

**PNAS PLUS** 

MICROBIOLOGY



Fig. 2. Generation of the gene tagged-mutants library and its screening in rabbit ligated ileal loops.

are conserved among the Firmicute phylum. LSEI\_0238, 0242, and 0247 orthologs are only found in the *Lactobacillus* genus. Seven genes, LSEI\_0135, 0156, 0806, 1461, 1790, 2553, and 2601, are specific to *L. casei* or related species such as *L. rhamnosus*, *Lactobacillus paracasei*, or *Lactobacillus zeae*.

Mutant Analysis. Housekeeping functions. Our screening identified several housekeeping genes that are likely to help maintain

physiological parameters during the changes to which bacteria are suddenly exposed in the in vivo environment (Table 1 and Fig. 3): LSEI\_0757 and 1539 (nucleotide metabolism), LSEI\_1615 and 1488 (DNA replication), LSEI\_1668 (transcription), LSEI\_1656 (translation), and LSEI\_1274 (possibly implicated in cell division). We also identified the RecG family helicase RecG (LSEI\_1615) and DinG (LSEI\_1488), primarily involved in DNA replication and possibly in DNA repair (28). Mutations in this category



Fig. 3. Main *L. casei* genes for basal metabolism, environmental adaptation, defense mechanisms, cell surface, transport and their role in *L. casei* establishment in the gut. Genes needed for gut persistence are red letters; genes not needed are in green letters; and genes for which no mutant was found in the library (genes not tested) are in black letters.

## Table 1. L. casei genes required for intestinal establishment

PNAS PNAS

Gene tag	Gene name	Operon prediction*
Functions lost in establishme	nt	
Housekeeping		
Cellular machinery		
LSEI_0757	Flavoprotein NrdI	No
LSEI_1274	ATPase for cell division	LSEI_1274-1279 (+)
LSEI_1488	Rad3-related DNA helicase (DinG)	LSEI_1488-1490 (-)
LSEI_1539	Guanosine polyphosphate pyrophosphohydrolase/synthetase	LSEI_1537-1539 (-)
LSEI_1600	RNA-binding protein	LSEI_1600-1601 (-)
LSEI_1615	ATP-dependent DNA helicase RecG	LSEI_1613-1617 (-)
LSEI_1656	tRNA delta (2)-isopentenyl pyrophosphate transferase	No
LSEI_1668	Transcription elongation factor GreA	LSEI_1668-1669 (-)
Adaptation to stress and	defense mechanisms	
LSEI_1313	Stress membrane GTPase	No
LSEI_1403	Tyrosine recombinase XerC subunit	LSEI_1402-1403 (+)
LSEI_1790	Predicted Mrr-like endonuclease	No
Metabolism		
Carbohydrate metabolis	m, transport and energy production	
LSEI_0174	Phosphoketolase	No
LSEI_0681	Lactose transport regulator	LSEI_0676-0681 (-)
LSEI_2549	L-lactate dehydrogenase (LdhL1)	No
Amino acid metabolism		
LSEI_0479	Homoserine transsuccinylase (MetA)	LSEI_0479-0480 (+)
LSEI_0480	Cysteine synthase (CysK)	LSEI_0479-0480 (+)
LSEI_1810	Dipeptidyl aminopeptidase/acylaminoacyl-peptidase	LSEI_1809-1810 (-)
LSEI_2162	Asparagine synthase (AsnB)	LSEI_2161-2162 (-)
Transport		
LSEI_0242	ABC-type Mn/Zn transport system, ATPase component	LSEI_0240-0243 (+)
LSEI_1000	ABC-type Na <sup>+</sup> efflux pump, permease component	LSEI_0999-1000 (+)
LSEI_1/38	Peptide ABC transporter permease	LSEI_1/38-1/39 (-)
LSEI_1/43	Cation transport AlPase (CopA)	LSEI_1/43-1/45 (-)
LSEI_1759	Phosphoenolpyruvate-protein phosphotransferase	LSEI_1759-1760 (-)
LSEI_2601	Multidrug ABC transporter ATPase/permease	LSEI_2599-2602 (+)
Biogenesis of the cell wall		
LSEI_0221	D-alanyl-D-alanine carboxypeptidase	LSEI_0219-0221 (+)
LSEI_0238	PST family polysaccharide transporter	LSEI_0238-0240 (+)
LSEI_0794	D-alanine-activating enzyme (DItA)	LSEI_0793-0797 (+)
LSEI_0796	D-alanyi carrier protein (DitC)	
	D-alanyi transfer protein (DitD)	
LSEI_2540	+	L3EI_2343-2346 (-)
Regulators	t	
	Transcriptional regulator	No
	Transcriptional regulator	
LSEI 2025	Fe2 $\pm$ /7n2 $\pm$ untake regulation protein	LJLI_0354-0550 (+)
Two-component systems		110
	DNA-binding response regulator	ISEL 0219-0221 (⊥)
LSEL_0275	Signal transduction histidine kinase	LSEL_0219-0221 (+)
General functions		
Predicted functions		
I SFL 0135	Diadenosine tetraphosphatase-like protein	No
LSEI 0781	Phosphoesterase. DHH family protein	LSEI 0780-0783 (+)
LSEI 1621	GTPase	LSEI 1619-1625 (-)
No predicted functions		
LSEI_0086	Hypothetical protein (transmembrane domain)	LSEI_0084-0086 (+)
LSEI 0156	Hypothetical protein	LSEI 0156-0158 (+)
LSEI_0806	Hypothetical protein (transmembrane domain)	LSEI_0805-0807 (+)
LSEI_1316	Hypothetical protein	LSEI_1314-1319 (+)
LSEI_1461	Hypothetical protein (transmembrane domain)	LSEI_1458-1461 (-)
LSEI_1710	Integral membrane domain	LSEI_1710-1712 (-)
LSEI_2262	Hypothetical protein (102aa)	LSEI_2262-2263 (-)
LSEI_2553	Hypothetical protein	LSEI_2551-2553 (+)
LSEI A13	Hypothetical protein	No

\*Two consecutive genes were considered to be in the same operon when they were separated by less than 100 bp and no transcription terminator can be predicted. (+) and (-) indicate the orientation of genes, taking the genome sequencing as the (+) orientation.

of genes, although not affecting bacterial viability in vitro, are likely to affect the bacterial physiology to a point that is incompatible with survival in the harsh conditions imposed by the in vivo environment.

Our screening identified only one general stress protein: LSEI 1313, a membrane GTPase. Other general stress-response genes such as ClpL (LSEI 2048) and ClpC (LSEI 2517) that encode proteases appeared to be dispensable (Fig. 3). Mutants for GroES, GroEL, and the three heat-shock proteins present in the L. casei genome were not available; their deletion probably is lethal. Also, enzymes implicated in the maintenance of bacterial redox homeostasis (glutathione peroxidase, glutathione reductase, and thioreductase; Table S2) as well as the redox-sensing transcriptional regulator Rex (LSEI 2245) were not needed for gut establishment, although LSEI 2245 was up-regulated in the presence of bile (29). Otherwise, bacteriocin production (LSEI 2374-2384, four mutants) is not essential during this step of establishment. Biogenesis of the cell wall. Cell wall-associated proteins have been favorite targets in Lactobacillus-host interactions (30). L. casei is one of the Lactobacillus species containing the largest set of genes encoding such proteins. Our screening did not identify mutants for a gene encoding a known cell wall protein motif

(LPxTG, LysM) or an eukaryotic molecule-binding motif or a sortase (Table S2). Nevertheless, other mutants in the input pool could compensate for this loss of function. Four hypothetical proteins displaying a transmembrane domain were identified as essential for establishment, and three of these, LSEI\_0806, 1461, and 2553, are specific to the *L. casei–L. rhamnosus* group.

Cell-surface-related genes identified in our screening are summarized in Fig. 3. Genes of the *dlt* operon (LSEI\_0793-0797) implicated in lipoteichoic acid biosynthesis and the polysaccharide transporters LSEI 0238 and LSEI 2546 that are responsible for the export of lipoteichoic acid are needed for gut establishment. Even the LSEI 0247 regulator, which belongs to the LytR-Cps2A-Psr (LCP) family and is described as a cell-enveloperelated transcriptional attenuator, is needed. In Bacillus subtilis, LCP molecules are required for transport of teichoic acids and capsular polysaccharides to the cell wall peptidoglycan (31). No mutant is available LSEI\_0885, in the third gene predicted to be involved in teichoic acid export. A mature cell wall, particularly proper biosynthesis and branching of the lipoteichoic acid, is a crucial element during L. casei establishment in the gut. This result is consistent with another contribution (32), which showed that a Lactobacillus reuteri dltA mutant was not competitive when newly introduced in the gastrointestinal tract of previously Lactobacillus-free mice. Our screening also reports the importance of one penicillin-binding protein (PBP): LSEI 0221, a D-alanyl-Dalanine carboxypeptidase that removes the last amino acid of the peptide bridge allowing the transpeptidation between two strands of peptidoglycan, is essential for gut establishment. Its upstream genes LSEI 0219 and 0220, encoding a two-component system (TCS) that may regulate its transcription, are necessary also. The five other mutagenized PBPs did not appear to be necessary for gut establishment (Fig. 3 and Table S2).

**Carbohydrate metabolism.** Carbohydrate metabolism and energy production are one of the main functional groups required for gut establishment of *L. casei* in this study. Two genes encoding key enzymes of the energy-producing heterolactic fermentation pathway appear crucial: a phosphoketolase (LSEI\_0174) and a lactate dehydrogenase (LSEI\_2549, *ldh1*) (Fig. 4). *L. casei* possesses five lactate dehydrogenase genes, four of which were mutated. LSEI\_2549 (*ldh1*) is the main L-lactate dehydrogenase and is the only one whose disruption led to a significant (25%) decrease in lactate production during growth in vitro (33) and to the synthesis of unusual end-fermentation products such as mannitol, acetoin, and ethanol (34). Thus, a complete and less efficient reorganization of sugar metabolism should strongly impact the adaptation of *L. casei* in the gut lumen, where the



**Fig. 4.** *L. casei* genes of heterolactic fermentation and their role in *L. casei* establishment in the gut. Genes needed for gut persistence are in red letters; genes not needed are in green letters; and genes for which no mutant was found in the library (genes not tested) in black letters. Metabolic pathways were annotated using Kyoto Eencyclopedia of Genes and Genome analysis (72).

concentration and variety of sugars is limited. Conversely, the pyruvate kinase, which is a key enzyme in sugar metabolism, is dispensable for establishment, indicating that *L. casei* can bypass this step, probably by synthesizing oxaloacetate as an intermediate between phosphoenolpyruvate and pyruvate (LSEI\_1820 and then LSEI 1855) (Table S2).

LSEI\_0681, a regulator of lactose or mannose/fructose metabolism of the DeoR repressor family, is needed for *L. casei* establishment (Fig. 5). According to LSEI\_0681's predicted function, its inactivation could up-regulate lactose and tagatose metabolic pathways (Fig. 5), thereby globally deregulating sugar metabolism. Another consequence could be the extinction of this pathway, because *deoR* is the first gene of the operon (Fig. 5).

Two transporters involved in energy production are required for establishment: an ATP-binding cassette (ABC)-type Na<sup>+</sup> efflux pump (LSEI\_1000, a permease component) and a phosphotransferase system (the *ptsI* kinase, LSEI\_1759, which acts with the phosphocarrier HPr, LSEI\_1760; no mutant available) as an energy coupling protein. The *ptsI*-deficient mutant of *L. casei* BL23 is unable to ferment a large number of carbohydrates (35). *Amino acid metabolism.* The synthesis of cysteine and asparagine appears to be essential for gut establishment, because mutants LSEI\_0479, 0480, and 2162 show a deficient phenotype. These



Fig. 5. L. casei genes of lactose and mannose metabolism and their role in L. casei establishment in the gut. (A) Metabolic pathway context. (B) Genetic environment. Genes needed for gut persistence are in red letters; genes not needed are in green letters; genes for which no mutant was found in the library (genes not tested) are in black letters.

amino acids may be a strong limiting factor in the ileal milieu. The *L. casei* ATCC334 strain can complete only the last step of cysteine synthesis; hence it is auxotrophic for this amino acid. Conversely, mutations in genes for tryptophane, alanine, and threonine synthesis did not affect establishment (Fig. 6 and Table S2). The cysteine synthase activity of the enzyme encoded by LSEI\_0480 was demonstrated experimentally (36). Because LSEI\_0479 and 0480 form an operon (Fig. 6), the phenotype of the LSEI\_0479 mutant could stem from an alteration of LSEI\_0480 transcription alone.

As do all the *Lactobacillus* species, *L. casei* shows numerous auxotrophies for amino acids which would be unfavorable in competitive environments. However, this deficiency is compensated by a large set of peptidases (27 genes) and ABC transporters predicted to transport amino acids or oligopeptides (37). Among the 15 mutated peptidase/protease genes, LSEI\_1810, a dipeptidase predicted to be secreted, is needed for establishment. It may have a high affinity for peptides containing limiting amino acids, such as cysteine and asparagine. Also two ABC transporters predicted to transport peptides, LSEI\_0242-0243 and LSEI\_1738-1739, are needed and could compensate auxotrophies (Fig. 3).

*Environmental adaptation.* Optimal bacterial adaptation to an environment is strongly associated with the expression of a series of sensors/regulators (38, 39). The *L. casei* genome contains 16 complete TCS and 124 transcriptional regulators contributing to its ability to sense and adapt to various environments (37). Among the 10 mutated TCS (a mutant of at least one of the two genes was available in our library) (Table S2), only the LSEI\_0219-0220 TCS, in the same genetic locus as a PBP gene, was needed for gut establishment, reinforcing the importance of the cell wall biogenesis in the ileal context.

Three transcriptional regulators among the 50 whose genes were mutagenized appeared to be required for optimal establishment of *L. casei* (Table S2): LSEI\_0247 (see the discussion of cell wall biogenesis, above), LSEI\_0394 (see the discussion of amino acids metabolism, above), and LSEI\_2025, which belongs to the ferric uptake regulator (Fur) family, a family of metal ion-dependent DNA-binding regulators that can sense metal ions (i.e.,  $Fe^{2+}$ ,  $Zn^{2+}$ ,  $Ni^{2+}$ , or  $Mn^{2+}$ ) or hydrogen peroxide.

Otherwise, regularly clustered, interspaced, short palindromic repeats (CRISPR)-Cas genes were reported to regulate gene expression during host-bacteria interaction (40). In this screening, the CRISPR-Cas locus (LSEI 0349-0356) is dispensable (two mutants were available). Among the six endonucleases of the genome, only one gene, LSEI 1790, was needed, in contrast to the two genes needed in the type I restriction modification system (LSEI 2094-2095, a mutant available for each gene) (Table S2). The L. casei genome encodes 105 transposases and 13 recombinases, but only one recombinase (LSEI 1403) was needed for gut establishment. This recombinase could support genetic adaptation to the gut, as shown in Bacillus fragilis in which recombinases modulate its surface in the gut (41). Moreover, a single point mutation in the EnvZ/OmpR two-component system affects the expression of more than 100 genes in Escherichia coli (42).

### Discussion

Here we report the first, to our knowledge, functional genomewide study of L. casei establishment in the gut. It was made possible by a technological breakthrough that allows transposonbased mutagenesis in this species (20). This approach was combined with differential tagging of the transposon, allowing the assembly of mutant pools that then could be challenged in vivo, according to the STM technique that allows the identification of loss-of-function mutants (43, 44). We sequenced the entire set of 9,250 mutants obtained in this study and assembled a library of 1,110 unique mutants in which the tagged transposon had integrated into a discrete gene. Considering that L. casei is expected to encode 2,929 genes (45), with a putative set of at least 400 essential genes (46, 47), one can consider that more than half of the genome was mutagenized. Because transposon insertions could have polar effect, one can safely state that nearly all the nonvital L. casei functions have been affected.

Our random mutagenesis method focused on identifying the genes necessary for bacterial establishment in the gut regardless

PNAS PLUS



Fig. 6. L. casei genes of amino acid metabolism and their role in L. casei establishment in the gut. (A) Metabolic pathway context. (B) Genetic environment. Genes needed for gut persistence are in red letters; genes not needed are in green letters; genes for which no mutant was found in the library (genes not tested) are in black letters.

the modulation of their in vivo expression, unlike global in vivo transcriptomics [e.g., in vivo expression technology (IVET) or recombination based in vivo expression technology (RIVET)] (11–13). Alternative approaches may call for (i) shotgun cloning of L. casei genome fragments into another bacterial species, its success depending upon the basal level of gut establishment of the recipient and the quality of heterospecific gene expression; or (ii) for comparative genomics when the studied strains are phylogenetically close and characterized according to their gut establishment. These approaches identified important genetic loci (14-16) but, unlike random mutagenesis, do not offer an exhaustive genomewide analysis search as does STM. Truly alternative approaches are based on quantification by massive deep sequencing of inoculated pools of thousands of different mutants grown together (17, 48). However, we favored STM followed by mutant quantification using DNA tags. Even though the generation of the mutant library was more labor intensive, it offered significant advantages for our experimental purpose: Mutants were grown individually; pools could be assembled extemporaneously; and limiting the pool size allowed a screening model in which the mutants could be exhaustively recovered. Moreover, because of the identification of mutations, mutants now are individually available for dedicated screenings, and the pools also can be assembled differently for other types of studies.

Our first aim was to investigate the functional genomics of the gut-colonization capacity of *L. casei*. We could not carry out

long-term colonization experiments in the mouse intestine because the level of colonization following intragastric administration was insufficient to allow the reliable tracking of each individual mutant composing the injected pool. As an alternative, we validated the rabbit ligated ileal loop model that has been shown to be well adapted to STM screening in *Shigella* (24– 26). Because colonization of ligated loops cannot be carried out for more than 16 h, this assay explores the initial survival of mutants to gut luminal conditions and their early steps of colonization, hence the term "establishment."

Sensing/Exchange with Environment. To adapt to ileal conditions, *L. casei* must sense environmental changes to protect itself from toxic compounds while maintaining a functional import of nutrients. Alteration of the cell wall structure may decrease the bacteria's sensing of its environment and resistance to the environment's harsh components (49). The consequences may encompass an increase in cellular permeability to environmental compounds and to protons (50), a decrease in resistance to autolysis (32), and a modification of surface properties and hence an enhanced susceptibility to bacteriocins and antimicrobial peptides. Also, some ABC transporters can be involved in resistance to antimicrobial or peptides by excluding these compounds. In fact, ABC transporters, in coordination with a TCS, were shown to be involved in resistance to antimicrobial peptides (51). Also, the mutant in LSEI\_0394, a gene encoding a regulator

of the AcrR family that is located in an operon encoding an ABC transporter (LSEI\_0395-0396) is impaired in its capacity of establishment, whereas the disruption of this transporter had no impact on gut establishment. It is likely that LSEI\_0394 encodes a repressor of this transporter, as is the case for AcrR in *E. coli* on the multidrug efflux pump *acrAB* (52). Therefore, derepression of the transporter becomes a handicap for the bacterium in the gut lumen. The control of exchanges with the ileal environment is fundamental for *L. casei* establishment to capture a maximum of nutrients without unbalancing the bacterial cell content or permitting the entrance of toxic molecules.

Sugar Metabolism. Previous transcriptomic studies carried out in Lactobacillus plantarum and L. johnsonii while colonizing the gut showed an up-regulation of a large set of genes related to carbohydrate transport and metabolism, indicating a global recruitment of sugar-utilization enzymes for energy supply and a change in carbohydrate-utilization pathways to adapt to sugar limitation (15, 53, 54). In L. plantarum, energy production from maltose, melibiose, and lactose is activated, as is the import of mannose and cellobiose in the cecum of monocolonized mice (54). Our results are fully consistent with these data. Also, in B. thetaiotaomicron, the transcription of numerous genes implicated in polysaccharide degradation is activated in the gut to degrade glycans that the competitive flora (Bacteroides longum or L. casei) cannot to metabolize (55). Symbionts thus adapt their profile of substrate utilization to the availability of these substrates in the ileal milieu and in response to the presence of other symbionts, suggesting the need to define the composition of the resident microbiota further in future studies.

**Metabolism and Link with Stress/Oxidation.** While initiating colonization, pathogens encounter potentially bactericidal components of the intestinal fluid (i.e., bile, lysozyme, trypsin) and also a strong host response, especially highly reactive oxidative stress molecules, and must respond accordingly to survive. However, it is not clear how much oxidative stress is imposed onto symbionts as they establish in the gut. This screening identified only one general stress protein and no gene encoding factors of the specific response to oxidative stress.

However, cysteine synthase is involved in establishment. Concerning sulfured amino acids, competition for nutrients and during the establishment of a commensal was described for B. thetaiotaomicron, particularly the need for vitamin B12 (17), an essential cofactor for methionine biosynthesis for most bacterial species, although not for L. casei, according to our genome analysis. Cysteine also constitutes a pool of sulfured molecules implicated in redox regulation of the host intestine and of bacterial cells. The main tandem compounds that permit redox homeostasis are cysteine/cystine, glutathione/glutathione disulfide, and thioredoxin/thioredoxin disulfide (56). Glutathione is synthesized from cysteine in epithelial cells and in some bacteria. L. casei is able to use it in complement to thioredoxin to control its redox balance but is unable to synthesize it (57). Because L. casei also is auxotrophic for cysteine, it is strongly dependent on the ileal content. Moreover, cysteine is known as the most limiting nonessential amino acid in human cells (58). Thus, L. casei must compete with epithelial cells and the endogenous flora for cysteine and glutathione. Because mutants in genes involved in the redox balance, particularly thioreductases, were not affected, we hypothesize that during establishment L. casei needs cysteine for nutrition rather than for the maintenance of its redox balance. This notion is consistent with the evidence for nutritional competition, particularly for cysteine, observed between intracellular pathogens and their hosts (59). Although it often involves essential amino acids, nutritional competition also can involve other amino acids, e.g., asparagine, which appears to be decisive for the virulence of Lactococcus garvieae (60) and Francisella tularensis (61). Thus, asparagine could be another nutritional requirement for which *L. casei* must compete with its host.

Iron limitation is a major signal in the virulence of mucosal pathogens. In most pathogens Fur proteins act as central regulators for successful colonization and virulence. They control genes involved in iron homeostasis and protection against reactive oxygen species damage (62) in Shigella (63), Salmonella (64), Vibrio cholerae (65), Pseudomonas (66), Listeria (67), and Helicobacter pylori (68). Therefore it is likely that L. casei is strongly challenged by iron-limiting conditions in the gut, particularly given the high levels of lactoferrin in intestinal secretions. It also is possible that the protein regulating  $Fe2^+/Zn2^+$  uptake (LSEI 2025) allows scavenging of other essential metals and provides L. casei protection against reactive oxygen species, although the latter benefit is unlikely because L. casei does not appear to activate oxidative stress defenses in the ileal context. Identification of copA (LSEI 1743), an ATPase predicted to be responsible for translocating copper, silver, and cadmium ions across biological membranes, emphasizes the importance of metal import. Accordingly, in Enterococcus hirae, copA supports bacterial survival in extremely low-copper environments (69). In L. plantarum, copA expression was highly up-regulated in the conventional mouse gut, and a copA mutant showed decreased colonization capacity (70, 71).

In conclusion, we identified 47 mutations affecting establishment that we organized into five functional groups: basic physiological processes (housekeeping), metabolism, cell wall biogenesis, environmental adaptation, and a remaining group of genes of unknown function. In summary, most genes linked to bacterial establishment are conserved among Firmicutes. In consequence, our library provides major information regarding the colonization potential of other Firmicutes. Regarding *L. casei*, we will better characterize the major pathways controlling bacterial establishment. We now are in a position to study the impact of controlled microbiota on the establishment of this model symbiont. Our annotated library also is available to study other phenotypes, particularly the identification of *L. casei* effectors involved in immune and metabolic functions in their colonized host.

### **Materials and Methods**

**Design of the** *L. casei* **Random Mutant Library.** For STM, 70 DNA tags, previously used for *Salmonella typhimurium* STM (21) were individually cloned into the EcoRI site of pV1110 to generate 70 differently tagged transposable vectors (Fig. S2 and Tables S3 and S4). The tagged mutant library in *L. casei* was obtained using the P<sub>junc</sub>-TpaselS<sub>1223</sub> system as recently described (20) and was ordered in pools of 70 mutants. After the transposon insertion sites were identified by by individual sequencing, mutants were reassembled to constitute a library of 1,110 gene mutants.

Screening for Bacterial Establishment. Each pool of mutants was challenged in rabbit ileal loops as previously described (25, 26) with the following modifications. In each loop, 0.5 mL of bacterial suspension was injected ( $5 \times 10^7$  cfu per loop). Challenges were carried out over 16 h. The whole intestinal loop was homogenized, diluted, and spread on agar plates to obtain isolated colonies and to proceed to DNA isolation. Quantitative PCR was used to measure the proportion of each tag corresponding to each mutant in injected and recovered pools. All mutants displaying at least a threefold decrease in quantity between injection and recovery were selected.

Detailed experimental procedures are described in *SI Text*.

ACKNOWLEDGMENTS. We thank Ellen Arena for careful reading of and pertinent suggestions regarding the manuscript, Cyril laconelli for advice in data processing, Christoph Tang for providing DNA tags, and the Pasteur Institute Micro-organism Collection for the *L. casei* type strain (CIP 107868, ATCC 334). This work and a postdoctoral fellowship (to H.L.-S.) were supported by an Advanced Grant of the European Research Council HOMEOEPITH, Grant Agreement ERC-232798 (to P.J.S.). P.J.S. is a Howard Hughes Medical Institute Senior Scholar. H.S. was supported by a doctoral fellowship from the Ministère de l'Enseignement Supérieur et de la Recherche.

- Savage DC (2001) Microbial biota of the human intestine: A tribute to some pioneering scientists. Curr Issues Intest Microbiol 2(1):1–15.
- 2. O'Hara AM, Shanahan F (2006) The gut flora as a forgotten organ. EMBO Rep 7(7):688–693.
- Brandtzaeg P (2013) Gate-keeper function of the intestinal epithelium. *Benef Microbes* 4(1):67–82.
  Bocci V (1992) The neglected organ: Bacterial flora has a crucial immunostimulatory
- role. Perspect Biol Med 35(2):251–260.
  Tremaroli V. Bäckhed F (2012) Functional interactions between the out microbiota
- and host metabolism. *Nature* 489(7415):242–249.
- Eberl G, Boneca IG (2010) Bacteria and MAMP-induced morphogenesis of the immune system. Curr Opin Immunol 22(4):448–454.
- 7. Qin J, et al.; MetaHIT Consortium (2010) A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* 464(7285):59–65.
- Packey CD, Sartor RB (2009) Commensal bacteria, traditional and opportunistic pathogens, dysbiosis and bacterial killing in inflammatory bowel diseases. Curr Opin Infect Dis 22(3):292–301.
- Arumugam M, et al.; MetaHIT Consortium (2011) Enterotypes of the human gut microbiome. Nature 473(7346):174–180.
- 10. Chow J, Lee SM, Shen Y, Khosravi A, Mazmanian SK (2010) Host-bacterial symbiosis in health and disease. *Adv Immunol* 107:243–274.
- O'Connell Motherway M, et al. (2011) Functional genome analysis of *Bifidobacterium* breve UCC2003 reveals type IVb tight adherence (Tad) pili as an essential and conserved host-colonization factor. Proc Natl Acad Sci USA 108(27):11217–11222.
- Bron PA, Grangette C, Mercenier A, de Vos WM, Kleerebezem M (2004) Identification of *Lactobacillus plantarum* genes that are induced in the gastrointestinal tract of mice. J Bacteriol 186(17):5721–5729.
- 13. Walter J, et al. (2003) Identification of *Lactobacillus reuteri* genes specifically induced in the mouse gastrointestinal tract. *Appl Environ Microbiol* 69(4):2044–2051.
- Kankainen M, et al. (2009) Comparative genomic analysis of Lactobacillus rhamnosus GG reveals pili containing a human- mucus binding protein. Proc Natl Acad Sci USA 106(40):17193–17198.
- Denou E, et al. (2008) Identification of genes associated with the long-gut-persistence phenotype of the probiotic *Lactobacillus johnsonii* strain NCC533 using a combination of genomics and transcriptome analysis. *J Bacteriol* 190(9):3161–3168.
- Lee SM, et al. (2013) Bacterial colonization factors control specificity and stability of the gut microbiota. *Nature* 501(7467):426–429.
- 17. Goodman AL, et al. (2009) Identifying genetic determinants needed to establish a human gut symbiont in its habitat. *Cell Host Microbe* 6(3):279–289.
- Karlsson CL, Molin G, Cilio CM, Ahrné S (2011) The pioneer gut microbiota in human neonates vaginally born at term-a pilot study. *Pediatr Res* 70(3):282–286.
- Fang F, O'Toole PW (2009) Genetic tools for investigating the biology of commensal lactobacilli. Front Biosci (Landmark Ed) 14:3111–3127.
- Licandro-Seraut H, et al. (2012) Development of an efficient in vivo system (P<sub>junc</sub>-TpaselS<sub>1223</sub>) for random transposon mutagenesis of Lactobacillus casei. Appl Environ Microbiol 78(15):5417–5423.
- Hensel M, et al. (1995) Simultaneous identification of bacterial virulence genes by negative selection. Science 269(5222):400–403.
- Commichau FM, Pietack N, Stülke J (2013) Essential genes in *Bacillus subtilis*: A reevaluation after ten years. *Mol Biosyst* 9(6):1068–1075.
- Caspi R, et al. (2010) The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Res* 38(Database issue):D473–D479.
- Sansonetti PJ, et al. (1983) Alterations in the pathogenicity of *Escherichia coli* K-12 after transfer of plasmid and chromosomal genes from *Shigella flexneri*. Infect Immun 39(3):1392–1402.
- Marteyn B, et al. (2010) Modulation of Shigella virulence in response to available oxygen in vivo. Nature 465(7296):355–358.
- West NP, et al. (2005) Optimization of virulence functions through glucosylation of Shigella LPS. Science 307(5713):1313–1317.
- Oozeer R, Mater DD, Goupil-Feuillerat N, Corthier G (2004) Initiation of protein synthesis by a labeled derivative of the *Lactobacillus casei* DN-114 001 strain during transit from the stomach to the cecum in mice harboring human microbiota. *Appl Environ Microbiol* 70(12):6992–6997.
- McRobbie AM, et al. (2012) Staphylococcus aureus DinG, a helicase that has evolved into a nuclease. Biochem J 442(1):77–84.
- Alcántara C, Zúñiga M (2012) Proteomic and transcriptomic analysis of the response to bile stress of *Lactobacillus casei* BL23. *Microbiology* 158(Pt 5):1206–1218.
- Kleerebezem M, et al. (2010) The extracellular biology of the lactobacilli. FEMS Microbiol Rev 34(2):199–230.
- Kawai Y, et al. (2011) A widespread family of bacterial cell wall assembly proteins. EMBO J 30(24):4931–4941.
- Walter J, et al. (2007) D-alanyl ester depletion of teichoic acids in *Lactobacillus reuteri* 100-23 results in impaired colonization of the mouse gastrointestinal tract. *Environ Microbiol* 9(7):1750–1760.
- Rico J, Yebra MJ, Pérez-Martínez G, Deutscher J, Monedero V (2008) Analysis of *ldh* genes in *Lactobacillus casei* BL23: Role on lactic acid production. *J Ind Microbiol Biotechnol* 35(6):579–586.
- Viana R, Yebra MJ, Galán JL, Monedero V, Pérez-Martínez G (2005) Pleiotropic effects of lactate dehydrogenase inactivation in *Lactobacillus casei*. Res Microbiol 156(5-6):641–649.
- Viana R, et al. (2000) Enzyme I and HPr from Lactobacillus casei: Their role in sugar transport, carbon catabolite repression and inducer exclusion. Mol Microbiol 36(3):570-584.
- Bogicevic B, Berthoud H, Portmann R, Meile L, Irmler S (2012) CysK from Lactobacillus casei encodes a protein with O-acetylserine sulfhydrylase and cysteine desulfurization activity. Appl Microbiol Biotechnol 94(5):1209–1220.

- Cai H, Thompson R, Budinich MF, Broadbent JR, Steele JL (2009) Genome sequence and comparative genome analysis of *Lactobacillus casei*: Insights into their nicheassociated evolution. *Genome Biol Evol* 1:239–257.
- Sturme MH, Francke C, Siezen RJ, de Vos WM, Kleerebezem M (2007) Making sense of quorum sensing in lactobacilli: A special focus on *Lactobacillus plantarum* WCFS1. *Microbiology* 153(Pt 12):3939–3947.
- Krell T, et al. (2010) Bacterial sensor kinases: Diversity in the recognition of environmental signals. Annu Rev Microbiol 64:539–559.
- Sampson TR, Saroj SD, Llewellyn AC, Tzeng YL, Weiss DS (2013) A CRISPR/Cas system mediates bacterial innate immune evasion and virulence. *Nature* 497(7448):254–257.
- Coyne MJ, Weinacht KG, Krinos CM, Comstock LE (2003) Mpi recombinase globally modulates the surface architecture of a human commensal bacterium. Proc Natl Acad Sci USA 100(18):10446–10451.
- 42. Giraud A, et al. (2008) Dissecting the genetic components of adaptation of *Escherichia* coli to the mouse gut. *PLoS Genet* 4(1):e2.
- Saenz HL, Dehio C (2005) Signature-tagged mutagenesis: Technical advances in a negative selection method for virulence gene identification. Curr Opin Microbiol 8(5):612–619.
- West NP, Sansonetti PJ, Frankel G, Tang CM (2003) Finding your niche: What has been learnt from STM studies on GI colonization. *Trends Microbiol* 11(7):338–344.
- Makarova K, et al. (2006) Comparative genomics of the lactic acid bacteria. Proc Natl Acad Sci USA 103(42):15611–15616.
- Acevedo-Rocha CG, Fang G, Schmidt M, Ussery DW, Danchin A (2013) From essential to persistent genes: A functional approach to constructing synthetic life. *Trends Genet* 29(5):273–279.
- Juhas M, Eberl L, Glass JI (2011) Essence of life: Essential genes of minimal genomes. Trends Cell Biol 21(10):562–568.
- Langridge GC, et al. (2009) Simultaneous assay of every Salmonella Typhi gene using one million transposon mutants. Genome Res 19(12):2308–2316.
- Jordan S, Hutchings MI, Mascher T (2008) Cell envelope stress response in Grampositive bacteria. FEMS Microbiol Rev 32(1):107–146.
- Boyd DA, et al. (2000) Defects in D-alanyl-lipoteichoic acid synthesis in Streptococcus mutans results in acid sensitivity. J Bacteriol 182(21):6055–6065.
- Revilla-Guarinos A, et al. (2013) Characterization of a regulatory network of peptide antibiotic detoxification modules in *Lactobacillus casei* BL23. *Appl Environ Microbiol* 79(10):3160–3170.
- Routh MD, Su CC, Zhang Q, Yu EW (2009) Structures of AcrR and CmeR: Insight into the mechanisms of transcriptional repression and multi-drug recognition in the TetR family of regulators. *Biochim Biophys Acta* 1794(5):844–851.
- Denou E, et al. (2007) Gene expression of commensal Lactobacillus johnsonii strain NCC533 during in vitro growth and in the murine gut. J Bacteriol 189(22):8109–8119.
- Marco ML, et al. (2009) Lifestyle of Lactobacillus plantarum in the mouse caecum. Environ Microbiol 11(10):2747–2757.
- Sonnenburg JL, Chen CT, Gordon JI (2006) Genomic and metabolic studies of the impact of probiotics on a model gut symbiont and host. *PLoS Biol* 4(12):e413.
- Circu ML, Aw TY (2012) Intestinal redox biology and oxidative stress. Semin Cell Dev Biol 23(7):729–737.
- Serata M, lino T, Yasuda E, Sako T (2012) Roles of thioredoxin and thioredoxin reductase in the resistance to oxidative stress in *Lactobacillus casei*. *Microbiology* 158(Pt 4):953–962.
- Young VR (1994) Adult amino acid requirements: The case for a major revision in current recommendations. J Nutr 124(8, Suppl):15175–15235.
   Abu Vusik X, Burgaro D. (2012) Microbiol guidt for food in vitual (Nutritional vitu).
- Abu Kwaik Y, Bumann D (2013) Microbial quest for food in vivo: 'Nutritional virulence' as an emerging paradigm. Cell Microbiol 15(6):882–890.
- Menéndez A, Fernández L, Reimundo P, Guijarro JA (2007) Genes required for Lactococcus garvieae survival in a fish host. Microbiology 153(Pt 10):3286–3294.
- Gesbert G, et al. (2014) Asparagine assimilation is critical for intracellular replication and dissemination of Francisella. *Cell Microbiol* 16(3):434–449.
- Troxell B, Hassan HM (2013) Transcriptional regulation by Ferric Uptake Regulator (Fur) in pathogenic bacteria. Front Cell Infect Microbiol 3:59.
- Africa LA, Murphy ER, Egan NR, Wigley AF, Wing HJ (2011) The iron-responsive Fur/RyhB regulatory cascade modulates the *Shigella* outer membrane protease lcsP. *Infect Immun* 79(11):4543–4549.
- Leclerc JM, Dozois CM, Daigle F (2013) Role of the Salmonella enterica serovar Typhi Fur regulator and small RNAs RfrA and RfrB in iron homeostasis and interaction with host cells. *Microbiology* 159(Pt 3):591–602.
- Mey AR, Wyckoff EE, Kanukurthy V, Fisher CR, Payne SM (2005) Iron and fur regulation in Vibrio cholerae and the role of fur in virulence. Infect Immun 73(12): 8167–8178.
- 66. Cornelis P, Matthijs S, Van Oeffelen L (2009) Iron uptake regulation in *Pseudomonas* aeruginosa. Biometals 22(1):15–22.
- McLaughlin HP, et al. (2012) A putative P-type ATPase required for virulence and resistance to haem toxicity in *Listeria monocytogenes*. PLoS ONE 7(2):e30928.
- Pich OQ, Merrell DS (2013) The ferric uptake regulator of *Helicobacter pylori*: A critical player in the battle for iron and colonization of the stomach. *Future Microbiol* 8(6):725-738.
- Solioz M, Stoyanov JV (2003) Copper homeostasis in Enterococcus hirae. FEMS Microbiol Rev 27(2-3):183–195.
  Her V 27(2-3):183–195.
- Marco ML, Bongers RS, de Vos WM, Kleerebezem M (2007) Spatial and temporal expression of *Lactobacillus plantarum* genes in the gastrointestinal tracts of mice. *Appl Environ Microbiol* 73(1):124–132.
- Bron PA, Meijer M, Bongers RS, de Vos WM, Kleerebezem M (2007) Dynamics of competitive population abundance of *Lactobacillus plantarum ivi* gene mutants in faecal samples after passage through the gastrointestinal tract of mice. J Appl Microbiol 103(5):1424–1434.
- Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M (2012) KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res* 40(Database issue): D109–D114.

MICROBIOLOGY