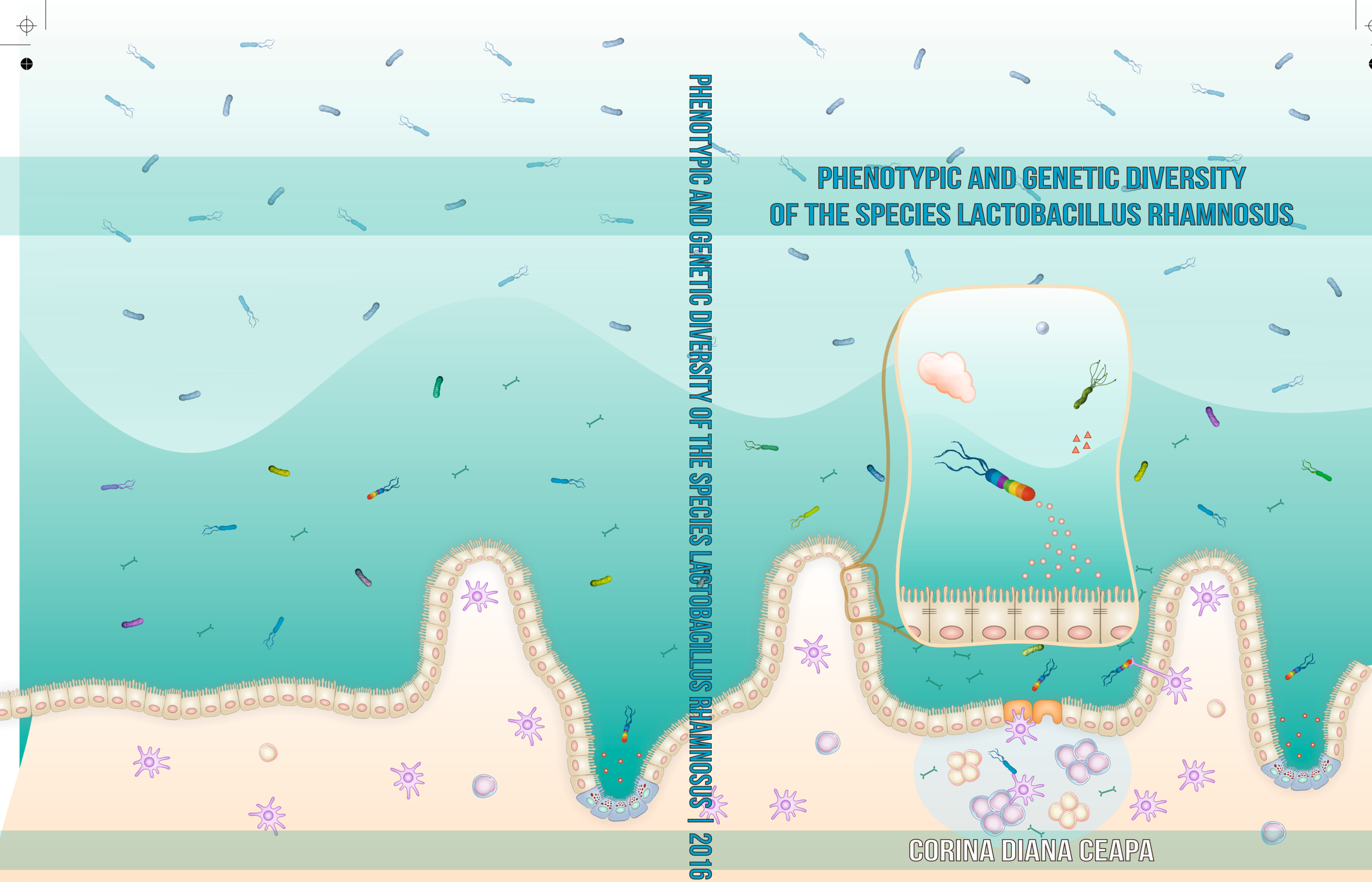
PHENOTYPIC AND GENETIC DIVERSITY OF THE SPECIES LACTOBACILLUS RHAMNOSUS | 2016

PHENOTYPIC AND GENETIC DIVERSITY OF THE SPECIES LACTOBACILLUS RHAMNOSUS



CORINA DIANA CEAPA

Propositions:

1. Small variations in gene content lead to large differences in recognition of bacterial strains by the immune system.
   (this thesis)

2. Fucose utilization is of such importance for gastro-intestinal tract adapted microbes that they have evolved to encode redundant kinase functions to support this capacity.
   (this thesis)

3. Assigning names is a human paradigm that needs to be approached in a more systematic way in the context of science.

4. Human biological, social and cultural complexity makes us intrinsically disorganized.

5. Adding an interactive component to scientific broadcasting could accelerate scientific progress as well as general public understanding of science.

6. Worldwide tax discounts for more sustainable ways of life rather than pollution taxation is a superior way to reduce climate change.

Propositions belonging to the thesis entitled:

"Phenotypic and genetic diversity of the species *Lactobacillus rhamnosus*"

Corina Diana Ceapa

Wageningen, 12 January 2016

Phenotypic and genetic diversity of the species *Lactobacillus rhamnosus*

Corina Diana Ceapa

To my loving family and dear Armando

Thesis committee

Promotor:
Prof. Dr. M. Kleerebezem
Professor of Bacterial Metagenomics,
Personal chair in the Host Microbe Interactomics Group,
Wageningen University

Thesis co-promotors:
Prof. Dr. J. Knol
Professor Early Life Microbiota, Wageningen University

Dr. J. Lambert,
Scientist Gut Biology & Microbiology
Danone Nutricia Research BV, Utrecht

Other members
Prof. Dr. T. Abee, Wageningen University
Prof. Dr. R.J. Siezen, Radboud University, Nijmegen
Dr. W. van Schaik, Utrecht University
Prof. dr. D. van Sinderen, University College Cork Ireland

This research was conducted under the auspices of the Graduate School VLAG (Voeding, Levensmiddelentechnologie, Agrobiotechnologie en Gezondheid).

Phenotypic and genetic diversity of the species *Lactobacillus rhamnosus*

Thesis

Submitted in fulfilment of the requirements for the degree of
doctor at Wageningen University
by the authority of the Rector Magnificus
Prof. Dr. Arthur Mol
in the presence of the
Thesis Committee appointed by the Academic Board
to be defended in public
on Tuesday 12th of January 2016
at 13.30 pm in the Aula.

Corina Diana Ceapa
Phenotypic and genetic diversity of the species *Lactobacillus rhamnosus*
195 pages

PhD Thesis, Wageningen University, Wageningen, NL (2016)
With references, with summaries in English and Dutch

ISBN 978-94-6257-628-5

Table of Contents

CHAPTER 1

Introduction

Delving into microbial diversity – studies on *Lactobacillus rhamnosus*

The human microbiota – an ecosystem of abundant diversity

The human microbiota is a complex and insufficiently understood group of microorganisms, which impacts on human's immunological, nutritional, and physiological processes [1]. The microbiota provides benefits to the physiology of their hosts in numerous ways: digestion of complex carbohydrates, colonization resistance against invading pathogens, maturation of the adaptive mucosal immune system and immune cells, and the production of primary and secondary metabolites like vitamins, carbohydrates, amino acids, and organic acids [1].

Owing to microbiota complexity, interactions of bacteria with their host are difficult to study [2, 3]. The rapid technological development during the last decades of the metagenomics field, as well as the implementation of post-metagenomic tools allowed more advanced analyses and sampling of these microorganisms by culture-independent methods [2]. This increased the availability of both culture independent metagenome as well as full genome sequence from cultured organisms information [4-7]. These developments bring the possibility to functionally explore members of the microbiota or whole communities [8-10], seeking a deeper understanding on how the microbiota composition and function is shaped by the diet and the host intestinal environment, and how its function may impact host physiology and susceptibility to disease [8, 11].

Studies on maintaining the homeostatic balance in host-microbe interactions may facilitate making choices for a healthy diet, and may prevent diseases and improve health in compromised subjects [22]. Recently, studies witnessed differences in community structure between anatomical sites [12, 13], between individuals [7], and between healthy and diseased states [14, 15], changing our view of human biology [16]. For instance, dysbiosis in the form of inflammatory bowel disease (IBD) and irritable bowel syndrome (IBS) [17-19] as well as obesity [7, 13, 20, 21] were correlated with a decrease in bacterial diversity, thus pointing to potential preventive therapies.

Advances in microbiota research also led to renewed interest in the non-pathogenic, commensal or even beneficial microbes that inhabit the human body [23], and the realization that we need to understand their co-evolution with the human host better to appreciate their role in human and animal health better [8]. Presence inside the GIT for a long period of time allowed for the evolution of multiple positive mechanisms to interact with the host. For example, intestinal epithelial and immune cells respond to microbiota-derived products: metabolites, proteins, cell-wall constituents or DNA [26] with inflammatory responses that, if uncontrolled, have the

potential to induce pathological responses, whereas other interactions enhance or promote anti-inflammatory mechanisms [27, 28]. Other adaptation mechanisms include managing stress (in the form of pH, presence of bile salts, antibacterial peptides, phages), attachment to epithelial surfaces and mucus and using the local carbon and nitrogen nutrients sources (utilization of different carbohydrates and proteins) [29] (Figure 1). Manipulating these processes can be a way to modulate the composition and therefore functionality of the microbiota. Intestinal bacteria are also key players in shaping the development of the mucosal immune system [38, 39].

Immunity is the composite interplay of the innate immune system that targets foreign entities and provides an unspecific protective response, and the adaptive immune system that defends against microbial challenges that can be recognized on basis of specific antigen ligands.

Through many benefits, including maximizing caloric exploitation of the diet and immune system development and maintenance, the microbiota contributes to our fitness such that we may be described as a human-microbial super organism [51, 52]. For this reason, the intestine, consisting of the highest populated bacterial ecosystem and a large percentage of the immune cells in our body, is considered a "newly discovered organ" [53]. The large intestine (LI) may include 100 trillion bacterial cells representing over 1000 different species from which approximately 80% are host-specific microbes [54, 55]. This bacterial diversity in the GIT is determined by environmental factors characteristic to the intestine: diet, the human host and the local bacterial community, among which **diet** is considered the main one [34, 59, 60]. Diet is a driving force of both microbial composition and host evolution, as dietary changes exert similar influences on gut microbiota in mammals from baboons to echidnas as they do in humans [34]. A wider variety of glycans in the diet leads to additional microbial niches and an increase in the number of species in the gut, as is the case of herbivores which have on average a phylogenetically richer microbiota than carnivores do, reflecting their more complex dietary substrates [61]. The impact of diet is evident also in humans, from alterations in gut microbiota composition in babies after weaning or in adults after dietary changes [62].

A different factor that substantially adds to diversity is the **individuality of the (human) host.** Despite the consistency of the main phyla present within the microbiota, the relative proportions and the species and strains that are present varies remarkably across individuals. Notably, monozygotic twins share more microbiota similarities than dizygotic twins, or siblings, which again share more similarity than marital partners [57].

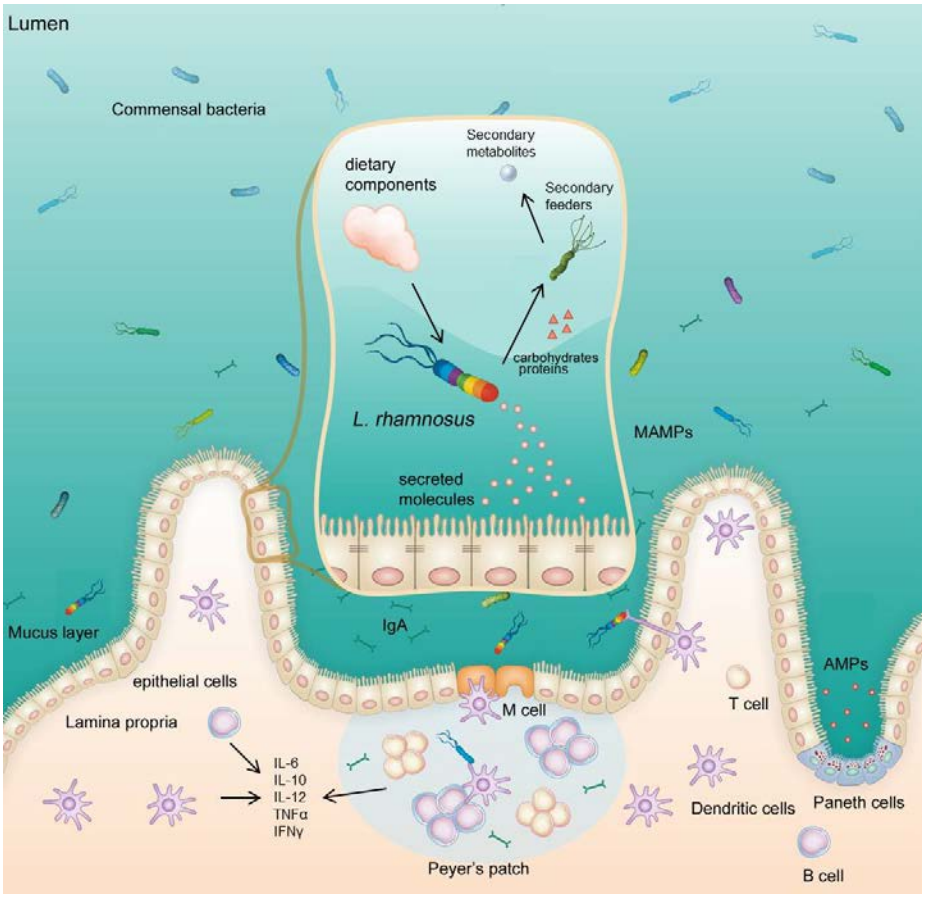


Figure 1. The intestinal environment modulates cellular differentiation in the immune system to control defense against pathogens and tolerance to commensal bacteria. Among others, suitable innate defense mechanisms limit microbial entry into intestinal tissues. Intestinal epithelial cells provide a physical barrier between the luminal microorganisms and the underlying intestinal tissues to control homeostasis and tolerance. Specialized epithelial cells produce a mucus layer (goblet cells) and secrete antimicrobial proteins (Paneth cells) that limit bacterial exposure to the epithelial cells. Production of large amounts of secretory immunoglobulin A (sIgA) by B cells provides additional protection from the luminal microbiota. Innate microbial sensing by epithelial cells, dendritic cells (DCs) and macrophages is mediated through pattern recognition receptors (PRRs) such as Toll-like receptors (TLRs). Activation of PRRs on innate immune cells normally induces pathways that mediate microbial killing and activate pro-inflammatory T cells and adaptive immune cells. During the maintenance of homeostasis and immune tolerance, however, activation of PRRs on macrophages and DCs in the intestinal lamina propria does not result in secretion of pro-inflammatory cytokines. DCs instead present antigen to T cells in the Peyer's patches and mesenteric lymph nodes, and this can lead to differentiation of regulatory T cell populations that are regulated by interleukin-10 (IL-10), transforming growth factor-β (TGF-β) and retinoic acid. Factors secreted by epithelial cells in the intestinal environment can contribute to tolerance of intestinal immune cells. AMP, antimicrobial peptides; IgA, immunoglobulin A; IL, interleukin; IFN-γ, interferon-γ; M cell, microfold cell; TNF-α, tumor necrosis factor alpha; MAMPs, microbial associated molecular patterns. Color variation on the surface of *L. rhamnosus* is used to imply the inherent genetic diversity of various strains.

These observations suggest that there is a prominent influence of host genetics on the microbiota composition, illustrating the bi-directionality of the interaction between microbes and hosts.

Despite the tremendous diversity of the microbiota composition in human individuals, metagenome sequence data analysis suggested that there might be three more prevalent or 'preferred' microbial ecosystem compositions states, designated "enterotypes", which are characterized by different abundance driver organisms, encompassing a Bacteroides, Prevotella or Ruminococcus driven enterotype [70].

Diversity in the intestinal niche changes the way we need to describe bacterial relationships, emphasizing functionality and gene content rather than belonging to a certain species. While genetic features are explored by sequencing, functional information comes, in part, from studying cultured isolates and *ex vivo* phenotypes [8, 75, 76]. These types of experiments with single isolates are extremely valuable for exploring the metabolic attributes of a particular microbe or set of microbes, and the genes involved in these functions [4, 77]. Expansion of our capacities to culture many representative species and strains derived from our gut microbial ecosystem, will enhance our capacity to understand their role in the context of the complex habitat they inhabit [126]. In addition, due to recent advances in metagenome sequencing capacity, and culture independent approaches the information on key-species from the intestinal tract is rapidly expanding [127, 128].

Particular microbiota members have been identified and extensive studies have characterized in detail how they establish specific mutualistic relationships with the host [90]. For example, *Bacteroides thetaiotaomicron* is a gram-negative anaerobe and a prominent component of the normal distal intestinal microbiota in both mice and humans. The monocolonization of *B. thetaiotaomicron* in germ-free animals was found to be able to reconstitute the entire fucosylation program of intestinal epithelial cells, and other commensal bacteria did not achieve this adaptation [91]. Similarly, prominent anti-inflammatory capacity has been assigned to *Faecalibacterium prausnitzii*, a member of the normal human microbiota that has been associated with extension of the period of remission in Crohn's disease patients [92]. Finally, human commensal *Bacteroides fragilis* produces polysaccharide A (PSA) that induces IL-10 expression from regulatory T cells, again supporting anti-inflammatory capacity and tolerance [93]. These examples provide initial molecular insights in the relationships established between typical intestinal commensals and their hosts, illustrating how specific molecular interactions can support homeostatic relationships between the host intestine and its microbiota.

Some groups of microbes associated with food products are probably transient inhabitants of the intestine rather than true commensals, and some have been associated with positive health effects on their consumers. Lilly and Stillwel [94] first described probiotics as selective nonpathogenic living microorganisms, which have beneficial effects on host health and disease prevention and/or treatment. Given the increasing interest in probiotic products from the consumer, and the continuing probiotic research in both basic sciences and clinical studies, the Food and Agriculture Organization of the United Nation–World Health Organization has now defined probiotics as "live microorganisms which, when consumed in adequate amounts as part of food, confer a health benefit on the host". Among the most commonly used probiotic bacteria are strains and species belonging to the genera of *Lactobacillus*, and *Bifidobacterium* [95].

While numerous studies to investigate their positive effects have been performed in human and animal models with different beneficial species, there is limited information about their diversity at strain level. Still, beneficial effects of probiotics are considered to be strain specific [22, 96] and environment dependent [97]. For example, several *Lactobacillus plantarum* strains showed marked differences when testing probiotic properties like tolerance to acid and bile salts, ability to ferment fructooligosaccharides (FOS), β-galactosidase activity and susceptibility to antibiotics *in vitro* [98].

Different strains of a *Lactobacillus* species can elicit a wide range of cytokine responses in immune cells and may have significantly different immunomodulatory effects *in vivo*. Lactobacilli differ considerably in their ability to trigger toll like receptor 2 (TLR2) signaling [27, 99]. Responsible for this high degree of variability can also be secreted proteins and molecules on the bacterial surface of Lactobacilli that seem to correlate with changes in the host. For example, a study of *Lactobacillus* strains that produced alternative lipoteichoic acid (LTA) variants has given important insight in the possible role of these canonical cell envelope components in host cell signaling. These studies illustrated that D-alanylation modifications of the LTA backbone increase its capacity to elicit anti-inflammatory immune effects [100], and *L. plantarum* WCFS1 [101] and *L. rhamnosus* GG [102] mutants deficient of D-alanylation of their LTA stimulated much lower TLR2-dependent secretion of pro-inflammatory cytokines compared to their proficient counterparts, which could be shown to translate into an improved protective effect of these deficient strains in a mouse colitis. This example is exemplary for the emerging notion that species and

strain specific effector molecules of probiotic bacteria are very often found in the bacterial cell envelope [129, 130].

Among the lactobacilli, *Lactobacillus rhamnosus* has a special place as it is among the most used and documented probiotic among the marketed lactobacilli, and recognized for its genetic and phenotypic diversity [104]. *L. rhamnosus* is a gram-positive organism with a low percent GC content [105] and while having been encountered as a natural inhabitant of the GI-tract of humans [106], many *L. rhamnosus* strains are isolated from the dairy environment, whereas more rarely also isolates are obtained from fermented products like kimchi or beer [107]. *L. rhamnosus*, together with *Lactobacillus casei*, *Lactobacillus paracasei*, and *Lactobacillus zeae*, constitutes a separate systematic group of the *Lactobacillus* genus [108]. Its main representative in the probiotic industry is *Lactobacillus rhamnosus* GG, which has been isolated from the human intestine and characterized extensively [105]. Human health benefits associated with LGG consumption include rotavirus diarrhea prevention [109] and a reduction of respiratory infections in children [110]. Due to strain GG utilization as a probiotic, the physiology of *L. rhamnosus* has started to be studied in detail and the species has become one of the model LABs for probiotic research.

The genome sequences of *Lactobacillus rhamnosus* strains have become publicly available [105, 108, 111-113], which provided a glimpse of the genomic diversity that may be present in this species. Their approximately 3 Mb genomes reflect some of the selective pressures that it encounters in its natural habitat and niches. For example, it contains an unusually large ensemble of genes involved in acquiring and metabolizing carbohydrates, which provides important fitness advantages in carbohydrate diverse niches like the intestine [111]. Recently, by comparing the 3.0-Mbp genome sequences of *L. rhamnosus* GG with the similarly sized genome of *L. rhamnosus* LC705, an adjunct starter culture exhibiting reduced binding to mucus, Kankainen and coworkers determined the molecular effector of the mucus binding ability – a pili-type structure formed by SpaCBA [114]. The SpaC protein is involved in mucus binding, and the expression of SpaCBA pilins was proposed to explain the longer persistence of this strain in the human intestine compared to the pilin deficient LC705 *L. rhamnosus* strain. Subsequent work has suggested that the same pilin like structure also plays a role in the capacity of the GG strain to modulate interleukin (IL)-8 induction in *in vitro* Caco-2 cell cultures [130].

At a molecular level, *L. rhamnosus* GG was proven to influence gut health by several other effector molecules. MBF, is another active mucus-specific surface

adhesin of *L. rhamnosus* GG, and contributes to the adherent mechanisms during intestinal colonization by this probiotic [115]. In addition, the p40 and p75 proteins that are released by *L. rhamnosus* GG in its environment and are derived from cell envelope muramidases, were shown to stimulate the Akt pathway, supporting apoptosis inhibition in tumor necrosis factor-induced mouse colon epithelial cells and cultured mouse colon explants. These proteinaceous factors, that appear to be conserved among *L. rhamnosus* and also *L. casei* strains are thereby shown to positively influence epithelial integrity [116, 117].

Currently there are 5 fully sequenced *L. rhamnosus* genomes within the public domain, and more than 100 strains were subjected to genome scanning and comparison to the GG genome by Illumina based sequence mapping to the GG template genome. Apart from *L. rhamnosus* GG, several strains of *Lactobacillus rhamnosus* have commonly been used as probiotics for various applications. Probiotic strain HN001 is studied for preventing the development of eczema [118] and strain GR-1 may possibly be developed into a cure for bacterial vaginosis [119]. These health effects appear strain specific. On the other hand, *L. rhamnosus* has also been reported in rare human clinical cases in predominantly immune compromised elderly and children [120]. Nevertheless, the spread of strain GG's utilization among the population was not shown not to be responsible for these cases, as bacteremia frequency didn't increase in the Finnish population during 20 years of *L. rhamnosus* GG consumption [121]. A more likely explanation for these cases is that the species can under specific conditions adapts its benign characteristics and break the intestinal barrier in immune-compromised individuals. Intriguingly in this context is that the before mentioned p40 and p75 were shown to attenuate epithelial barrier repair by a MAP kinase-dependent mechanism *in vitro* [122], whereas other studies show their activation of EGF-receptor signaling that plays an important role in epithelial wound repair [131], illustrating the variability and subtlety involved in mechanistic model design and use. Still these mechanistic insights do not fully explain the complex interaction of *L. rhamnosus* with the host mucosa, or the observed strain specificity of effects in vivo, which provides a rational for the prominent interest for more extensive comparative genomic analysis with more strains, to enable the specification of strain specific effects better and strengthen our knowledge of the *L. rhamnosus* strains that are employed by the food industry as probiotic or starters bacteria.

Thesis outline:

The research project that this thesis is based on was funded by the FP7 Marie Curie ITN project Cross-Talk (project number 21553-2) and has been established as a collaboration between Danone Nutricia Research and Wageningen University.

The overall goal of the Cross Talk project (comprised of 18 individual research fellowships) is to answer key questions on the role of the host-gut microbiota cross-talk in the development and maintenance of health by (i) identification of biological components and strains with high health potential, (ii) proposition of design for rational selection and use of probiotics, and (iii) formulation of recommendations regarding *in vitro* and *in vivo* approaches for future studies.

Danone's overall objective is bringing health through food to as many people as possible. The company has a keen interest to develop, test and validate hypotheses resulting in scientific insights, principles and concepts that can be transformed into safe and responsible specialized nutritional products for the most vulnerable groups in our society; babies, toddlers, patients and elderly.

Within this framework, the main goal of the current project is to determine the impact of the diversity within a probiotic *Lactobacillus* species in relation to its general biotechnological and immune-modulatory capacities. For this purpose, we selected *L. rhamnosus* as a species of interest for different reasons, including its genome size, natural genetic diversity, environmental versatility, and capacity of some strains to modulate the host immune system *in vivo* [123, 124]. A correlation between the genetic content of microbial strains and their effect on immune health was already proven for strains of the species [125].

Gut and microbiology research, including the role of intestinal microbiota in human health, as well as probiotic and prebiotic functionality, are at the heart of current food innovations. This is why we decided to give an overview of the current knowledge and practices in pre-, probiotics and fermented milks research in **Chapter 2**. Probiotics and prebiotics have a large spectrum of applications for human health: they can have a positive effect on the immune system, on digestive comfort and reduce risks of infection, allergy or excessive inflammatory reactions. Alterations in gut microbiota have been connected to many diseases such as obesity and intestinal bowel disease. We show here that understanding the relationship between gut microbiota, nutrition and human health, at the physiological but also molecular level is essential for developing future concepts and intervention strategies.

L. rhamnosus strains were subjected to a detailed *in vitro* characterization including the profiling of immune modulation, pathogen inhibition, surface properties

and carbohydrate utilization. The latter phenotypic aspect was analyzed in depth in **Chapter 3** and identified the metabolic cluster identifying carbohydrates L-Sorbose and α-Methyl–D-Glucoside and the genes of L. rhamnosus involved in the utilization of these carbohydrates. Several other phenotypic traits appeared to be much less variable are reported in **Chapter 7**, and the lack of variation disallowed the identification of the genetic basis involved in these phenotypes.

Another gene – trait-matching analysis focused on the capacity of individual *L. rhamnosus* strains that were able to grow on L-Fucose, allowing the identification of the predicted 5 gene fucose operon. The functionality of this operon was established by a fucose kinase deletion mutant strategy, followed by the phenotypic comparison of this mutant with the *L. rhamnosus* GG parental strain and underpinning the involvement of the operon in fucose utilization (**Chapter 4**).

An ambition of the project was also to establish reliable methodology for the accurate *in vitro* profiling of immunomodulatory capacities of individual bacterial strains. **Chapter 5** shows that uncontrolled overnight grown batch-cultures of bacteria are more variable in their immunomodulatory phenotype as compared to bacteria obtained from pH controlled batch cultures that are harvested in early stationary phase of growth. These results clearly illustrate that to obtain accurate and reproducible results, a higher degree of control of the precise conditions of growth of the bacteria should be imposed, as compared to what is common practice in studies of the interplay between bacteria and immune-cells.

To expand our insight in the genomic diversity of the *L. rhamnosus* species, the genomes of 40 strains were compared and species and strain specific genetic characteristics are discussed in **Chapter 6**. This large scale genome comparison allowed us to pinpoint genetic entities and genomic regions of plasticity within the species that are prone to be targeted by genomic rearrangement and/or horizontal gene transfer. Functions frequently encoded within these regions include carbohydrate metabolism, transport and mobile genetic elements, and are in agreement with an evolutionary view of the *L. rhamnosus* species that maintains the genetic / metabolic machinery to adapt effectively to a large range of environments, potentially by preventing high-frequency intra-species gene exchanges by the CRISPR-CAS system. Moreover, these genome analyses revealed also a repertoire of *L. rhamnosus* surface proteins and genes that are potentially influencing the interaction with the (human) host mucosa. This comparative genomics information can further fuel the quest for probiotic effector molecules and provides an important foundation for future research into probiotic functionality.

Finally, **Chapter 7** discusses the results of the thesis in the context of our current knowledge of species diversity and uniformity and its role in the capacity of intestinal lactobacilli to survive in the complex environment of the gut, and identifies perspectives for future work.

References:

1. Clemente, Jose C., et al., *The Impact of the Gut Microbiota on Human Health: An Integrative View.* Cell, 2012. 148(6): p. 1258-1270.
2. Blottière, H.M., et al., *Human intestinal metagenomics: state of the art and future.* Current opinion in microbiology, 2013. 16: p. 232-9.
3. van Baarlen, P., M. Kleerebezem, and J.M. Wells, *Omics approaches to study host–microbiota interactions.* Curr Opin Microbiol, 2013. 16(3): p. 270-277.
4. Siezen, R.J. and J.E. van Hylckama Vlieg, *Genomic diversity and versatility of Lactobacillus plantarum, a natural metabolic engineer*, in *Microbial Cell Factories*. 2011. p. S3.
5. Smokvina, T., et al., *Lactobacillus paracasei comparative genomics: towards species pan-genome definition and exploitation of diversity.* PloS one, 2013. 8: p. e68731.
6. Soon, W.W., M. Hariharan, and M.P. Snyder, *High-throughput sequencing for biology and medicine.* Molecular systems biology, 2013. 9: p. 640.
7. Turnbaugh, P.J., et al., *A core gut microbiome in obese and lean twins.* Nature, 2009. 457: p. 480-484.
8. Bienenstock, J., et al., *New insights into probiotic mechanisms: a harvest from functional and metagenomic studies.* Gut Microbes, 2013. 4(2): p. 94-100.
9. Douillard, F.P., et al., *Comparative genomic and functional analysis of 100 Lactobacillus rhamnosus strains and their comparison with strain GG.* PLoS genetics, 2013. 9: p. e1003683.
10. Pretzer, G., et al., *Biodiversity-Based Identification and Functional Characterization of the Mannose-Specific Adhesin of Lactobacillus plantarum.* Journal of bacteriology, 2005. 187: p. 6128-6136.
11. Aidy, S.E., et al., *The gut microbiota and mucosal homeostasis Colonized at birth or at adulthood , does it matter ? © 2013 Landes Bioscience . Do not distribute.* 2013. 4: p. 118-124.
12. Morgan, X.C., N. Segata, and C. Huttenhower, *Biodiversity and functional genomics in the human microbiome*, in *Trends in Genetics*. 2013. p. 51-58.
13. Ursell, L.K., et al., *The interpersonal and intrapersonal diversity of human-associated microbiota in key body sites.* The Journal of allergy and clinical immunology, 2012. 129: p. 1204-8.
14. Dupont, A.W. and H.L. Dupont, *The intestinal microbiota and chronic disorders of the gut.* Nature reviews. Gastroenterology & hepatology, 2011. 8: p. 523-531.
15. Abt, M.C. and D. Artis, *The intestinal microbiota in health and disease: the influence of microbial products on immune cell homeostasis.* Current opinion in gastroenterology, 2009. 25: p. 496-502.
16. Mani, S., U.a. Boelsterli, and M.R. Redinbo, *Understanding and modulating mammalian-microbial communication for improved human health.* Annual review of pharmacology and toxicology, 2014. 54: p. 559-80.
17. Noor, S.O., et al., *Ulcerative colitis and irritable bowel patients exhibit distinct abnormalities of the gut microbiota.* BMC Gastroenterol, 2010. 10: p. 134.

18. Carroll, I.M., et al., *Alterations in composition and diversity of the intestinal microbiota in patients with diarrhea-predominant irritable bowel syndrome.* Neurogastroenterol Motil, 2012. 24(6): p. 521-30, e248.

19. Simren, M., et al., *Intestinal microbiota in functional bowel disorders: a Rome foundation report*, in *Gut*. 2012.

20. Le Chatelier, E., et al., *Richness of human gut microbiome correlates with metabolic markers.* Nature, 2013. 500: p. 541-6.

21. Kinross, J.M., A.W. Darzi, and J.K. Nicholson, *Gut microbiome-host interactions in health and disease.* Genome medicine, 2011. 3: p. 14.

22. Ceapa, C., et al., *Influence of fermented milk products, prebiotics and probiotics on microbiota composition and health.* Best Pract Res Clin Gastroenterol, 2013. 27(1): p. 139-55.

23. van Baarlen, P., et al., *Human mucosal in vivo transcriptome responses to three lactobacilli indicate how probiotics may modulate human cellular pathways.* Proc Natl Acad Sci U S A, 2011. 108 Suppl 1: p. 4562-9.

24. Hooper, L.V. and J.I. Gordon, *Commensal host-bacterial relationships in the gut.* Science, 2001. 292(5519): p. 1115-8.

25. Ley, R.E., D.A. Peterson, and J.I. Gordon, *Ecological and evolutionary forces shaping microbial diversity in the human intestine.* Cell, 2006. 124(4): p. 837-48.

26. Bron, P.a., P. van Baarlen, and M. Kleerebezem, *Emerging molecular insights into the interaction between probiotics and the host intestinal mucosa.* Nature reviews. Microbiology, 2012. 10: p. 66-78.

27. van Bergenhenegouwen, J., et al., *Extracellular Vesicles Modulate Host-Microbe Responses by Altering TLR2 Activity and Phagocytosis.* PLoS One, 2014. 9(2): p. e89121.

28. Jones, S.E., et al., *Protection from intestinal inflammation by bacterial exopolysaccharides.* Journal of immunology (Baltimore, Md. : 1950), 2014. 192: p. 4813-20.

29. Lebeer, S., J. Vanderleyden, and S.C. De Keersmaecker, *Host interactions of probiotic bacterial surface molecules: comparison with commensals and pathogens.* Nat Rev Microbiol, 2010. 8(3): p. 171-84.

30. Schluter, J. and K.R. Foster, *The evolution of mutualism in gut microbiota via host epithelial selection.* PLoS Biol, 2012. 10(11): p. e1001424.

31. Sachs, J.L., R.G. Skophammer, and J.U. Regus, *Evolutionary transitions in bacterial symbiosis.* Proc Natl Acad Sci U S A, 2011. 108 Suppl 2: p. 10800-7.

32. Leigh, E.G., Jr., *The evolution of mutualism.* J Evol Biol, 2010. 23(12): p. 2507-28.

33. Ley, R.E., D.a. Peterson, and J.I. Gordon, *Ecological and evolutionary forces shaping microbial diversity in the human intestine.* Cell, 2006. 124: p. 837-48.

34. Muegge, B.D., et al., *Diet Drives Convergence in Gut Microbiome Functions Across Mammalian Phylogeny and Within Humans.* Science, 2011. 332(6032): p. 970-974.

35. Ley, R.E., et al., *Worlds within worlds: evolution of the vertebrate gut microbiota.* Nature reviews. Microbiology, 2008. 6: p. 776-788.

36. Hooper, L.V., et al., *Host-microbial symbiosis in the mammalian intestine: exploring an internal ecosystem.* BioEssays : news and reviews in molecular, cellular and developmental biology, 1998. 20: p. 336-343.

37. Maynard, C.L., et al., *Reciprocal interactions of the intestinal microbiota and immune system.* Nature, 2012. 489: p. 231-41.

38. Owen, J.L. and M. Mohamadzadeh, *Microbial Activation of Gut Dendritic Cells and the Control of Mucosal Immunity.* J Interferon Cytokine Res, 2013.

39. Wu, H.J. and E. Wu, *The role of gut microbiota in immune homeostasis and autoimmunity.* Gut Microbes, 2012. 3(1): p. 4-14.

51. Qin, J., et al., *A human gut microbial gene catalogue established by metagenomic sequencing.* 2010. 464(7285): p. 59-65.

52. Lederberg, J., *Infectious history.* Science, 2000. 288(5464): p. 287-93.

53. Willyard, C., *Microbiome: Gut reaction*, in *Nature*. 2011. p. S5-S7.

54. Turnbaugh, P.J., et al., *The Human Microbiome Project.* 2007. 449(7164): p. 804-810.

55. Chung, H., et al., *Gut Immune Maturation Depends on Colonization with a Host-Specific Microbiota.* Cell, 2012. 149(7): p. 1578-1593.

56. Eckburg, P.B., et al., *Diversity of the human intestinal microbial flora.* Science, 2005. 308(5728): p. 1635-8.

57. Zoetendal, E.G., et al., *The host genotype affects the bacterial community in the human gastrointestinal tract.* Microbial Ecology in Health and Disease, 2001. 13(3): p. 129-134.

58. Ley, R.E., et al., *Obesity alters gut microbial ecology.* Proc Natl Acad Sci U S A, 2005. 102(31): p. 11070-5.

59. Scott, K.P., et al., *The influence of diet on the gut microbiota.* Pharmacological Research, 2013. 69(1): p. 52-60.

60. Ley, R.E., et al., *Evolution of mammals and their gut microbes.* Science (New York, N.Y.), 2008. 320: p. 1647-1651.

61. Ley, R.E., et al., *Evolution of mammals and their gut microbes.* Science, 2008. 320(5883): p. 1647-51.

62. Rothe, M. and M. Blaut, *Evolution of the gut microbiota and the influence of diet.* Benef Microbes, 2013. 4(1): p. 31-7.

63. Flint, H.J., et al., *Microbial degradation of complex carbohydrates in the gut.* gutmicrobes, 2012. 3(4): p. 289-306.

64. Hooper, L.V., et al., *A molecular sensor that allows a gut commensal to control its nutrient foundation in a competitive ecosystem.* Proc Natl Acad Sci U S A, 1999. 96(17): p. 9833-8.

65. Hoskins, L.C., *Mucin degradation in the human gastrointestinal tract and its significance to enteric microbial ecology.* European Journal of Gastroenterology & Hepatology, 1993. 5(4): p. 205-213.

66. Scott, K.P., et al., *Whole-genome transcription profiling reveals genes up-regulated by growth on fucose in the human gut bacterium "Roseburia inulinivorans".* J Bacteriol, 2006. 188(12): p. 4340-9.

67. Bienenstock, J., et al., *Fucosylated but Not Sialylated Milk Oligosaccharides Diminish Colon Motor Contractions.* PLoS ONE, 2013. 8(10): p. e76236.

68. Schloissnig, S., et al., *Genomic variation landscape of the human gut microbiome.* Nature, 2013. 493: p. 45-50.

69. Lozupone, C.a., et al., *Diversity, stability and resilience of the human gut microbiota.* Nature, 2012. 489: p. 220-30.

70. Arumugam, M., et al., *Enterotypes of the human gut microbiome.* 2011. 473(7346): p. 174-180.

71. Candela, M., et al., *Intestinal microbiota is a plastic factor responding to environmental changes*, in *Trends in Microbiology*. 2012. p. 385-391.

72. Zoetendal, E.G., et al., *Molecular ecological analysis of the gastrointestinal microbiota: a review.* The Journal of nutrition, 2004. 134: p. 465-472.

73. Matamoros, S., et al., *Development of intestinal microbiota in infants and its impact on health*, in *Trends in Microbiology*. 2013. p. 167-173.

74. Walter, J. and R. Ley, *The human gut microbiome: ecology and recent evolutionary changes.* Annu Rev Microbiol, 2011. 65: p. 411-29.

75. Dutilh, B.E., et al., *Explaining microbial phenotypes on a genomic scale: GWAS for microbes.* Briefings in functional genomics, 2013. 12: p. 366-80.

76. Ottman, N., et al., *The function of our microbiota: who is out there and what do they do?*, in *Frontiers in Cellular and Infection Microbiology*. 2012.

77. Kleerebezem, M. and E.E. Vaughan, *Probiotic and gut lactobacilli and bifidobacteria: molecular approaches to study diversity and activity.* Annual review of microbiology, 2009. 63: p. 269-90.

78. Kirsch, P., J. Jores, and L.H. Wieler, *[Plasticity of bacterial genomes: pathogenicity islands and the locus of enterocyte effacement (LEE)].* Berl Munch Tierarztl Wochenschr, 2004. 117(3-4): p. 116-29.

79. van Schaik, W. and R.J. Willems, *Genome-based insights into the evolution of enterococci.* Clin Microbiol Infect, 2010. 16(6): p. 527-32.

80. Pillar, C.M. and M.S. Gilmore, *Enterococcal virulence--pathogenicity island of E. Faecalis.* Front Biosci, 2004. 9: p. 2335-46.

81. Fabrega, A. and J. Vila, *Salmonella enterica serovar Typhimurium skills to succeed in the host: virulence and regulation.* Clin Microbiol Rev, 2013. 26(2): p. 308-41.

82. Siriken, B., *[Salmonella pathogenicity islands].* Mikrobiyol Bul, 2013. 47(1): p. 181-8.

83. Ho, N.K., et al., *Pathogenicity, host responses and implications for management of enterohemorrhagic Escherichia coli O157:H7 infection.* Canadian Journal of Gastroenterology, 2013. 27(5): p. 281-285.

84. Baba, T., et al., *Genome sequence of Staphylococcus aureus strain newman and comparative analysis of staphylococcal genomes: Polymorphism and evolution of two major pathogenicity islands.* J Bacteriol, 2008. 190(1): p. 300-310.

85. Gerlach, R.G. and M. Hensel, *Salmonella Pathogenicity Islands in host specificity, host pathogen-interactions and antibiotics resistance of Salmonella enterica.* Berl Munch Tierarztl Wochenschr, 2007. 120(7-8): p. 317-327.

86. Kreft, J., et al., *Pathogenicity islands and other virulence elements in Listeria*, in *Curr Top Microbiol Immunol*. 2002. p. 109-125.

87. Doumith, M., et al., *New Aspects Regarding Evolution and Virulence of Listeria monocytogenes Revealed by Comparative Genomics and DNA Arrays.* Infection and Immunity, 2004. 72(2): p. 1072-1083.

88. Milillo, S.R., et al., *A Review of the Ecology, Genomics, and Stress Response of Listeria innocua and Listeria monocytogenes.* Critical Reviews in Food Science and Nutrition, 2011. 52(8): p. 712-725.

89. Achtman, M. and M. Wagner, *Microbial diversity and the genetic nature of microbial species.* Nat Rev Microbiol, 2008. 6(6): p. 431-40.

90. Ivanov, I.I. and K. Honda, *Intestinal commensal microbes as immune modulators*, in *Cell Host and Microbe*. 2012. p. 496-508.

91. Wrzosek, L., et al., *Bacteroides thetaiotaomicron and Faecalibacterium prausnitzii influence the production of mucus glycans and the development of goblet cells in the colonic epithelium of a gnotobiotic model rodent.* BMC biology, 2013. 11: p. 61.

92. Sokol, H., et al., *Faecalibacterium prausnitzii is an anti-inflammatory commensal bacterium identified by gut microbiota analysis of Crohn disease patients.* Proceedings of the National Academy of Sciences of the United States of America, 2008. 105: p. 16731-16736.

93. Mazmanian, S.K., J.L. Round, and D.L. Kasper, *A microbial symbiosis factor prevents intestinal inflammatory disease.* Nature, 2008. 453: p. 620-625.

94. Lilly, D.M. and R.H. Stillwell, *Probiotics: Growth-Promoting Factors Produced by Microorganisms.* Science, 1965. 147(3659): p. 747-8.

95. Vanderpool, C., F. Yan, and D.B. Polk, *Mechanisms of probiotic action: Implications for therapeutic applications in inflammatory bowel diseases.* Inflamm Bowel Dis, 2008. 14(11): p. 1585-96.

96. Lee, I.C., et al., *The quest for probiotic effector molecules--unraveling strain specificity at the molecular level.* Pharmacol Res, 2013. 69(1): p. 61-74.

97. van Baarlen, P., et al., *Differential NF-kappaB pathways induction by Lactobacillus plantarum in the duodenum of healthy humans correlating with immune tolerance.* Proc Natl Acad Sci U S A, 2009. 106(7): p. 2371-6.

98. Cebeci, A. and C. Gürakan, *Properties of potential probiotic Lactobacillus plantarum strains.* Food Microbiology, 2003. 20: p. 511-518.

99. Remus, D.M., et al., *Impact of 4 Lactobacillus plantarum capsular polysaccharide clusters on surface glycan composition and host cell signaling.* Microbial cell factories, 2012. 11: p. 149.

100. Lebeer, S., I.J.J. Claes, and J. Vanderleyden, *Anti-inflammatory potential of probiotics: Lipoteichoic acid makes a difference.* Trends Microbiol, 2012. 20(1): p. 5-10.

101. Bron, P.A., et al., *Lactobacillus plantarum possesses the capability for wall teichoic acid backbone alditol switching.* Microb Cell Fact, 2012. 11.

102. Claes, I.J.J., et al., *Impact of lipoteichoic acid modification on the performance of the probiotic Lactobacillus rhamnosus GG in experimental colitis.* Clin Exp Immunol, 2010. 162(2): p. 306-314.

103. Zadeh, M., et al., *Induction of intestinal pro-inflammatory immune responses by lipoteichoic acid.* J Inflamm (Lond), 2012. 9.

104. Douillard, F.P., et al., *Comparative genomic and functional analysis of 100 Lactobacillus rhamnosus strains and their comparison with strain GG.* PLoS Genet, 2013. 9(8): p. e1003683.

105. Morita, H., et al., *Complete genome sequence of the probiotic Lactobacillus rhamnosus ATCC 53103.* Journal of bacteriology, 2009. 191: p. 7630-7631.

106. Heilig, H.G., et al., *Molecular diversity of Lactobacillus spp. and other lactic acid bacteria in the human intestine as determined by specific amplification of 16S ribosomal DNA.* Appl Environ Microbiol, 2002. 68(1): p. 114-23.

107. Pittet, V., et al., *Genome Sequence of Lactobacillus rhamnosus ATCC 8530*, in *Journal of Bacteriology*. 2012. p. 726-726.

108. Toh, H., et al., *Genomic Adaptation of the Lactobacillus casei Group.* PloS one, 2013. 8: p. e75073.

109. Szajewska, H., M. Wanke, and B. Patro, *Meta-analysis: the effects of Lactobacillus rhamnosus GG supplementation for the prevention of healthcare-associated diarrhoea in children.* Aliment Pharmacol Ther, 2011. 34(9): p. 1079-87.

110. Liu, S., et al., *Lactobacillus rhamnosus GG supplementation for preventing respiratory infections in children: a meta-analysis of randomized, placebo-controlled trials.* Indian Pediatr, 2013. 50(4): p. 377-81.

111. Douillard, F.P., et al., *Comparative genomic and functional analysis of Lactobacillus casei and Lactobacillus rhamnosus strains marketed as probiotics.* Applied and environmental microbiology, 2013. 79: p. 1923-33.

112. Pittet, V., et al., *Genome sequence of Lactobacillus rhamnosus ATCC 8530.* J Bacteriol, 2012. 194(3): p. 726.

113. Kankainen, M., et al., *Comparative genomic analysis of Lactobacillus rhamnosus GG reveals pili containing a human- mucus binding protein.* Proceedings of the National Academy of Sciences of the United States of America, 2009. 106: p. 17193-17198.

114. Kankainen, M., et al., *Comparative genomic analysis of Lactobacillus rhamnosus GG reveals pili containing a human-mucus binding protein.* Proceedings of the National Academy of Sciences of the United States of America, 2009. 106(40): p. 17193-17198.

115. von Ossowski, I., et al., *Functional characterization of a mucus-specific LPXTG surface adhesin from probiotic Lactobacillus rhamnosus GG.* Applied and Environmental Microbiology, 2011. 77(13): p. 4465-4472.

116. Bäuerl, C., et al., *Functional analysis of the p40 and p75 proteins from lactobacillus casei BL23.* J Mol Microbiol Biotechnol, 2011. 19(4): p. 231-241.

117. Yan, F., et al., *Soluble proteins produced by probiotic bacteria regulate intestinal epithelial cell survival and growth.* Gastroenterology, 2007. 132(2): p. 562-75.

118. Wickens, K., et al., *Early supplementation with Lactobacillus rhamnosus HN001 reduces eczema prevalence to 6 years: does it also reduce atopic sensitization?* Clin Exp Allergy, 2013. 43(9): p. 1048-57.

119. Hummelen, R., et al., *Lactobacillus rhamnosus GR-1 and L. reuteri RC-14 to prevent or cure bacterial vaginosis among women with HIV.* Int J Gynaecol Obstet, 2010. 111(3): p. 245-8.

120. Vancanneyt, M., et al., *Intraspecific genotypic characterization of Lactobacillus rhamnosus strains intended for probiotic use and isolates of human origin.* Appl Environ Microbiol, 2006. 72(8): p. 5376-83.

121. Salminen, M.K., et al., *Lactobacillus bacteremia during a rapid increase in probiotic use of Lactobacillus rhamnosus GG in Finland.* Clinical infectious diseases: an official publication of the Infectious Diseases Society of America, 2002. 35: p. 1155-1160.

122. Seth, A., et al., *Probiotics ameliorate the hydrogen peroxide-induced epithelial barrier disruption by a PKC- and MAP kinase-dependent mechanism.* Am J Physiol Gastrointest Liver Physiol, 2008. 294(4): p. G1060-9.

123. Botina, S.G., et al., *Genetic diversity of the genus Lactobacillus bacteria from the human gastrointestinal microbiome.* Russian Journal of Genetics, 2011. 46(12): p. 1399-1406.

124. Gu, R., et al., *Analysis of different Lactobacillus rhamnosus by random amplified polymorphic DNA.* Wei Sheng Wu Xue Bao, 2008. 48(4): p. 426-431.

125. Kankainen, M., et al., *Comparative genomic analysis of Lactobacillus rhamnosus GG reveals pili containing a human- mucus binding protein.* Proc Natl Acad Sci U S A, 2009. 106(40): p. 17193-8.

126. Vartoukian, S. R., Palmer, R. M. & Wade, W. G. *Strategies for culture of 'unculturable' bacteria.* FEMS Microbiol. Lett. 309, 1–7 (2010).

127. Qin, J. et al. *A human gut microbial gene catalogue established by metagenomic sequencing.* Nature 464, 59–65 (2010).

128. Gevers, D. et al. *The Human Microbiome Project: A community resource for the healthy human microbiome.* PLoS Biol. 10, (2012).

129. Bron, P. A., Tomita, S., Mercenier, A. & Kleerebezem, M. *Cell surface-associated compounds of probiotic lactobacilli sustain the strain-specificity dogma.* Curr. Opin. Microbiol. 16, 262–269 (2013).

130. Bron, P. a, van Baarlen, P. & Kleerebezem, M. *Emerging molecular insights into the interaction between probiotics and the host intestinal mucosa.* Nat. Rev. Microbiol. 10, 66–78 (2012).

131. Zhang, L., Li, N., Caicedo, R. & Neu, J. *Alive and dead Lactobacillus rhamnosus GG decrease tumor necrosis factor-alpha-induced interleukin-8 production in Caco-2 cells.* J. Nutr. 135, 1752–1756 (2005).

CHAPTER 2

Influence of fermented milk products, prebiotics and probiotics on microbiota composition and health

Published as: Ceapa C, Wopereis H, Rezaïki L, Kleerebezem M, Knol J, Oozeer R; Influence of fermented milk products, prebiotics and probiotics on microbiota composition and health. Best Pract Res Clin Gastroenterol 2013, 27:139-155.

Abstract

The gut microbiota is a highly diverse and relative stabile ecosystem increasingly recognized for its impact on human health. The homeostasis of microbes and the host is also referred to as eubiosis. In contrast, deviation from the normal composition, defined as dysbiosis, is often associated with localized diseases such as inflammatory bowel disease or colonic cancer, but also with systemic diseases like metabolic syndrome and allergic diseases. Modulating a gut microbiota dysbiosis with nutritional concepts may contribute to improving health status, reducing diseases or disease symptoms or supporting already established treatments. The gut microbiota can be modulated by different nutritional concepts, varying from specific food ingredients to complex diets or by the ingestion of particular live microorganisms. To underpin the importance of bacteria in the gut, we describe molecular mechanisms involved in the crosstalk between gut bacteria and the human host, and review the impact of different nutritional concepts such as pre-, pro- and synbiotics on the gastrointestinal ecosystem and their potential health benefits. The aim of this review is to provide examples of potential nutritional concepts that target the gut microbiota to support human physiology and potentially health outcomes.

Keywords: fermented milk, probiotic, prebiotic, microbiota, mechanisms of action, clinical evidence.

Introduction

Gut microbiota changes in health and disease

The human body hosts roughly ten times more microorganisms than eukaryotic human cells. These organisms are part of a complex ecosystem comprising more than 3.3 million genes and corresponding to a large spectrum of enzymatic activities leading to molecular signals and metabolites that may directly influence our health and well-being. It is clear that our diet can have a significant impact on the composition and functionality of the gut microbiota and in this way can influence our health status [1-3]. As long as recorded human history goes, a strong relation was recognized between diet and our state of health [4, 5].

During the last decade the human intestinal microbiota has gained increased interest for its postulated impact on human health. Its potential implication in diseases within the gastrointestinal tract and beyond has been widely reported [6-10]. Among these, especially the immune related diseases such as allergy, inflammatory bowel disease, but also metabolic and degenerative diseases, typically increasing in industrialized societies, have been associated with altered patterns in the gut microbiota [11, 12].

Even though it is difficult to demonstrate causative relationships for specific commensal bacterial species in health and disease, there is emerging evidence for certain gut microbial species being involved in disease etiology [13-17]. In addition, in many cases, reduced microbiota diversity can be correlated to compromised health, implicating this more generic microbiota-related parameter in health and disease [18-20] . The diversity, defined as the observed number of types of species and or genes in the gut ecosystem, is generally reduced in obese individuals and IBD patients when compared to healthy controls, for example [21].

Interestingly, several microbiota transplantation studies in animals have shown that the transmission of a dysbiosed gut microbiota to their healthy counterparts is sufficient to induce the disease outcome, indicating indeed a causative relationship [22]. More recently, transplantation of a human microbiota from a lean donor to obese subjects induced an improvement of insulin-resistance confirming that microbial imbalance is not solely a secondary consequence, but can contribute to the etiology of certain diseases [23].

In-depth genetic characterization of the microbiota has recently demonstrated that human beings can be divided in three different clusters based on their microbiota composition [24]. If distinct microbiota patterns, also referred to as enterotypes, can be recognized, one can hypothesize that based on the microbiota composition different concepts can be developed to target health benefits for the different groups of individuals. Understanding the microbiota composition in relation to health and disease may allow targeted nutritional approaches to improve health outcomes or to reduce disease severity. Some of the possibilities to modulate the microbiota host interactions are reviewed here.

Manipulating human microbiota, the specific role of pro-, pre- and synbiotics

The human microbiota composition is the result of a bi-directional interaction between the host and its microbial consortium. Immune factors such as secretory IgA [25, 26] or endogenous secretions ending up in the intestine have been shown to affect

the intestinal microbiota. Besides these endogenous modulations, the microbiota composition and stability is also determined by nutrition or other factors such as antibiotics, drugs, or disease. Established high resolution tools such as 16S pyrosequencing or quantitative full metagenomic sequencing allowed a detailed assessment of the effect of different nutritional intakes on the microbiota and its gene content [27]. Non-digestible oligosaccharides are major drivers of the colonic microbiota composition by promoting the saccharolytic activity of the microbiota. As such, non-digestible oligosaccharides (NDO) have been largely used to selectively promote microbiota enrichment for lactobacilli and/or bifidobacteria and to stimulate the production of certain types of short chain fatty acids (SCFA). These NDO offer an exhaustive array of molecules with different lengths, solubility and sugar composition, and form a diverse source of substrates to alter the gut microbiota and its activities. The use of specific mixtures of NDO leading to specific changes in the gut microbiota and concomitantly confer health benefits for the host is referred to as prebiotics [28].

Elie Metchnikoff, who is considered as the founding father of the probiotics concept (concept defined below), first made reference to the properties of fermented milk (containing lactic acid bacteria) that the native long living Bulgarian populations used and linked it to increased well-being. He described the responsible microorganisms and their effects in his book, “The Prolongation of Life”, setting the base for other studies on positive effects of bacteria [29].

Later on, finding evidence of the early colonization of the GIT of infants with a new bacterium he then named *Escherichia coli*, Theodor Escherich started recommending the use of the bacteria in digestion afflictions when he discovered it in 1885 [30]. The discovery of *Bifidobacteria* in the microbiota of human milk - fed infants by Henri Tissier at the Pasteur Institute led to the recommendation to administer bifidobacteria to infants with diarrhea already in the 1950s [31, 32].

Nowadays the most common probiotic products contain bifidobacteria and/or lactobacilli, but also include other lactic acid bacteria such as lactococci and streptococci. Other promising probiotic strains include organisms of the genera *Bacillus*, *Enterococcus*, *Escherichia*, *Propionibacterium*, and the yeast genus *Saccharomyces*. Probiotics have to be isolated and characterized as pure microorganisms, and single strains or combinations of strains must be tested in appropriate human trials to give specific and measurable health benefits.

Well-characterized probiotic strains for *in vivo* survival, anti-microbial and immune properties appear to provide a relevant tool for the specific modulation of the human gut microbiota. Manipulating the microbiota with probiotics can therefore be

complementary to the application of prebiotic supplementation. The combination of both pre- and probiotics, also referred to as synbiotics, constitutes another nutritional tool to modulate the microbiota. Below we have reviewed literature on the health benefits of microorganisms that are expected to exert their benefits in the complex and dynamic environment of the gut including the residing microbiota.

Microbiota and Probiotics

Clinical evidence for probiotic use

Probiotics are defined as "live microorganisms that, when administered in adequate amounts, confer a health benefit on the consumer" (World Health Organization) [33]. Advantages of using probiotics are various: among them, lowering risk and therapy support for gastrointestinal diseases [34, 35], enhancing the immune responses and maintaining urogenital health are particularly relevant for the health care system [36]. In addition, probiotics have been demonstrated to have an adjuvant effect on vaccination [37-41], acting both locally [42] and at organism scale [43].

In present times, even if the importance of probiotics is recognized in the multitude of clinical trials and other functionally based studies, the interaction between probiotics, diet and host remains only partially understood due to the niche's complexity [44, 45]. The relationship between intestinal microbiota and the host has drawn both scientific and industrial interest to unravel molecular mechanisms of action [46]. Developments of this field could, in due course, support a more disease targeted and/or personalized therapeutic or prophylactic application of bacterial strains, with strong mechanistic and scientific support [47]. In view of personalized nutrition, we would need to adapt probiotics quantity and characteristics to host-responses and therefore increase treatment effectiveness. Examination of molecular mechanisms of action of probiotics has been described in several recent extensive reviews [46, 48-55].

There is an abundance of clinical evidence regarding the use of probiotics in health or disease that is hard to interpret and summarize because of the heterogeneity of methods and results. The Cochrane Collaboration, a rigorous data analysis organization, has taken up the task, and published topic-organized results in several reviews over the last 5 years. Their findings regarding significance of results and safety (comprising studies until December 2011) are summarized in Table 1.

The Cochrane reviews highlight that probiotic effects are strain specific and cannot be extrapolated even at the species level. The relevance of strain specificity is supported by a recent meta-study on all gastroenterological related diseases.

Table 1. The Cochrane Collaboration meta-analysis of probiotic effects in gastroenterology conditions.

Published	Disease	Effect of probiotics	Ref.
Jul-08	Active Crohn's disease	Insufficient evidence	[170]
Jul-08	*Clostridium difficile* associated colitis in adults	Insufficient evidence	[16]
Oct-08	Active ulcerative colitis	Limited evidence that probiotics may reduce disease activity; not enough evidence to recommend the use of probiotics for the treatment of active UC	[171]
Jan-09	Allergic disease and food hypersensitivity	Insufficient evidence	[172]
Jun-10	Pouchitis	Oral probiotic therapy with VSL#3 appears to be effective for acute and/or chronic pouchitis	[173]
Nov-10	Persistent diarrhoea in children	Few trials with small number of participants for a clear effect, probiotics shorten the duration of diarrhoea and reduce the stool frequency on day 5.	[174]
Dec-10	Acute infectious diarrhoea	Significantly shortened duration of diarrhoea and reduced stool frequency compared to controls, no adverse effects	[175]
Mar-11	Necrotizing enterocolitis in preterm infants	Use of probiotics reduces the occurrence of NEC and death in premature infants born less than 1500 grams	[176]
Nov-11	Prevention of pediatric antibiotic-associated diarrhea (AAD)	Some probiotic strains are effective for preventing AAD	[177]
Dec-11	Maintenance of remission in ulcerative colitis	No definite conclusion, probiotics were as ineffective as drug therapy	[178]

It summarizes 74 studies from 1970 until 2012, concluding that six of the eight diseases: pouchitis, infectious diarrhoea, irritable bowel syndrome (IBS), helicobacter pylori infection, *Clostridium difficile* infection and antibiotic associated diarrhoea showed positive effects while traveller's diarrhoea did not show significant improvement from probiotics.

For necrotizing enterocolitis, probiotics were only effective for infants with a birth weight lower than 1500g. The probiotic species and strains that were reported with a positive effect were VSL#3 (a mixture of several lactic acid bacteria and Bifidobacteria), *Lactobacillus rhamnosus GG (LGG), Saccharomyces boulardii, Bifidobacterium infantis, Lactobacillus acidophilus, Lactobacillus casei, Clostridium butyricum, Enterococcus faecum, Lactobacillus plantarium, Bifidobacterium lactis and Lactobacillus acidophilus* combined with *Bifidobacterium infantis* [56].

While clinical studies propose prospective benefits for probiotics in a variety of gastrointestinal, pancreatic and liver diseases, as well as systemic disorders like obesity [57-60] and allergy [61-63] that may have gastrointestinal symptoms, the most the most convincing evidence to date remains in the areas of infection, allergy and irritable bowel syndrome (IBS).

The mechanisms by which probiotics bacteria promote health remain speculative but appear to be a combination of direct interaction with the host and indirect effects by modulation of the GI-tract microbiota. Below we have summarized several mechanisms by which microbes directly affect health and discussed the potential of ingested microorganisms to confer health benefits by affecting the gut microbiota composition and activities.

Microbiota modulation by probiotics: molecular mechanisms of action

A unique advantage of probiotic therapy is that these living organisms are their own delivery system and potentially bring a broad repertoire of anti-pathogenic and anti-inflammatory potential into play. Possible mechanisms of action may include: (1) enhancing the natural barrier function of the normal intestinal mucosa, (2) modulation of the immune system (3), antagonism of pathogens and (4) production of enzymatic activities and/or beneficial metabolites for the host [64]. Direct host – bacteria cross-talk is reported in both clinical and pre-clinical studies. While clinical studies provide mainly insights into symptoms alleviation and therapeutic and/ or prophylactic efficacy, the real basis of the health-beneficial effect can only be obtained at the molecular level. Mechanistic insights have the potential to be further developed into therapeutic concepts. The main sites of reported probiotic action are the mucosal

interface with its immune component, the small intestine and the colon, mechanisms characteristic for each site being further developed here.

Impact of probiotics on the gut barrier

The mucus layer, the epithelial lining of the mucosal tissues as well as the immune cells, present at sub-epithelial level, are all part of the mucosal barrier. Thus, modulation at all these levels can positively affect barrier robustness and thereby influence disease state(s). Notably, several local phenomena that are dependent on each other have been reported: an increase in gut permeability [65, 66], higher mucosal inflammation [67, 68] and changes in the mucus structure and quantity [69]. At a cellular level, epithelial cells are at the center stage of the barrier effect, receiving molecular signals from the gut lumen, exchanging signals with the underlying immune cells but also communicating with the entire organism by means of circulating signaling molecules. The gut barrier plays a crucial role in the pathogenesis of numerous gastrointestinal diseases such as inflammatory bowel disease (IBD), irritable bowel syndrome (IBS), celiac disease and infectious enterocolitis [70-73]. Therefore selecting probiotic strains that can promote the gut barrier appears to be a relevant strategy with broad impact on different types of disease.

Several studies using Caco-2 intestinal cells and mice showed that *L. rhamnosus GG* (LGG) or the probiotic mix VSL#3 could interact directly with intestinal epithelial cells and maintain the integrity of the epithelial barrier. LGG persistence capacity in the GIT was linked to its *in vivo* expression of pili containing a mucus binding domain [74, 75]. In addition, LGG and its soluble factors (p75 and p40) were shown to prevent epithelial cell apoptosis *in vitro* through activating anti-apoptotic Akt and suppression of NF-κB. An additional effect observed in the study was that LGG enhances mucin secretion by epithelial cells [76], an outcome that was also observed for the Gram negative probiotic strain *E. coli* Nissle *in vitro* [77]. These effects can potentially contribute to pathogen exclusion and maintenance of homeostasis if reproducible *in vivo*. In addition, it shows that probiotic strains affect the same tissue – in this case the epithelium – by different pathways, all contributing to the preservation of the barrier effect.

In the clinical context, administration of *Lactobacillus plantarum* in the small intestine of healthy subjects induced structural changes in epithelial tight junctions, resulting in increased tight junction specific proteins occludin and zonula occludens-1. The results were reproduced in an *in vitro* model for human intestinal epithelium – the Caco2 cell line - and this significantly protected against chemically induced tight

junction damage [66]. Since loss of tight junction integrity and the resulting increased intestinal permeability to macromolecules are associated with several diseases such as IBD, IBS and celiac disease, the data obtained with the *L. plantarum* strain provide relevant information towards an intervention in the corresponding subjects [78].

In order to better characterize how probiotics strains affect the mucosal barrier, van Baarlen et al. looked at in vivo duodenal mucosal transcriptional responses of healthy adults after several probiotics interventions with strains from the species *L. plantarum*, *L. acidophilus*, *L. casei*, and *L. rhamnosus* [79, 80]. The study shows that different treatments/ strains induce differential gene-regulatory networks and pathways in the human mucosa. For instance, mucosal responses to *L. acidophilus* included up-regulation of IL-1β, an activator of NF-κB signaling cascade, which may drive the transcription of genes involved in lymphogenesis and B-cell maturation, thus contributing to enhancement of barrier function. *L. rhamnosus* consumption led to differential expression of genes involved in wound repair and healing, angiogenesis, IFN response, calcium signaling and ion homeostasis, relevant for the vascularization / nourishment of epithelial cells [80]. The observed changes in transcriptional networks display similarity with responses obtained with bioactive molecules and drugs, which may point to possible novel application areas for probiotics in either therapeutic or prophylactic nutritional regimes, aiming to strengthen the mucosal barrier.

Direct impact of probiotics on the immune system

Microbiota has been suggested as one of the main actors in the initiation or maintenance of GIT related immune diseases like atopic diseases and food allergy. Several consistent studies showed specific commensal bacterial species to exert a central role in inducing sIgA production [81] as well as in maintaining the homeostasis of several T cell populations like regulatory T cells (Treg), T helper 1(TH1) and 17 (TH17) [51, 82, 83].

Regarding probiotics strains, multiple studies have investigated their impact on sIgA. As an example, in the clinical setting, *L. rhamnosus HN001* has been shown to modulate intestinal immunity *in vivo* by increasing the levels of sIgA and other immunoglobulin secreting cells in the intestinal mucosa. However the exact mechanism by which the bacterium promote sIgA remains speculative [84-86]. *L. rhamnosus* strain GG has been shown to improve the rotavirus-specific IgA response in children with rotavirus induced diarrhea [87].

The impact of probiotics on macrophages and dendritic cells has been shown *in vivo* for some strains such as *L. casei* Shirota and *L. rhamnosus* Lr23. Both strains

trigger formation of regulatory dendritic cells and stimulate macrophages to produce TNF-α [88]. These responses are potential mechanisms by which these strains may fine-tune local immune system components and thus lowering the chances for allergy.

When looking at the molecular evidence, one of the most documented bacterial effectors on the host immune system concerns a commensal bacterial polysaccharide (PSA) from *Bacteroides fragilis* [20]. Mazmanian *et al.* established that intestinal dendritic cells appear to be vital to these effects. They are presenting PSA to CD4+ T cells and induce naive T-cell differentiation towards TH1 and Treg cells that sustain the production of appropriate cytokine profiles in the host tissues of germ-free animals [89]. In connection to that, it was speculated that changes in the bacterial cell-wall but also single surface protein structures will influence probiotic effectiveness. The idea is supported by an *in vitro* study on *L. plantarum* cell wall components. Bron and colleagues report a link between the type of teichoic acids functional groups a cell expresses on its surface and recognition by the immune system. By the production of wall teichoic acids (WTA) bacteria can shield relevant molecules on the surface and modulate the host immune response by affecting the secretion of inflammatory cytokines by dendritic cells [90]. The discovery of specific immune system effectors offers interesting perspectives that can either steer the identification of food-grade species harboring similar properties and capable of modulating the immune system in a similar way, or may stimulate the application of purified effectors as novel bioactive ingredients.

Direct cell-to-cell contact can skew immune responses as well, and can involve intermediary roles of receptor-mediated immune-recognition. Pattern recognition receptors (PRRs) are a family of receptors responsible for the detection of “microbe associated molecular patterns” (MAMPs) or host derived “damage associated molecular patterns” (DAMPs) which induce innate immune signaling [91]. A large part of the cellular immune system of the host is located below the epithelium of the gut. It is continuously challenged with signals coming from the lumen. In normal healthy conditions the mix of molecules coming from the diet and the commensal bacteria is promoting homeostasis. In the case of pathogen overload, the body senses the threat by measuring the quantity and structures of the MAMPs it receives and produces a response. This early response can be manipulated with the use of bacterial strains. As an example of such interaction, *Bifidobacterium* and *Lactobacillus* strains were shown to influence immune responses of circulating immune cells (peripheral blood mononuclear cells or PBMCs) *in vitro*. When challenged with either molecules coming from pathogens or whole pathogenic microorganisms such as *C. albicans*, PBMCs pre-

treated with probiotics s respond with different cytokine profiles that are skewed towards tolerance [92]. This may have a therapeutic potential, as dysregulation of the innate pathogen recognition system was linked to an increase in IBD symptoms [93].

The point has been made on the importance of baseline heterogeneity of human subjects when interpreting their response to probiotics treatment [94]. Inter-subject variation has been consolidated based on *in vivo* duodenal transcriptomics studies showing that despite conserved response patterns upon probiotic consumption can be observed, the baseline mucosal-molecular state of individual humans is considerably different and the conserved responses to probiotics may correlate with physiological perceivable consequences in only a susceptible subpopulation [46, 48, 80]. Although molecular responses to probiotics appear to have a significant level of conservation between individuals, the baseline variation of these same individuals may explain the distinction between responders and non-responders in probiotic trials [95], this leads to the suggestion that future use of probiotics could benefit from selection of suitable bacterial strains for administration to specific subgroups of individuals, or subsets of patient cohorts, that are stratified on basis of molecular-diagnostics [46, 48, 96].

Overall, the influence of probiotics on immune markers is extensively documented on basis of in vitro models, and several clinical trials have confirmed these observations [97-100]. Because of baseline heterogeneity of human subjects their responses to probiotic intervention and concomitant modulation of the commensal microbiota further explorative work is required to decipher these interactions and eventually optimize the clinical outcomes of probiotic intervention studies [96].

Potential health promoting metabolites produced by probiotics

On top of the molecular interactions between bacterial molecules and immune response, probiotics can exert beneficial effects through the production of bacterial metabolites. Metabolic functions of the microbiota that positively influences the host include pH changes [101], production of vitamins [102], fatty acids [103] and bile acid transformation [104], some relevant cases being discussed hereafter.

Indrio and colleagues assessed the metabolic activity of the microbiota by measuring infant fecal pH and showed similar results between infants who were human milk fed or fed with a formula fermented by BbC50 and ST065, a profile that differed from the formula without added probiotics [101]. A diet with a lower pH was correlated with protection from pathogenic gastric and pulmonary challenge in rabbits [105]. In infants, the microbiota is not mature, and the concentration and diversity of bacterial groups may not be sufficient to oppose colonization by a newly introduced

member [106]. Therefore, increased protection from pathogens for infants that cannot be human milk fed is a relevant benefit.

While pH changes can result from normal growth of most lactic acid bacteria, also a variety of members of the commensal bacteria are capable of producing vitamins or degrading bile salts. These commensal microbiota characteristics are good examples of the mutualistic relationship of the mammalian host with its microbiota, especially since humans (and many other mammals) cannot synthesize many hydrosoluble vitamins. Among vitamins of the B complex, folic acid was shown to lower colon cancer risk [107, 108]. Folate biosynthesis by the colonic microbiota was shown to be performed by several *Bifidobacterium* strains *in vitro* (e.g. *Bifidobacterium bifidum* and *Bifidobacterium longum subsp. infantis*) [109] and also *Streptococcus thermophilus* [110], *Bacillus subtilis* and *E. coli* [111]. The finding was confirmed *in vivo* when administration of high-folate producing strains increased fecal levels of folate in humans, which is especially useful at this level for the homeostasis of mucosal enterocytes of the colon [108, 112]. Exploiting the capacity of the microbiota to deliver vitamins may represent a more natural way of vitamin supplementation, compared to chemically synthesized vitamins, and provides additional health-benefits to fermented products while not affecting production costs [102]. As an important part of the diet, the dietary carbohydrates are known to influence microbial metabolism in the intestine. Dietary glycan degradation results mainly in the formation of short chain fatty acids and gases. Major bacterial fermentation products are acetate, propionate and butyrate, and their production tends to lower the colonic pH, also influencing the luminal capacity for pathogen antagonism. These weak acids influence the microbial composition and directly affect host health, with butyrate being the preferred energy source for colonocytes. Certain bacterial species in the colon are nourished through cross-feeding, using either the breakdown products of complex carbohydrate degradation or fermentation products such as lactic acid for growth [103].

Highly linked to metabolization of dietary lipids and providing an essential role in energy harvest, bile acids are synthesized from cholesterol in the liver and further metabolized by the gut microbiota into secondary bile acids. It has recently been proposed that the gut microbiota influences the host by modulating bile acid synthesis. The microbiota is capable of changing the bile acid pool composition in the small intestine by altering the host expression profile of genes involved in bile acid synthesis, conjugation, and reabsorption [104]. This is a clear example of commensal host - microbiota adaptation, as bile salts serve both a metabolic role and as a (secondary) bacterial signal to the host for microbiota status / luminal content composition.

An additional benefit probiotics can have on gut metabolism is degradation of lactose (the major sugar present in milk) in into D-glucose and D-galactose – the so-called lactase activity [113]. *Streptococcus thermophilus* produces a β-galactosidase (lactase) in the intestinal tract of mice and its presence correlates with a local reduction of the lactose content [114]. This shows that bacteria must be alive in order to help with lactose digestion. The resulting benefit is highly relevant for lactose intolerant patients alleviating abdominal pain, diarrhea, and flatulence and was observed also after consumption of yoghurt [114].

Analysis of host–microbe interactions can thus contribute to the understanding of metabolic activities of single or mixes of probiotics and facilitate the development of dietary interventions for metabolically linked disorders.

Probiotics health benefits associated to microbiota modulation

As seen above, potential health benefits of probiotics appear to depend on direct effects of probiotic strains by means of secreted cell components, metabolic effects and cell to cell interactions. The impact of probiotics strains on the human GIT microbiota seems to rely on changes in the microbial network interactions while quantitative changes appear to be moderate and poorly documented. The intricacy of the niche, the unavailability of tools that go deep enough with the analysis or the management of data with many confounding variables but especially the fact that no current consensus exists about what is a healthy microbiota makes it hard to gather this type of evidence in humans.

Probiotic bacteria, generally ingested at a level of 10^{8-9} cells, reach the colon in an amount based on survival rate in stomach and small intestine. The impact of ingested probiotics on the colonic environment is essentially attributed to the fecal persistence of the ingested strains. They colonize the gut temporarily and disappear once the consumption stops. *Lactobacillus* or *Bifidobacterium* probiotic strains can be recovered at a level ranging from 10^7 to 10^9 cells per gram corresponding to less than 0.1% of the fecal microbiota [115, 116].

Modulation of commensal microbiota by transiting probiotics can be expected due to anti-microbial compounds with broad spectrum such as reuterin [117] or plantaricins [118] or indirectly through modulation of the immune system or gut barrier function. The production of lactate during the transit may affect specifically the microbiota by promoting lactate users such as *Roseburia intestinalis [119]* or *Eubacterium halii* [120]. These commensals, as mentioned above, will produce different types of SCFA by secondary fermentation of lactate and other primary fermentation metabolites. The

impact on the microbiota of probiotics can be leveraged through food matrix fermentation process. For instance, it has been shown that it is possible to steer *S. thermophilus* to degrade lactose and generate galacto-oligosaccharides (GOS) [121]. GOS accumulates as kinetic intermediates of lactose hydrolysis and are hydrolyzed by β-galactosidases only when lactose conversion approaches 100% [122]. The main impact of fermented products is thought to come from cellular components of bacteria, metabolites and degradation of milk proteins during the fermentation process. On top of GOS, other beneficial metabolites, such as acetic acid and lactic acid result from this fermentation process [123].

Overall, few studies have reported slight changes in the fecal microbiota associated with probiotics ingestion in humans [124-129]. One of the obvious example concerns the use of probiotics in the prevention of antibiotic-induced diarrhea and acute infectious diarrhea [130, 131]. Meta-analyses and randomized controlled trials (RCTs) substantiate the use of organisms like *Saccharomyces boulardii* and *Lactobacillus rhamnosus GG* in the prevention of antibiotic-associated diarrhea (AAD) that are frequently caused by outgrowth of *C. difficile*. Antibiotherapy is associated with a significant alteration of the microbiota and probiotic interventions were observed to positively affect the recovery towards a normal microbiota following the termination of treatment by preventing outgrowth of opportunistic pathogens like *C. difficile*. Similarly other gut related disease such as IBS or colics that have been associated with microbiota dysbiosis [132-134]. VSL3 has also been shown, in a different study, to significantly improve gut comfort of IBS subjects [135].

A recent intervention looking at a probiotic effects on overall microbiota composition focused on patients with diarrhea-dominant IBS (IBS-D) treated with a probiotic mixture of *L. acidophilus, L. plantarum, L. rhamnosus, Bifidobacterium breve, B. lactis, B. longum* and *Streptococcus thermophilus.* Interestingly, fecal microbiota profiling showed a more similar microbial composition in probiotics-treated patients than that of the placebo group and patient's relief of symptoms correlated with uniformization of fecal microbiota profiles. This study is one of the first to suggest that microbial community composition is more stable during the period of probiotics treatment and that it positively correlates with improvement of disease symptoms [136].

A similar effect of stabilization of local ecology of the gut was observed after daily supplementation of the diet of infants at high risk for asthma development with of *L. rhamnosus GG* (LGG) from birth until 6 months of age. The global microbiota analysis

associated LGG abundance with a distinct community composition characterized by a higher diversity, also linking it to a reduced incidence of the allergic symptoms [137].

Most evidence available on the impact of probiotics microorganisms on the microbiota composition and functions has been obtained by using methods targeting specific bacterial genera like *Lactobacillus* and *Bifidobacteria* [138, 139] while this type of nutrition may have very subtle influence on other relevant genera as well. For example, significant reduction of bacterial diversity of members of the *Clostridium* cluster IV and significant reduction in the abundance of bacteria involved in butyrate and propionate metabolism, including *Ruminococcus bromii, Eubacterium rectale, Roseburia sp., and Akkermansia sp.* are markers of dysbiosis in ulcerative colitis (UC). Increased abundance of (opportunistic) pathogens including *Fusobacterium sp., Peptostreptococcus sp., Helicobacter sp., and Campylobacter sp.* as well as *Clostridium difficile* were found to be associated with UC [140, 141]. It remains to be established if particular species would need to be followed in specific patients or subpopulation groups. The use of new sequencing technologies will bring new insights in this direction.

Microbiota dysbiosis in immune-related disease such as allergy or IBD has been convincingly demonstrated in humans [142-147]. Thus a successful probiotic intervention may be associated with a targeted modulation of the microbiota to repress specific pathobionts or stimulate endogenous beneficial groups on top of direct molecular interaction with immune cells in the small intestine. Transiting probiotics are therefore not always expected to affect the global intestinal microbiota structure in a major way, but rather to directly modulate with the immune system and miscellaneous epithelial receptors all along the digestive tract. As a consequence low abundance but metabolically active bacteria can still be meaningful in microbiota modulation, by for example modulating existing microbiota interactive metabolic networks. All evidence taken together, probiotic strains that are able to combine specific and direct interaction with the host with transient impacts on the residing microbiota can elicit complex multifaceted but more optimal health benefits.

Microbiota and Pre- and Syn-biotics

The prebiotics concept

Modulation of the gut microbiota by applying specific non-digestible carbohydrates (NDO) has received a lot of interest since the introduction of the prebiotic concept by Gibson and Roberfroid in 1995 [148]. In 2008, the most recent definition of the prebiotic concept is formulated as a selectively fermented dietary

ingredient that results in specific changes, in the composition and activity of the gastrointestinal microbiota, thus conferring benefits upon host health [149]. This definition focuses specifically on the gut ecosystem with its indigenous microbiota as the niche of action. [150]. Other niches may be considered in the future with a similar concept, however the large intestine forms an ideal environment for microbial growth and fermentation of non-digestable dietary ingredients, since it has a slow transit time, readily available nutrients and a favorable pH [151].

The majority of scientific data have been obtained using food ingredients/supplements belonging to two chemical groups namely inulin-type fructans (ITF) and the galacto-oligosaccharides (GOS). These have repeatedly been demonstrated to selectively stimulate the growth of *Bifidobacteria* and, in some cases, lactobacilli leading to a significant change in gut microbiota composition. The concepts and their health effects have been extensively reviewed by Gibson and Roberfroid in 2010 [152]. Here we highlight in the major supposed health benefits as shown in human studies.

The supposed benefits of selectively promoting the growth of lactobacilli and in particular bifidobacteria are linked to the fact that these bacteria enact both a saccharolytic metabolism and relatively large proteolytic activities, leading to enhanced levels of lactic acid, acetate and lactate and reduced colonic pH [150]. These ecophysiological changes have been linked to an improved protection against potential pathogens [148], reduction of diarrhoea [153], improved digestion and absorption [154] and immunostimulation [155].

The prebiotic concept is of particular interest in early life, especially because human milk fed infants are dominated by bifidobacteria in contrast to infants fed cow's milk based standard formula. Human milk differs substantially from cow's milk, which is generally the basis for infant formulas (IF). While NDO are virtually absent from cow's milk, it represents the third most abundant fraction after lactose and lipids in human milk [156]. Use of a prebiotic mixture of short chain galacto-oligosaccharides (scGOS) and long chain fructo-oligosaccharides (lcFOS) (in proportions of 9:1) showed prevention of allergies and infections in newborns with effects lasting beyond the intervention period [157-159]. This finding underpins the importance of early microbial colonization and how it can influence a healthy development. Clinical data also revealed that supplementation with scGOS/lcFOS increased stool frequency and stool softness in both term and preterm infants, and similar to what is observed in human milk fed infants [160].

The adult microbiota is more complex in contrast to that of infants and is no longer dominated by *Bifidobacteria*. Nevertheless, many prebiotics cause a promotion of this genus in the colon of adults. More recently it was shown that ITF (inulin-type fructans) selectively changes the gut microbiota in obese women, leading to modest changes in key metabolites associated with obesity and diabetes [161]. Interestingly DeWulf et al. showed ITF to not only selectively promote Bifidobacteria but also *Faecalibacterium prausnitzii*. *F. prausnitzii* is regarded as beneficial in IBD patients, due to its anti-inflammatory effects [162-164]. The increase of this species elicited by the consumption of prebiotics may be explained by increased levels of acetate produced by Bifidobacteria that may act as metabolic intermediate for this secondary-fermentor, and butyrate-producing organism. The recent revolution of "omics"-approaches will define the human microbiota more and more precisely in relation to health and disease, and will help to understand how prebiotics can help in preventing or treating diseases associated with gut microbiota dysbiosis.

The synbiotic concept

A synbiotic is a combination of pro- and prebiotics. Current available combinations include bifidobacteria and fructooligosaccharides (FOS), *Lactobacillus* rhamnosus GG and inulin, and bifidobacteria and lactobacilli with FOS or inulin [165]. Although the field of synbiotics is just developing many applications have been proposed already. A few examples are presented below.

The combination of scGOS/lcFOS (9:1) and *B. breve M16-V* in a 12-week intervention in infants around 5 months of age showed reduced severity of atopic dermatitis in a subgroup of infants with elevated IgE levels but not in the whole study group. However at one year of age it was found that the synbiotic group showed attenuated use of asthma medication and lower prevalence of asthma-like symptoms in the whole study group at one year of age suggesting, long-term effects of the intervention early in life [166, 167].

Fermented milk supplemented with 2 probiotic strains, *Bifidobacterium lactis* Bi-07 and *Lactobacillus acidophilus* NCFM, and a prebiotic, isomalto-oligosaccharide, was orally administered to healthy adults and mice, and immune as well as fecal bacteria analyses were conducted using the same culturing methods. The same effects on the composition of the intestinal microbiota were observed in man and mice: increases in fecal bifidobacteria and lactobacilli and decrease of fecal enterobacteria compared to control [168].

Although, probiotics can have complementary effect to prebiotics, the future opportunities of improved health benefits may lie in supplying the combination of both. Synbiotic concepts may therefore optimize the global efficiency of modulating the microbiota in a positive way.

Concluding remarks

Understanding the complexity of the gut microbiota composition and functionality, in relation to health will offer opportunities for directed approaches through either food or pharma to improve health in the general population. The microbiota can be used to develop diagnostic tools to characterize disease status or disease risk. Modulating the microbiota with nutritional concept or drugs to cure or prevent diseases seems a target within reach [169]. The general use of fecal transplantation strategies seems unrealistic for many applications, whereas defined and accepted food strategies seem more appropriate. The concept necessary to reach the optimal effects need to be determined and could vary from simple prebiotics or single probiotics strains to more sophisticated concepts that include complex mixtures of viable micro-organisms and/or prebiotics in synbiotic concepts. The target could be general well-being, or to modulate or correct endogenous host microbe interaction in a more specific way either in upper or lower parts of the intestine. It is clear that understanding the taxonomic composition of the microbiota is as relevant as understanding the functionalities of the microbiota, which is a reflection of the ecosystems capacity to interact with specific target pathways in the host organism. The development of high throughput molecular technologies for microbiota functionality characterization will certainly catalyze the discovery of new targets for nutrition interventions that can improve health outcomes also in a clinical setting.

References:

1. Ottman, N., et al., *The function of our microbiota: who is out there and what do they do?* Front Cell Infect Microbiol, 2012. **2**: p. 104.
2. Scott, K.P., et al., *Nutritional influences on the gut microbiota and the consequences for gastrointestinal health.* Biochemical Society Transactions, 2011. **39**: p. 1073-1078.
3. O'Toole, P.W. and J.C. Cooney, *Probiotic bacteria influence the composition and function of the intestinal microbiota.* Interdiscip Perspect Infect Dis, 2008. **2008**: p. 175285.
4. Smolyansky, J., W. Ronald Ross, and R.P. Victor, *Chapter 3 - Probiotics: A Historical Perspective*, in *Bioactive Foods in Promoting Health*, Academic Press: Boston. p. 43-45.
5. Gordon, S.M., *Elie Metchnikoff: Father of natural immunity.* European Journal of Immunology, 2008. **38**(12): p. 3257-3264.
6. Clarke, G., et al., *Review article: probiotics for the treatment of irritable bowel syndrome - focus on lactic acid bacteria.* Alimentary Pharmacology & Therapeutics, 2012. **35**(4): p. 403-413.
7. Rauch, M. and S.V. Lynch, *The potential for probiotic manipulation of the gastrointestinal microbiome.* Current Opinion in Biotechnology, 2012. **23**(2): p. 192-201.

8. Orlando, A. and F. Russo, *Intestinal Microbiota, Probiotics and Human Gastrointestinal Cancers.* J Gastrointest Cancer, 2012.

9. Almansa, C., A. Agrawal, and L.A. Houghton, *Intestinal microbiota, pathophysiology and translation to probiotic use in patients with irritable bowel syndrome.* Expert Rev Gastroenterol Hepatol, 2012. **6**(3): p. 383-98.

10. Gentschew, L. and L.R. Ferguson, *Role of nutrition and microbiota in susceptibility to inflammatory bowel diseases.* Mol Nutr Food Res, 2012. **56**(4): p. 524-35.

11. Holmes, E., et al., *Therapeutic Modulation of Microbiota-Host Metabolic Interactions.* Science Translational Medicine, 2012. **4**(137).

12. Ha, E.-M., *The Impact of Gut Microbiota in Human Health and Diseases: Implication for Therapeutic Potential.* Biomolecules & Therapeutics, 2011. **19**(2): p. 155-173.

13. Possemiers, S., et al., *The intestinal environment in health and disease - recent insights on the potential of intestinal bacteria to influence human health.* Curr Pharm Des, 2009. **15**(18): p. 2051-65.

14. Lingwood, C.A., *Glycolipid receptors for verotoxin and Helicobacter pylori: role in pathology.* Biochim Biophys Acta, 1999. **1455**(2-3): p. 375-86.

15. Borg, S., et al., *Novel Salmonella typhimurium properties in host--parasite interactions.* Immunol Lett, 1999. **68**(2-3): p. 247-9.

16. Pillai, A. and R. Nelson, *Probiotics for treatment of Clostridium difficile-associated colitis in adults.* Cochrane Database Syst Rev, 2008(1): p. CD004611.

17. Fujimoto, T., et al., *Decreased abundance of Faecalibacterium prausnitzii in the gut microbiota of Crohn's disease.* J Gastroenterol Hepatol, 2013. **28**(4): p. 613-9.

18. Storro, O., E. Avershina, and K. Rudi, *Diversity of intestinal microbiota in infancy and the risk of allergic disease in childhood.* Curr Opin Allergy Clin Immunol, 2013.

19. Cozen, W., et al., *Fecal microbiota diversity in survivors of adolescent/young adult Hodgkin lymphoma: a study of twins.* Br J Cancer, 2013. **108**(5): p. 1163-7.

20. Ismail, I.H., et al., *Reduced gut microbial diversity in early life is associated with later development of eczema but not atopy in high-risk infants.* Pediatr Allergy Immunol, 2012. **23**(7): p. 674-81.

21. Luoto, R., et al., *Initial dietary and microbiological environments deviate in normal-weight compared to overweight children at 10 years of age.* J Pediatr Gastroenterol Nutr, 2011. **52**(1): p. 90-5.

22. Turnbaugh, P.J., et al., *Diet-induced obesity is linked to marked but reversible alterations in the mouse distal gut microbiome.* Cell Host Microbe, 2008. **3**(4): p. 213-23.

23. van Nood, E., et al., *Duodenal Infusion of Donor Feces for Recurrent Clostridium difficile.* New England Journal of Medicine. **368**(5): p. 407-415.

24. Arumugam, M., et al., *Enterotypes of the human gut microbiome.* Nature, 2011. **473**(7346): p. 174-80.

25. Kawamoto, S., et al., *The inhibitory receptor PD-1 regulates IgA selection and bacterial composition in the gut.* Science, 2012. **336**(6080): p. 485-9.

26. Hapfelmeier, S., et al., *Reversible microbial colonization of germ-free mice reveals the dynamics of IgA immune responses.* Science, 2010. **328**(5986): p. 1705-9.

27. Barrett, E., et al., *gamma-Aminobutyric acid production by culturable bacteria from the human intestine.* Journal of Applied Microbiology, 2012. **113**(2): p. 411-417.

28. Sim, K., et al., *Improved Detection of Bifidobacteria with Optimised 16S rRNA-Gene Based Pyrosequencing.* PLoS One, 2012. **7**(3): p. e32543.

29. Metchnikoff, E., *The prolongation of life. Optimistic studies.* William Heinemann, London, UK, 1907.

30. I. Goktepe, K.V., M.A. Juneja, *Probiotics in food safety and human health* CRC Press, Boca Raton, FL 2006.

31. Schrezenmeir, J. and M. de Vrese, *Probiotics, prebiotics, and synbiotics--approaching a definition.* Am J Clin Nutr, 2001. **73**(2 Suppl): p. 361S-364S.

32. Tissier, H., *Taxonomy and ecology of bifidobacteria.* Bifidobacteria Microflora, 1984. **3**: p. 11-28.

33. FAO/WHO, *Guidelines for the evaluation of probiotics in food.* 2002: London, Ontario.

34. Truusalu, K., et al., *Immunological, antioxidative, and morphological response in combined treatment of ofloxacin and Lactobacillus fermentum ME-3 probiotic in Salmonella Typhimurium murine model.* APMIS, 2010. **118**(11): p. 864-72.

35. Siggers, R.H., et al., *Early administration of probiotics alters bacterial colonization and limits diet-induced gut dysfunction and severity of necrotizing enterocolitis in preterm pigs.* J Nutr, 2008. **138**(8): p. 1437-44.

36. Aureli, P., et al., *Probiotics and health: an evidence-based review.* Pharmacol Res. **63**(5): p. 366-76.

37. Roos, T.B., et al., *The immune modulation of Bacillus cereus var. Toyoi in mice immunized with experimental inactivated Bovine Herpesvirus Type 5 vaccine.* Vaccine, 2012. **30**(12): p. 2173-7.

38. Moussu, H., et al., *Bifidobacterium bifidum NCC 453 promotes tolerance induction in murine models of sublingual immunotherapy.* Int Arch Allergy Immunol, 2012. **158**(1): p. 35-42.

39. Licciardi, P.V. and M.L. Tang, *Vaccine adjuvant properties of probiotic bacteria.* Discov Med, 2011. **12**(67): p. 525-33.

40. Medina, M., et al., *Lactococcus lactis as an adjuvant and delivery vehicle of antigens against pneumococcal respiratory infections.* Bioeng Bugs, 2010. **1**(5): p. 313-25.

41. Zhang, W., et al., *Probiotic Lactobacillus acidophilus enhances the immunogenicity of an oral rotavirus vaccine in gnotobiotic pigs.* Vaccine, 2008. **26**(29-30): p. 3655-61.

42. Isolauri, E., et al., *A human Lactobacillus strain (Lactobacillus casei sp strain GG) promotes recovery from acute diarrhea in children.* Pediatrics, 1991. **88**(1): p. 90-7.

43. Kaila, M., et al., *Enhancement of the circulating antibody secreting cell response in human diarrhea by a human Lactobacillus strain.* Pediatr Res, 1992. **32**(2): p. 141-4.

44. Kleerebezem, M. and E.E. Vaughan, *Probiotic and gut lactobacilli and bifidobacteria: molecular approaches to study diversity and activity.* Annu Rev Microbiol, 2009. **63**: p. 269-90.

45. Lebeer, S., J. Vanderleyden, and S.C. De Keersmaecker, *Genes and molecules of lactobacilli supporting probiotic action.* Microbiol Mol Biol Rev, 2008. **72**(4): p. 728-64, Table of Contents.

46. Klaenhammer, T.R., et al., *The impact of probiotics and prebiotics on the immune system.* Nat Rev Immunol, 2012. **12**(10): p. 728-34.

47. Remus, D.M., M. Kleerebezem, and P.A. Bron, *An intimate tete-a-tete - How probiotic lactobacilli communicate with the host.* European Journal of Pharmacology, 2011. **668**: p. S33-S42.

48. Bron, P.A., P. van Baarlen, and M. Kleerebezem, *Emerging molecular insights into the interaction between probiotics and the host intestinal mucosa.* Nature Reviews Microbiology, 2012. **10**(1): p. 66-U90.

49. Remus, D.M., M. Kleerebezem, and P.A. Bron, *An intimate tete-a-tete - how probiotic lactobacilli communicate with the host.* Eur J Pharmacol. **668 Suppl 1**: p. S33-42.

50. Siciliano, R.A. and M.F. Mazzeo, *Molecular mechanisms of probiotic action: A proteomic perspective.* Current Opinion in Microbiology, 2012. **15**(3): p. 390-396.

51. Kelly, D., M.I. Delday, and I. Mulder, *Microbes and microbial effector molecules in treatment of inflammatory disorders.* Immunological Reviews, 2012. **245**(1): p. 27-44.

52. De Keersmaecker, S.C.J., S. Lebeer, and J. Vanderleyden, *Genes and Molecules of Lactobacilli Supporting Probiotic Action.* Microbiology and Molecular Biology Reviews, 2008. **72**(4): p. 728-+.

53. Lee, I.C., et al., *The quest for probiotic effector molecules-Unraveling strain specificity at the molecular level.* Pharmacol Res, 2012.

54. Ciorba, M.A., *A gastroenterologist's guide to probiotics.* Clin Gastroenterol Hepatol. **10**(9): p. 960-8.

55. Shen, J., M.S. Obin, and L. Zhao, *The gut microbiota, obesity and insulin resistance.* Mol Aspects Med. **34**(1): p. 39-58.

56. Ritchie, M.L. and T.N. Romanuk, *A Meta-Analysis of Probiotic Efficacy for Gastrointestinal Diseases.* PLoS ONE, 2012. **7**(4).

57. Delzenne, N.M. and P.D. Cani, *Interaction Between Obesity and the Gut Microbiota: Relevance in Nutrition*, in *Annual Review of Nutrition, Vol 31*, R.J. Cousins, D.M. Bier, and B.A. Bowman, Editors. 2011. p. 15-31.

58. Delzenne, N.M., et al., *Targeting gut microbiota in obesity: effects of prebiotics and probiotics.* Nature Reviews Endocrinology, 2011. **7**(11): p. 639-646.

59. Kootte, R.S., et al., *The therapeutic potential of manipulating gut microbiota in obesity and type 2 diabetes mellitus.* Diabetes Obesity & Metabolism, 2012. **14**(2): p. 112-120.

60. Million, M., et al., *Obesity-associated gut microbiota is enriched in Lactobacillus reuteri and depleted in Bifidobacterium animalis and Methanobrevibacter smithii.* International Journal of Obesity, 2012. **36**(6): p. 817-825.

61. Kalliomaki, M., et al., *Guidance for Substantiating the Evidence for Beneficial Effects of Probiotics: Prevention and Management of Allergic Diseases by Probiotics.* Journal of Nutrition, 2010. **140**(3): p. 713S-721S.

62. Gourbeyre, P., S. Denery, and M. Bodinier, *Probiotics, prebiotics, and synbiotics: impact on the gut immune system and allergic reactions.* Journal of Leukocyte Biology, 2011. **89**(5): p. 685-695.

63. Ly, N.P., et al., *Gut microbiota, probiotics, and vitamin D: Interrelated exposures influencing allergy, asthma, and obesity?* Journal of Allergy and Clinical Immunology, 2011. **127**(5): p. 1087-1094.

64. McFarland, L.V., *Evidence-based review of probiotics for antibiotic-associated diarrhea and Clostridium difficile infections.* Anaerobe, 2009. **15**(6): p. 274-280.

65. Liu, Z., et al., *Randomised clinical trial: the effects of perioperative probiotic treatment on barrier function and post-operative infectious complications in colorectal cancer surgery - a double-blind study.* Aliment Pharmacol Ther, 2011. **33**(1): p. 50-63.

66. Karczewski, J., et al., *Regulation of human epithelial tight junction proteins by Lactobacillus plantarum in vivo and protective effects on the epithelial barrier.* American Journal of Physiology - Gastrointestinal and Liver Physiology, 2010. **298**(6): p. G851-G859.

67. Jonkers, D., et al., *Probiotics in the Management of Inflammatory Bowel Disease A Systematic Review of Intervention Studies in Adult Patients.* Drugs, 2012. **72**(6): p. 803-823.

68. Appleyard, C.B., et al., *Pretreatment with the probiotic VSL#3 delays transition from inflammation to dysplasia in a rat model of colitis-associated cancer.* Am J Physiol Gastrointest Liver Physiol, 2011. **301**(6): p. G1004-13.

69. Dorofeyev, A., *Microbiological and mucus barrier modifications in patients with inflammatory bowel diseases and extraintestinal manifestations* Journal of Crohn's and Colitis Supplements, 2008. **2**(1): p. 94-95.

70. Zakostelska, Z., et al., *Lysate of Probiotic Lactobacillus casei DN-114 001 Ameliorates Colitis by Strengthening the Gut Barrier Function and Changing the Gut Microenvironment.* PLoS One, 2011. **6**(11).

71. Blaut, M. and S. Klaus, *Intestinal microbiota and obesity.* Handb Exp Pharmacol, 2012(209): p. 251-73.

72. Iannitti, T. and B. Palmieri, *Therapeutical use of probiotic formulations in clinical practice.* Clin Nutr, 2010. **29**(6): p. 701-25.

73. Song, C.H., et al., *Probiotics promote endocytic allergen degradation in gut epithelial cells.* Biochemical and Biophysical Research Communications, 2012. **426**(1): p. 135-140.

74. Reunanen, J., et al., *Characterization of the SpaCBA pilus fibers in the probiotic Lactobacillus rhamnosus GG.* Appl Environ Microbiol, 2012. **78**(7): p. 2337-44.

75. Lebeer, S., et al., *Functional analysis of Lactobacillus rhamnosus GG pili in relation to adhesion and immunomodulatory interactions with intestinal epithelial cells.* Appl Environ Microbiol, 2011. **78**(1): p. 185-93.

76. Yan, F. and D.B. Polk, *Probiotic bacterium prevents cytokine-induced apoptosis in intestinal epithelial cells.* J Biol Chem, 2002. **277**(52): p. 50959-65.

77. Hafez, M.M., *Upregulation of Intestinal Mucin Expression by the Probiotic Bacterium E. coli Nissle 1917.* Probiotics and Antimicrobial Proteins, 2012. **4**(2): p. 67-77.

78. Sawada, N., et al., *Tight junctions and human diseases.* Med Electron Microsc, 2003. **36**(3): p. 147-56.

79. van Baarlen, P., et al., *Differential NF-kappaB pathways induction by Lactobacillus plantarum in the duodenum of healthy humans correlating with immune tolerance.* Proc Natl Acad Sci U S A, 2009. **106**(7): p. 2371-6.

80. van Baarlen, P., et al., *Human mucosal in vivo transcriptome responses to three lactobacilli indicate how probiotics may modulate human cellular pathways.* Proceedings of the National Academy of Sciences of the United States of America, 2011. **108**: p. 4562-4569.

81. Holscher, H.D. and K.A. Tappenden, *Bifidobacterium lactis BB12 enhances intestinal antibody response in formula-fed infants.* Journal of Pediatric Gastroenterology and Nutrition, 2010. **51**(, SUPPL. 2): p. E26.

82. Round, J.L. and S.K. Mazmanian, *Inducible Foxp3+ regulatory T-cell development by a commensal bacterium of the intestinal microbiota.* Proc Natl Acad Sci U S A, 2010. **107**(27): p. 12204-9.

83. Ivanov, I.I., et al., *Induction of Intestinal Th17 Cells by Segmented Filamentous Bacteria.* Cell, 2009. **139**(3): p. 485-498.

84. Wickens, K., et al., *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**(7): p. 1071-9.

85. Gill, H.S., K.J. Rutherfurd, and M.L. Cross, *Dietary probiotic supplementation enhances natural killer cell activity in the elderly: an investigation of age-related immunological changes.* J Clin Immunol, 2001. **21**(4): p. 264-71.

86. Sheih, Y.H., et al., *Systemic immunity-enhancing effects in healthy subjects following dietary consumption of the lactic acid bacterium Lactobacillus rhamnosus HN001.* J Am Coll Nutr, 2001. **20**(2 Suppl): p. 149-56.

87. Isolauri, E., *Probiotics in the prevention and treatment of allergic disease.* Pediatr Allergy Immunol, 2001. **12 Suppl 14**: p. 56-9.

88. Matsuguchi, T., et al., *Lipoteichoic acids from Lactobacillus strains elicit strong tumor necrosis factor alpha-inducing activities in macrophages through Toll-like receptor 2.* Clin Diagn Lab Immunol, 2003. **10**(2): p. 259-66.

89. Mazmanian, S.K., et al., *An immunomodulatory molecule of symbiotic bacteria directs maturation of the host immune system.* Cell, 2005. **122**(1): p. 107-18.

90. Bron, P.A., et al., *Lactobacillus plantarum possesses the capability for wall teichoic acid backbone alditol switching.* Microb Cell Fact. **11**(1): p. 123.

91. Walsh, D., et al., *Pattern recognition receptors-Molecular orchestrators of inflammation in inflammatory bowel disease.* Cytokine Growth Factor Rev. **24**(2): p. 91-104.

92. Plantinga, T.S., et al., *Modulation of Toll-like receptor ligands and Candida albicans-induced cytokine responses by specific probiotics.* Cytokine. **59**(1): p. 159-65.

93. Abraham, C. and R. Medzhitov, *Interactions Between the Host Innate Immune System and Microbes in Inflammatory Bowel Disease.* Gastroenterology. **140**(6): p. 1729-1737.

94. van Baarlen, P., J.M. Wells, and M. Kleerebezem, *Regulation of intestinal homeostasis and immunity with probiotic lactobacilli.* Trends Immunol.

95. Boessen, R., et al., *Classifying responders and non-responders; does it help when there is evidence of differentially responding patient groups?* J Psychiatr Res, 2012. **46**(9): p. 1169-73.

96. Hafrmannsperger, G., T. Clavel, and D. Haller, *Gut matters: Microbe-host interactions in allergic diseases.* Journal of Allergy and Clinical Immunology. **129**(6): p. 1452-1459.

97. Schunter, M., et al., *Randomized pilot trial of a synbiotic dietary supplement in chronic HIV-1 infection.* BMC Complement Altern Med, 2012. **12**: p. 84.

98. Gianotti, L., et al., *A randomized double-blind trial on perioperative administration of probiotics in colorectal cancer patients.* World J Gastroenterol, 2010. **16**(2): p. 167-75.

99. Boyle, R.J., et al., *Effects of Lactobacillus GG treatment during pregnancy on the development of fetal antigen-specific immune responses.* Clin Exp Allergy, 2008. **38**(12): p. 1882-90.

100. Snel, J., et al., *Strain-specific immunomodulatory effects of Lactobacillus plantarum strains on birch-pollen-allergic subjects out of season.* Clin Exp Allergy. **41**(2): p. 232-42.

101. Indrio, F., et al., *Effect of a fermented formula on thymus size and stool pH in healthy term infants.* Pediatr Res, 2007. **62**(1): p. 98-100.

102. LeBlanc, J.G., et al., *Bacteria as vitamin suppliers to their host: a gut microbiota perspective.* Current Opinion in Biotechnology. **24**(2): p. 160-168.

103. Scott, K.P., et al., *The influence of diet on the gut microbiota.* Pharmacological Research. **69**(1): p. 52-60.

104. Sayin, SamaÂ I., et al., *Gut Microbiota Regulates Bile Acid Metabolism by Reducing the Levels of Tauro-beta-muricholic Acid, a Naturally Occurring FXR Antagonist.* Cell Metabolism. **17**(2): p. 225-235.

105. Boneti, C., et al., *Probiotic acidified formula in an animal model reduces pulmonary and gastric bacterial load.* Journal of Pediatric Surgery, 2009. **44**(3): p. 530-533.

106. Matamoros, S., et al., *Development of intestinal microbiota in infants and its impact on health.* Trends in Microbiology, 2013. **21**(4): p. 167-173.

107. Hsu, S.P., et al., *236 Molecular Mechanism Underlying Folic Acid-inhibited Colon Cancer Growth.* European Journal of Cancer. **48, Supplement 5**(0): p. S57-S58.

108. Kim, Y.I., *Role of folate in colon cancer development and progression.* J Nutr, 2003. **133**(11 Suppl 1): p. 3731S-3739S.

109. Pompei, A., et al., *Folate Production by Bifidobacteria as a Potential Probiotic Property.* Applied and Environmental Microbiology, 2007. **73**(1): p. 179-185.

110. Iyer, R., et al., *Probiotic properties of folate producing Streptococcus thermophilus strains.* Food Research International, 2010. **43**(1): p. 103-110.

111. Burgess, C.M., E.J. Smid, and D. van Sinderen, *Bacterial vitamin B2, B11 and B12 overproduction: An overview.* International Journal of Food Microbiology, 2009. **133**(12): p. 1-7.

112. Strozzi, G.P. and L. Mogna, *Quantification of folic acid in human feces after administration of Bifidobacterium probiotic strains.* J Clin Gastroenterol, 2008. **42 Suppl 3 Pt 2**: p. S179-84.

113. Zhong, Y., et al., *The role of colonic microbiota in lactose intolerance.* Dig Dis Sci, 2004. **49**(1): p. 78-83.

114. Drouault, S., J. Anba, and G.r. Corthier, *Streptococcus thermophilus Is Able To Produce a B-Galactosidase Active during Its Transit in the Digestive Tract of Germ-Free Mice.* Applied and Environmental Microbiology, 2002. **68**(2): p. 938-941.

115. Gerritsen, J., et al., *Intestinal microbiota in human health and disease: the impact of probiotics.* Genes Nutr. **6**(3): p. 209-40.

116. Aureli, P., et al., *Probiotics and health: An evidence-based review.* Pharmacological Research. **63**(5): p. 366-376.

117. Talarico, T.L., et al., *Production and isolation of reuterin, a growth inhibitor produced by Lactobacillus reuteri.* Antimicrob Agents Chemother, 1988. **32**(12): p. 1854-8.

118. Holo, H., et al., *Plantaricin W from Lactobacillus plantarum belongs to a new family of two-peptide lantibiotics.* Microbiology, 2001. **147**(Pt 3): p. 643-51.

119. Duncan, S.H., P. Louis, and H.J. Flint, *Lactate-utilizing bacteria, isolated from human feces, that produce butyrate as a major fermentation product.* Appl Environ Microbiol, 2004. **70**(10): p. 5810-7.

120. Hold, G.L., et al., *Oligonucleotide Probes That Detect Quantitatively Significant Groups of Butyrate-Producing Bacteria in Human Feces.* Applied and Environmental Microbiology, 2003. **69**(7): p. 4320-4324.

121. GÃ¤nzle, M.G., *Enzymatic synthesis of galacto-oligosaccharides and other lactose derivatives (hetero-oligosaccharides) from lactose.* International Dairy Journal, 2012. **22**(2): p. 116-122.

122. Torres, D.P.M., et al., *Galacto-Oligosaccharides: Production, Properties, Applications, and Significance as Prebiotics.* Comprehensive Reviews in Food Science and Food Safety. **9**(5): p. 438-454.

123. Soukoulis, C., et al., *Proton transfer reaction time-of-flight mass spectrometry monitoring of the evolution of volatile compounds during lactic acid fermentation of milk.* Rapid Commun Mass Spectrom, 2010. **24**(14): p. 2127-3134.

124. Martin, F.-P.J., et al., *Dietary Modulation of Gut Functional Ecology Studied by Fecal Metabonomics.* Journal of Proteome Research, 2010. **9**(10): p. 5284-5295.

125. Flier, J.S. and J.J. Mekalanos, *Gut check: testing a role for the intestinal microbiome in human obesity.* Sci Transl Med, 2009. **1**(6): p. 6ps7.

126. Shahinas, D., et al., *Toward an understanding of changes in diversity associated with fecal microbiome transplantation based on 16S rRNA gene deep sequencing.* MBio. **3**(5).

127. McNulty, N.P., et al., *The impact of a consortium of fermented milk strains on the gut microbiome of gnotobiotic mice and monozygotic twins.* Sci Transl Med. **3**(106): p. 106ra106.

128. Tilg, H. and A. Kaser, *Gut microbiome, obesity, and metabolic dysfunction.* J Clin Invest. **121**(6): p. 2126-32.

129. Rampelli, S., et al., *A probiotics-containing biscuit modulates the intestinal microbiota in the elderly.* J Nutr Health Aging, 2013. **17**(2): p. 166-72.

130. Billoo, A.G., et al., *Role of a probiotic (Saccharomyces boulardii) in management and prevention of diarrhoea.* World J Gastroenterol, 2006. **12**(28): p. 4557-60.

131. Thibault, H., C. Aubert-Jacquin, and O. Goulet, *Effects of long-term consumption of a fermented infant formula (with Bifidobacterium breve c50 and Streptococcus thermophilus 065) on acute diarrhea in healthy infants.* J Pediatr Gastroenterol Nutr, 2004. **39**(2): p. 147-52.

132. Cremon, C., et al., *Intestinal dysbiosis in irritable bowel syndrome: etiological factor or epiphenomenon?* Expert Rev Mol Diagn. **10**(4): p. 389-93.

133. Whelan, K., *Probiotics and prebiotics in the management of irritable bowel syndrome: a review of recent clinical trials and systematic reviews.* Curr Opin Clin Nutr Metab Care. **14**(6): p. 581-7.

134. Roos, S., et al., *454 Pyrosequencing Analysis on Faecal Samples from a Randomized DBPC Trial of Colicky Infants Treated with Lactobacillus reuteri DSM 17938.* PLoS One, 2013. **8**(2).

135. Agrawal, A., et al., *Clinical trial: the effects of a fermented milk product containing Bifidobacterium lactis DN-173 010 on abdominal distension and gastrointestinal transit in irritable bowel syndrome with constipation.* Alimentary Pharmacology & Therapeutics, 2009. **29**(1): p. 104-114.

136. Ki Cha, B., et al., *The effect of a multispecies probiotic mixture on the symptoms and fecal microbiota in diarrhea-dominant irritable bowel syndrome: a randomized, double-blind, placebo-controlled trial.* J Clin Gastroenterol. **46**(3): p. 220-7.

137. Cox, M.J., et al., *Lactobacillus casei Abundance Is Associated with Profound Shifts in the Infant Gut Microbiome.* PLoS One. **5**(1): p. e8745.

138. Rautava, S., et al., *Probiotics modulate host-microbe interaction in the placenta and fetal gut: a randomized, double-blind, placebo-controlled trial.* Neonatology, 2012. **102**(3): p. 178-84.

139. Savard, P., et al., *Impact of Bifidobacterium animalis subsp. lactis BB-12 and, Lactobacillus acidophilus LA-5-containing yoghurt, on fecal bacterial counts of healthy adults.* International Journal of Food Microbiology, 2011. **149**(1): p. 50-57.

140. Belzer, C. and W.M. de Vos, *Microbes inside--from diversity to function: the case of Akkermansia.* ISME J, 2012. **6**(8): p. 1449-58.

141. Rajilic-Stojanovic, M., et al., *Phylogenetic analysis of dysbiosis in ulcerative colitis during remission.* Inflamm Bowel Dis, 2013. **19**(3): p. 481-8.

142. Toh, Z.Q., et al., *Probiotic therapy as a novel approach for allergic disease.* Front Pharmacol. **3**: p. 171.

143. Bushman, F.D., J.D. Lewis, and G.D. Wu, *Diet, the human gut microbiota, and IBD.* Anaerobe, (0).

144. Comito, D. and C. Romano, *Dysbiosis in the pathogenesis of pediatric inflammatory bowel diseases.* Int J Inflam. **2012**: p. 687143.

145. Elson, C.O. and Y. Cong, *Host-microbiota interactions in inflammatory bowel disease.* Gut Microbes. **3**(4): p. 332-44.

146. Kaur, N., et al., *Intestinal dysbiosis in inflammatory bowel disease.* Gut Microbes. **2**(4): p. 211-6.

147. Fava, F. and S. Danese, *Intestinal microbiota in inflammatory bowel disease: friend of foe?* World J Gastroenterol. **17**(5): p. 557-66.

148. Gibson, G.R. and M.B. Roberfroid, *Dietary Modulation of the Human Colonic Microbiota: Introducing the Concept of Prebiotics.* J. Nutr., 1995. **125**(6): p. 1401-1412.

149. ISAPP, *6th Meeting of the International Scientific Association of Probiotics and Prebiotics, London, Ontario.* 2008.

150. Gibson, G.R., et al., *Dietary prebiotics: current status and new definition.* IFIS Functional Foods Bulletin, 2010. **7**: p. 1-19.

151. Cummings, J.H. and G.T. Macfarlane, *The control and consequences of bacterial fermentation in the human colon.* Journal of Applied Microbiology, 1991. **70**(6): p. 443-459.

152. Roberfroid, M., et al., *Prebiotic effects: metabolic and health benefits.* British Journal of Nutrition, 2010. **104**(SupplementS2): p. S1-S63.

153. de Vrese, M. and J. Schrezenmeir, *Probiotics, prebiotics, and synbiotics.* Adv Biochem Eng Biotechnol, 2008. **111**: p. 1-66.

154. Fioramonti, J., V. Theodorou, and L. Bueno, *Probiotics: what are they? What are their effects on gut physiology?* Best Practice & Research Clinical Gastroenterology, 2003. **17**(5): p. 711-724.

155. Veereman, G., *Pediatric applications of inulin and oligofructose.* J Nutr, 2007. **137**(11 Suppl): p. 2585S-2589S.

156. Oozeer, R., *Intestinal microbiology in early life - Specific prebiotics can have similar functionalities as human milk oligosaccharides.* The American Journal of Clinical Nutrition, 2012.

157. Arslanoglu, S., G.E. Moro, and G. Boehm, *Early supplementation of prebiotic oligosaccharides protects formula-fed infants against infections during the first 6 months of life.* J Nutr, 2007. **137**(11): p. 2420-4.

158. Grathwohl, D., *Early supplementation of prebiotic oligosaccharides protects formula-fed infants against infections during the first six months of life.* J Nutr, 2008. **138**(8): p. 1520; author reply 1521.

159. Moro, G., et al., *A mixture of prebiotic oligosaccharides reduces the incidence of atopic dermatitis during the first six months of age.* Arch Dis Child, 2006. **91**(10): p. 814-9.

160. Moro, G., et al., *Dosage-related bifidogenic effects of galacto- and fructooligosaccharides in formula-fed term infants.* J Pediatr Gastroenterol Nutr, 2002. **34**(3): p. 291-5.

161. Dewulf, E.M., et al., *Insight into the prebiotic concept: lessons from an exploratory, double blind intervention study with inulin-type fructans in obese women.* Gut, 2012.

162. Martinez-Medina, M., et al., *Abnormal microbiota composition in the ileocolonic mucosa of Crohn's disease patients as revealed by polymerase chain reaction-denaturing gradient gel electrophoresis.* Inflamm Bowel Dis, 2006. **12**(12): p. 1136-45.

163. Frank, D.N., et al., *Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases.* Proc Natl Acad Sci U S A, 2007. **104**(34): p. 13780-5.

164. Sokol, H., et al., *Faecalibacterium prausnitzii is an anti-inflammatory commensal bacterium identified by gut microbiota analysis of Crohn disease patients.* Proc Natl Acad Sci U S A, 2008. **105**(43): p. 16731-6.

165. Tsilingiri, K. and M. Rescigno, *Postbiotics: what else?* Benef Microbes: p. 69-75.

166. van der Aa, L.B., et al., *Effect of a new synbiotic mixture on atopic dermatitis in infants: a randomized-controlled trial.* Clin Exp Allergy, 2010. **40**(5): p. 795-804.

167. van der Aa, L.B., et al., *Synbiotics prevent asthma-like symptoms in infants with atopic dermatitis.* Allergy, 2011. **66**(2): p. 170-7.

168. Wang, S., et al., *Fermented milk supplemented with probiotics and prebiotics can effectively alter the intestinal microbiota and immunity of host animals.* Journal of dairy science. **95**(9): p. 4813-4822.

169. Khanna, S., et al., *The epidemiology of community-acquired Clostridium difficile infection: a population-based study.* Am J Gastroenterol, 2012. **107**(1): p. 89-95.

170. Butterworth, A.D., A.G. Thomas, and A.K. Akobeng, *Probiotics for induction of remission in Crohn's disease.* Cochrane Database Syst Rev, 2008(3): p. CD006634.

171. Mallon, P., et al., *Probiotics for induction of remission in ulcerative colitis.* Cochrane Database Syst Rev, 2007(4): p. CD005573.

172. Osborn, D.A. and J.K. Sinn, *Probiotics in infants for prevention of allergic disease and food hypersensitivity.* Cochrane Database Syst Rev, 2007(4): p. CD006475.

173. Holubar, S.D., et al., *Treatment and prevention of pouchitis after ileal pouch-anal anastomosis for chronic ulcerative colitis.* Cochrane Database Syst Rev, 2010(6): p. CD001176.

174. Bernaola Aponte, G., et al., *Probiotics for treating persistent diarrhoea in children.* Cochrane Database Syst Rev, 2010(11): p. CD007401.

175. Allen, S.J., et al., *Probiotics for treating acute infectious diarrhoea.* Cochrane Database Syst Rev, 2010(11): p. CD003048.

176. Alfaleh, K., et al., *Probiotics for prevention of necrotizing enterocolitis in preterm infants.* Cochrane Database Syst Rev, 2011(3): p. CD005496.

177. Johnston, B.C., et al., *Probiotics for the prevention of pediatric antibiotic-associated diarrhea.* Cochrane Database Syst Rev, 2011(11): p. CD004827.

178. Naidoo, K., et al., *Probiotics for maintenance of remission in ulcerative colitis.* Cochrane Database Syst Rev, 2011(12): p. CD007443.

CHAPTER 3

Correlation of *Lactobacillus rhamnosus* genotypes and carbohydrate utilization signatures determined by phenotype profiling

Published as: 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.

Abstract:

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 α-methyl-D-glucoside. From the analysis, candidate genes were identified that correlate with L-sorbose and α-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.

Introduction

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 non-optimal 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 fermentations 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].

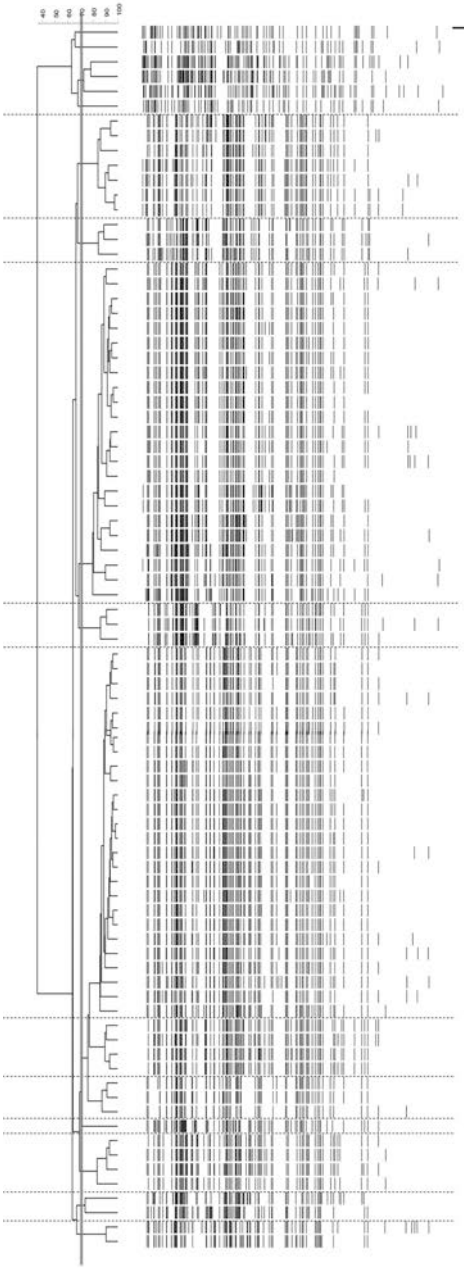
In human-associated niches, LAB can contribute to the metabolic capacities of the resident microbial ecosystem [7]. Moreover, they can interact with the host's mucosal tissues and the immune system [8]. Genotypic and phenotypic high-throughput analyses targeted several *Lactobacillus* species [4, 5]. 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.

Table 1. *L. rhamnosus* strains used in this study and availability of genome, coverage and number of contigs, origin, and phenotype data availability.

*Boldface indicates strains whose genome was included in the GTM analysis.

Strain ID	Origin	Biolog	Number of genes	Number of contigs	Coverage	Bioproject ID	Biosample ID
Lr003	Probiotic product	no					
Lr004	Unknown	no					
Lr009	Vegetable drink	yes					
Lr010	Cheese	no					
Lr026	Unknown	yes					
Lr032	**Human baby faeces**	**yes**	**3357**	**1104**	**6,5**	SUB583571	SAMN03196659
Lr035	Human faeces	no					
Lr037	Fermented milk	yes					
Lr040	Unknown	yes					
Lr044	**Cheese**	**yes**	**3344**	**2050**	**10,5**	SUB583571	SAMN03196658
Lr047	Cheese	yes					
Lr053 (NCIMB 8608)	**Unknown**	**yes**	**3260**	**1209**	**11,2**	SUB583571	SAMN03196655
Lr055 (HN001)	**Cheese**	**yes**	**2811**	**94**	**30**	PRJNA55109	SAMN02469790
Lr064	Unknown	yes					
Lr071	**Human faeces**	**yes**	**3289**	**690**	**7,9**	SUB583571	SAMN03196657
Lr072	Human faeces	no					
Lr073	**Human faeces**	**yes**	**3280**	**808**	**7,7**	SUB583571	SAMN03196656
Lr074	Human faeces	yes					
Lr075	Human faeces	yes					
Lr076	Human faeces	no					
Lr077	Human faeces	no					
Lr078	Human faeces	no					
Lr079	Human faeces	no					
Lr080	Human faeces	no					
Lr081	Human faeces	no					
Lr082	Human faeces	no					
Lr083	Human faeces	no					
Lr084	Human faeces	no					
Lr085	Human faeces	no					
Lr086	Human faeces	no					
Lr088	Human faeces	no					
Lr089	Human faeces	no					
Lr090	Human faeces	no					
Lr091	Human faeces	no					
Lr092	Human faeces	no					
Lr093	Human faeces	no					
Lr094	Human faeces	no					
Lr095	Human faeces	no					
Lr096	Human faeces	no					
Lr097	Human faeces	no					
Lr098	Human faeces	no					
Lr099	Human faeces	no					
Lr100	Human faeces	no					
Lr101	Human faeces	no					
Lr102	Human faeces	no					
Lr103	Human faeces	no					
Lr104	Human faeces	no					
Lr105	Human faeces	no					
Lr106	Human faeces	no					
Lr107	Dairy	no					
Lr108	**Human baby faeces**	**yes**	**2649**	**161**	**12,5**	SUB583571	SAMN03196652
Lr109	Human baby faeces	no					
Lr110	Goat faeces	yes					
Lr111	Goat faeces	no					
Lr132	Unknown	no					
Lr133	Human vagina	no					
Lr134	Saliva	yes					
Lr135	Vegetable probiotic drink	yes					
Lr136	Soy sauce	yes					
Lr137	Cheese	yes					
Lr138	**Cheese**	**yes**	**3085**	**585**	**8,5**	SUB583571	SAMN03196654
Lr139	Cheese	yes					
Lr140	**Fermented milk**	**yes**	**3401**	**703**	**8,2**	SUB583571	SAMN03196660
Lr141	Fermented milk	yes					
Lr142 (LGG)	**Healthy human intestine**	**yes**	**2985**	**1**	**30**	PRJNA59313	SAMEA2272375



AFLP group	Strain ID	Origin
outgroup	Lr003	*L. casei*
outgroup	LMG22920	*L. casei*
outgroup	LMG13087T	*L. casei*
outgroup	LMG11407	*L. casei*
outgroup	LMG13789	*L. casei*
outgroup	LMG9436	*L. casei*
1	Lr047	Cheese
1	Lr089	Human faeces
1	Lr047	Cheese
1	Lr072	Human faeces
1	Lr026	Unknown
1	Lr044	Cheese
1	Lr044	Cheese
2	Lr134	Human saliva
2	Lr075	Human faeces
2	Lr085	Human faeces
3	Lr140	Fermented milk
3	Lr004	Unknown
3	Lr053	NCIMB 8608, Unknown
3	Lr010	Cheese
3	Lr135	Vegetable probiotic drink
3	Lr073	Human faeces
3	Lr110	Goat faeces
3	Lr074	Human faeces
3	Lr136	Soy sauce
3	LMG6400T	Unknown, type strain derivative
3	Lr037	Fermented milk
3	Lr064	ATCC 7469 Type strain, Unknown
3	Lr091	Human faeces
3	Lr092	Human faeces
3	Lr093	Human faeces
3	Lr094	Human faeces
3	Lr095	Human faeces
3	Lr096	Human faeces
3	Lr111	Goat faeces
3	Lr132	Unknown
3	Lr133	Feminin urethra
3	Lr064	ATCC 7469 Type strain, Unknown
3	LMG10770	Human faeces
3	LMG10775	Human faeces
4	Lr032	Human baby faeces
4	Lr009	Vegetable drink
4	LMG12166	Cheese
5	Lr035	Human faeces
5	Lr142	GG strain, healthy human intestine
5	Lr040	Unknown
5	Lr080	Human faeces
5	Lr081	Human faeces
5	Lr088	Human faeces
5	Lr082	Human faeces
5	Lr090	Human faeces
5	Lr103	Human faeces
5	Lr090	Human faeces
5	Lr090	Human faeces
5	Lr105	Human faeces
5	Lr103	Human faeces
5	Lr098	Human faeces
5	Lr090	Human faeces
5	Lr101	Human faeces
5	Lr106	Human faeces
5	Lr090	Human faeces
5	Lr099	Human faeces
5	Lr077	Human faeces
5	Lr090	Human faeces
5	Lr090	Human faeces
5	Lr086	Human faeces
5	Lr083	Human faeces
5	Lr076	Human faeces
5	Lr100	Human faeces
6	Lr078	Human faeces
6	Lr084	Human faeces
6	Lr097	Human faeces
6	Lr104	Human faeces
7	Lr071	Human faeces
7	Lr102	Human faeces
7	Lr079	Human faeces
8	Lr055	HN001, Cheese
9	Lr141	Fermented milk
9	Lr137	Cheese
9	Lr139	Cheese
9	LMG18030	Yogurt
10	Lr138	Cheese
10	Lr107	Dairy
11	Lr108	Human baby faeces
11	Lr109	Human baby faeces

Figure 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	Relevant feature(s)	Reference(s)
pIL253	Ery^R; cloning vector; medium copy in gram positive bacteria	[1]
pSOR253	Ery^R; pIL253 derivative containing the L-Sorbose candidate operon	This study
pAMG253	Ery^R; pIL253 derivative containing the α-Methyl-D-Glucoside candidate operon	This study

Ery^r, erythromycin resistant

Primer	Sequence (5′ to 3′)
Primer E01	GACTGCGTACCAATTCA
Primer T11	GTTTCTTATGAGTCCTGACCGAA
Primer T13	GTTTCTTATGAGTCCTGACCGAG
Adaptor (5' to 3') EcoRI	CTCGTAGACTGCGTACC
Adaptor (3' to 5') EcoRI	CTGACGCATGGTTAA
Adaptor (5' to 3') TaqI	GACGATGAGTCCTGAC
Adaptor (3' to 5') TaqI	TACTCAGGACTGGC
αMetDglu_F	CCGGATCGGGAAACTGAT
αMetDglu_R	CGAACACAATCTTGCCGTTC
LSor_F	CAATTTCTTCTCTGCATCGC
LSor_R	AGTATCCATATTGGCATCCG

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 metabolization of specific carbon sources, which were validated by

genetic complementation of genes encoding proteins/enzymes for L-sorbose and α-methyl-glucoside utilization.

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 ± 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-dioxiadenosine and α-methyl-D-galactoside). Carbohydrate cluster 2 contained 38 carbon sources with a relatively high average level of growth (114 ± 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, α-D-glucose, L-lyxose, salicin, D-mannitol, L-arabinose, 2-deoxy-D-ribose, and 3-*O*-α-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-γ-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, α-methyl-D-glucoside, β-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*-α-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, α- and β-methyl-D-glucosides, turanose, maltitol, arbutin, mannitol, trehalose, and tagatose but not cellobiose, D-galactonic acid-γ-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).

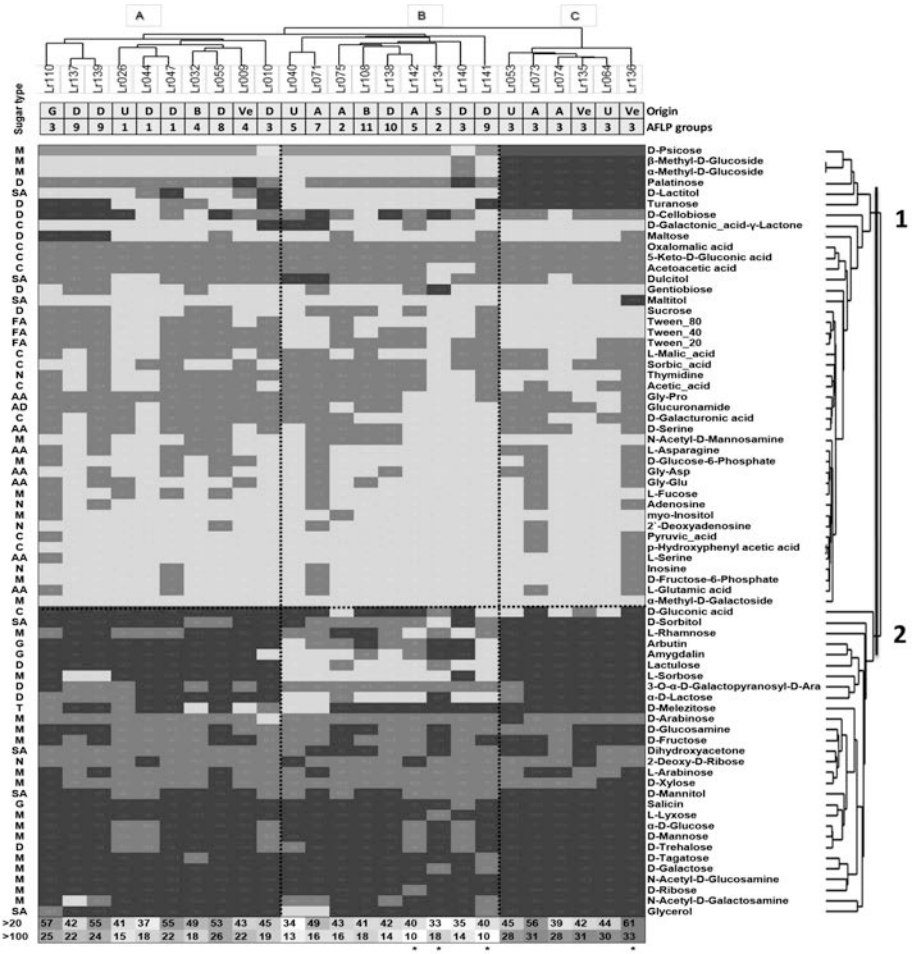


Figure 2 Heat map and clustering of *L. rhamnosus* strains and their substrates based on growth endpoint data (Phenotype MicroArray Biolog data) for carbohydrates in clusters 1 and 2. The shading of the heat map refers to the growth level: black, >100 OU; dark gray, 100 to 20 OU; light gray, _20 OU. 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 below 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 20OUand above 100OU for each *L. rhamnosus* strain, and the highest and lowest values are marked with an asterisk.

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 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 non-utilizing (<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, α-methyl-D-glucoside, 3-*O*-β-D-galactopyranosyl-D-arabinose, L-sorbose, dulcitol, and D-galactonic acid-γ-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 α-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, α-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 α-methyl-D-glucoside utilization), respectively.

Table 3 Genes selected for expression by complementation after the GTM analysis for α-methyl-D-glucoside and L-sorbose. [a] MFS, major facilitator superfamily.

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*a* transporter	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[a] transporter BenK	No	GG
6.145	α-Methyl-D-glucoside	4	Transporter	LC705_02793	GG
6.145	α-Methyl-D-glucoside	4	Hypothetical protein	LC705_02792	GG
6.145	α-Methyl-D-glucoside	4	ATPase	LC705_02791	GG

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 contrast to the strain transformed with the empty pIL253 vector, the pSOR253- and pAMG253-harboring derivatives were able to grow on sorbose and α-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.

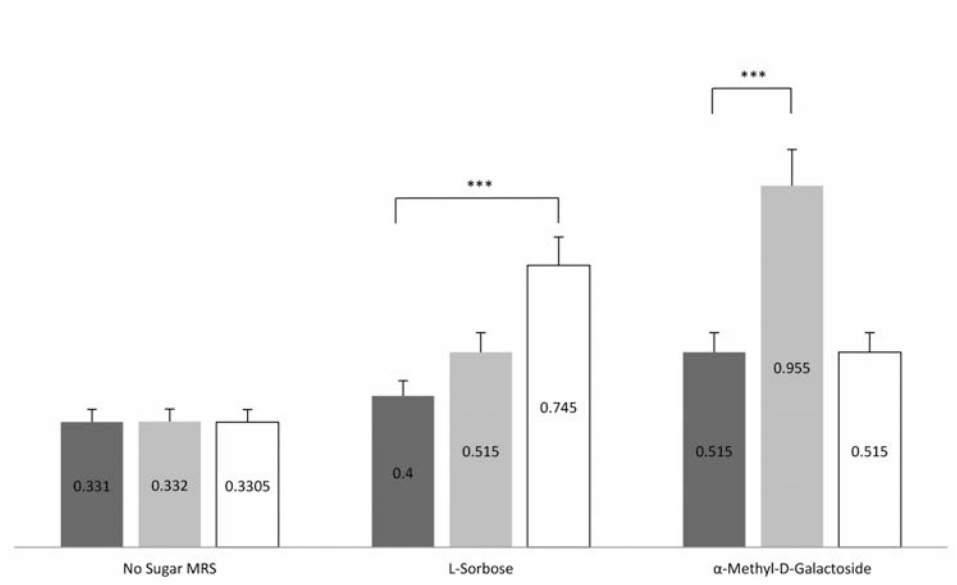


Figure 3. Growth of the wild type GG strain - dark grey - and expression strains M12 (transformed with pSOR253 containing the L-Sorbose operon) – white - and M17 (transformed with pAMG253 containing the α-Methyl-D-Glucoside operon) – light grey - on L-Sorbose and α-Methyl-D-Glucoside in batch cultures. The numbers represent OD values, per mL, for the cultures, in sugar-free MRS supplemented with either L-Sorbose or α-Methyl-D-Glucoside.

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 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 α-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.

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-Sharpe (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 μg · 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_{600}] 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 μl 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-carboxyfluorescein [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 (Applera, 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 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 re-annotated 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**. 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 threshold 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 co-occurrence 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 α-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 α-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 ScaI-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 μg · 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 α-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 α-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), JUIN00000000 (*Lactobacillus rhamnosus* Lr032), and JUIO00000000 (*Lactobacillus rhamnosus* Lr140) (Table 1). The versions described in this paper have the numbers "XXXX01000000," where XXXX represents the first four letters of each accession number.

References:

1. Barraclough T, Balbi K, Ellis R. 2012. *Evolving concepts of bacterial species.* Evol Biol 39:148–157.

2. Di Cagno R. et al. 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.

3. Moraes P. et al. 2013. *Comparison of phenotypic and molecular tests to identify lactic acid bacteria.* Braz J Microbiol 44:109–112.

4. Siezen RRJ. et al. 2010. *Phenotypic and genomic diversity of Lactobacillus plantarum strains isolated from various environmental niches.* Environ Microbiol.

5. Smokvina T. et al. 2013. *Lactobacillus paracasei comparative genomics: towards species pan-genome definition and exploitation of diversity.* PLoS One 8:e68731.

6. Goh Y. et al. *Specialized adaptation of a lactic acid bacterium to the milk environment: the comparative genomics of Streptococcus thermophilus LMD-9.* Microb Cell Fact 10:S22.

7. Den Besten G. et al. 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.

8. 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.

9. Pretzer G. et al. 2005. *Biodiversity-based identification and functional characterization of the mannose-specific adhesin of Lactobacillus plantarum.* J Bacteriol 187:6128–6136.

10. Adlerberth I. et al. 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.

11. Gross G. et al. 2008. *Mannose-specific interaction of Lactobacillus plantarum with porcine jejunal epithelium.* FEMS Immunol Med Microbiol 54:215–223.

12. Wickens K. et al. 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.

13. Hummelen R. et al. 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.

14. 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.

15. Kankainen M. at al. 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.

16. Pittet V, Ewen E, Bushell BR, Ziola B. 2012. *Genome sequence of Lactobacillus rhamnosus ATCC 8530.* J Bacteriol 194:726.

17. Yu B. et al. 2011. *Genome sequence of Lactobacillus rhamnosus strain CASL, an efficient L-lactic acid producer from cheap substrate cassava.* J Bacteriol 193:7013–7014.

18. Douillard FP. et al. 2013. *Comparative genomic and functional analysis of 100 Lactobacillus rhamnosus strains and their comparison with strain GG.* PLoS Genet 9:e1003683.

19. Gevers D, Huys G, Swings J. 2001. *Applicability of rep-PCR fingerprinting for identification of Lactobacillus species.* FEMS Microbiol Lett 205:31–36.

20. Vos P. et al. 1995. *AFLP: a new technique for DNA fingerprinting.* Nucleic Acids Res 23:4407–4414.

21. Aziz RK. et al. 2008. *The RAST server: rapid annotations using subsystems technology.* BMC Genomics 9:75. 10.1186/1471-2164-9-75.

22. Chen F, Mackey AJ, Stoeckert CJ, Roos DS. 2006. *OrthoMCL-DB: querying a comprehensive multi-species collection of ortholog groups.* Nucleic Acids Res 34:D363–D368.

23. 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.

24. Bochner BR, Gadzinski P, Panomitros E. 2001. *Phenotype microarrays for high-throughput phenotypic testing and assay of gene function.* Genome Res 11:1246–1255.

25. Hammer Ø, Harper DAT, Ryan PD. 2001. *PAST: paleontological statistics software package for education and data analysis.* Palaeontol Electron 4:1–9.

26. Enke H. 1979. *The information in contingency tables.* Biom J 21:94–95.

27. Brown RP. 1997. Testing for heterogeneity among phenotypic correlations: a comparison of methods using Monte Carlo simulations. Genetica 101:67–74.

28. McHugh ML. 2013. *The chi-square test of independence.* Biochem Med (Zagreb) 23:143–149.

29. Tohsato Y, Mori H. 2008. *Phenotype profiling of single gene deletion mutants of E. coli using Biolog technology.* Genome Inform 21:42–52.

30. Toh H. et al. 2013. *Genomic adaptation of the Lactobacillus casei group.* PLoS One 8:e75073.

31. 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.

32. 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.

33. McLeod A. et al. 2008. *Diversity of Lactobacillus sakei strains investigated by phenotypic and genotypic methods.* Syst Appl Microbiol 31:393–403.

34. Vancanneyt M. et al. 2006. *Intraspecific genotypic characterization of Lactobacillus rhamnosus strains intended for probiotic use and isolates of human origin.* Appl Environ Microbiol 72:5376–5383.

35. Oh PL. et al. 2010. *Diversification of the gut symbiont Lactobacillus reuteri as a result of host-driven evolution.* ISME J 4:377–387.

36. Tanigawa K, Watanabe K. 2011. *Multilocus sequence typing reveals a novel subspeciation of Lactobacillus delbrueckii.* Microbiology 157:727–738.

37. Gancheva A. et al. 1999. *A polyphasic approach towards the identification of strains belonging to Lactobacillus acidophilus and related species.* Syst Appl Microbiol 22:573–585.

38. Torriani S. et al. 2001. *Differentiation of Lactobacillus plantarum, L. pentosus and L. paraplantarum species by RAPD-PCR and AFLP.* Syst Appl Microbiol 24:554–560.

39. Siezen RJ. et al. 2011. *Genome-scale diversity and niche adaptation analysis of Lactococcus lactis by comparative genome hybridization using multi-strain arrays.* Microb Biotechnol 4:383–402.

40. Kütahya OE. et al. 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.

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.

42. Succi M. et al. 2005. *Bile salt and acid tolerance of Lactobacillus rhamnosus strains*

isolated from Parmigiano Reggiano cheese. FEMS Microbiol Lett 244:129–137.

43. Flint HJ. et al. 2012. *Microbial degradation of complex carbohydrates in the gut.* Gut Microbes 3:289–306.

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. et al. 2009. *Genome sequence and comparative genome analysis of Lactobacillus casei: insights into their niche-associated evolution.* Genome Biol Evol 1:239–257.

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.

47. Broadbent JR. et al. 2012. *Analysis of the Lactobacillus casei supragenome and its influence in species evolution and lifestyle adaptation.* BMC Genomics 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.

49. Chiaramonte F. et al 2010. *Analysis of Lactobacillus sakei mutants selected after adaptation to the gastrointestinal tracts of axenic mice.* Appl Environ Microbiol 76:2932–2939.

50. Douillard FP. et al. 2013. *Comparative genomic and functional analysis of Lactobacillus casei and Lactobacillus rhamnosus strains marketed as probiotics.* Appl Environ Microbiol 79:1923–1933.

51. Bayjanov JR. et al. 2013. *Genotype-phenotype matching analysis of 38 Lactococcus lactis strains using random forest methods.* BMC Microbiol 13:68.

52. Khatri B. et al. 2013. *High throughput phenotypic analysis of Mycobacterium tuberculosis and Mycobacterium bovis strains' metabolism using biolog phenotype microarrays.* PLoS One 8:e52673.

53. Bachmann H. et al. 2009. *Regulatory phenotyping reveals important diversity within the species Lactococcus lactis.* Appl Environ Microbiol 75:5687–5694.

54. Bayjanov JR. et al. 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.

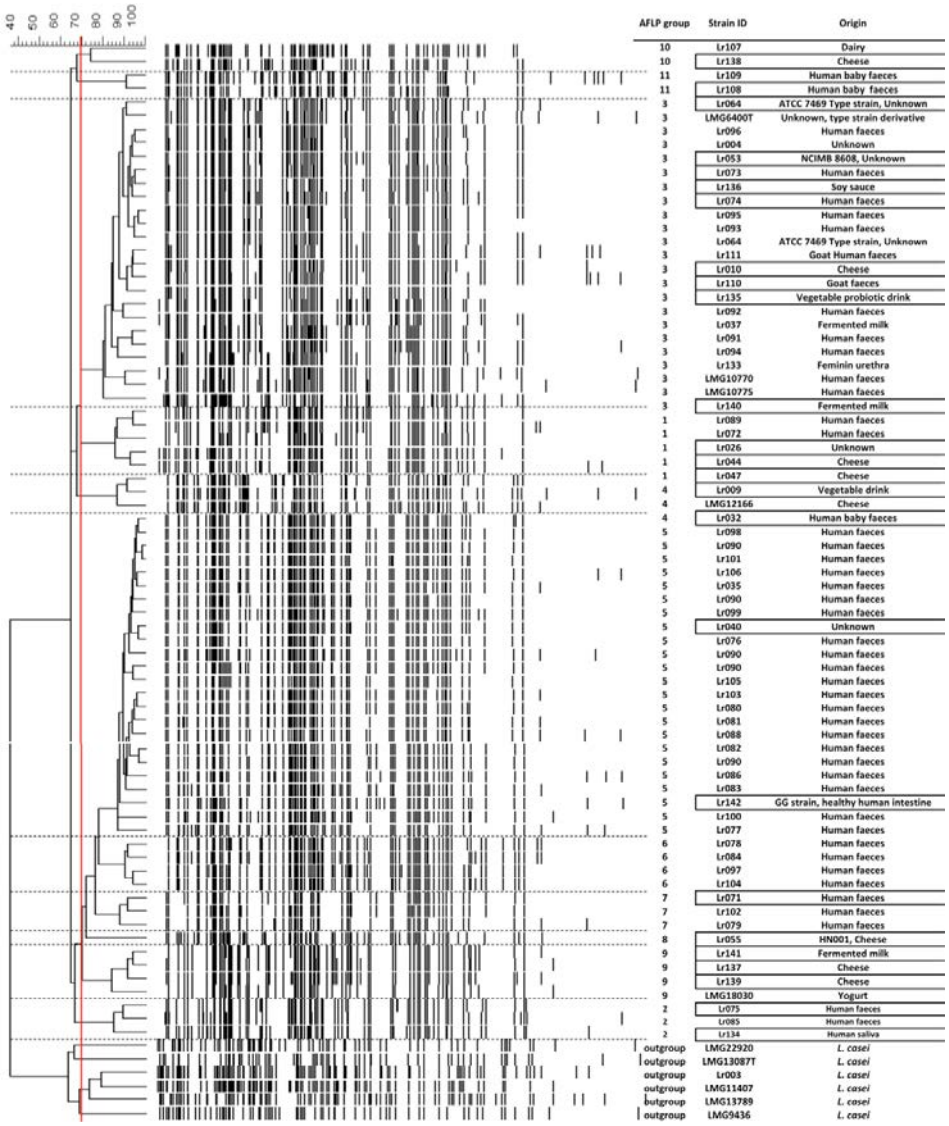
55. Bachmann H. et al. 2012. *Microbial domestication signatures of Lactococcus lactis can be reproduced by experimental evolution.* Genome Res 22:115–124.

56. Van Bokhorst-van de Veen H. et al. 2012. *Congruent strain specific intestinal persistence of Lactobacillus plantarum in an intestine-mimicking in vitro system and in human volunteers.* PLoS One 7:e44588.

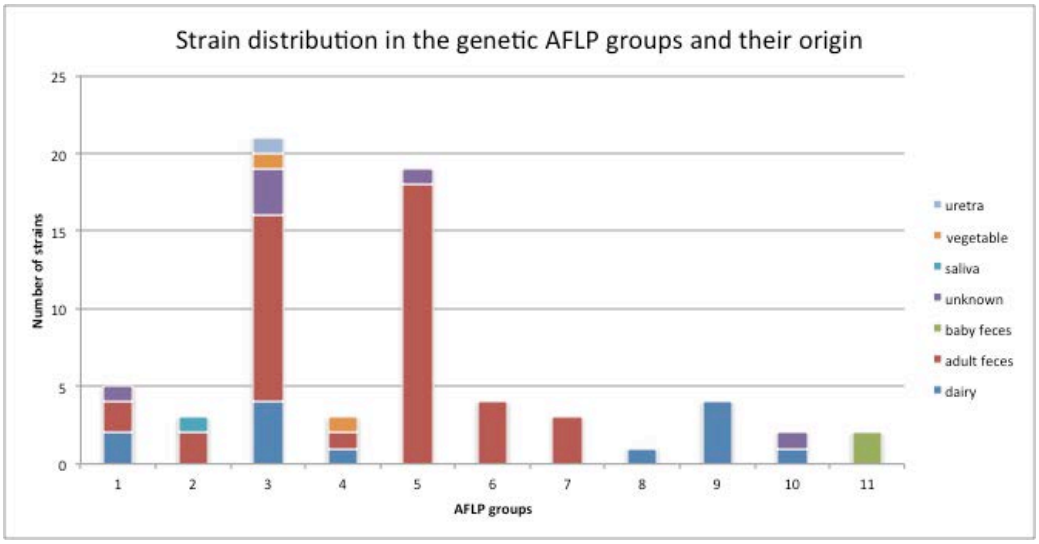
57. Brandsma JB. et al. 2012. *Arginine metabolism in sugar deprived Lactococcus lactis enhances survival and cellular activity, while supporting flavour production.* Food Microbiol 29:27–32.

58. Dhaisne A. et al. 2013. *Genotypic and phenotypic analysis of dairy Lactococcus lactis biodiversity in milk: volatile organic compounds as discriminating markers.* Appl Environ Microbiol 79:4643–4652.

Appendix 1. Supplementary data:



Supplementary Figure 1. Dendrogram based on the analysis of AFLP patterns of the primer combination E01/T11, with visualization of the banding patterns. The strains origin and clustering (at a similarity level of 70%) is represented in the columns. The 25 strains selected for Omnilog growth experiments are highlighted with a box. The type strain ATCC7469 was obtained from both the Danone and Numico culture collections and appears twice in the analysis.



Supplementary Figure 2. Strain distribution in the genetic AFLP groups and their origin.

Supplementary Figure 3. Clustering of *L. rhamnosus* strains and their substrates based on raw kinetic data (Phenotype MicroArray™ Biolog data) for all carbohydrates. The clustering of the strains and substrates was created in PAST using the Euclidian distance matrix based on UPGMA clustering. The colouring of the clusters refers to metabolic groups 1 (green), 2 (blue) and 3 (orange). Available for download at library.wur.nl.

Supplementary Table 1. Sørensen–Dice similarity matrix of strains against all other strains using profiles created from Biolog data. Values of 0.95 or higher represent a significant level of similarity. The results are the basis of the MG grouping and are also reflected in the Biolog based clustering (Figure 2).

AFLP Group	1	2	**3**	4	**5**	**6**	**7**	8	**9**	10	**11**	Sum	Monte Carlo p (p<0.05)	Cramer's V (0 to 1)
adult faeces	2	2	11	0	18	4	3	0	0	0	0	40	**2.71E-05**	**0.60**
baby faeces	0	0	0	1	0	0	0	0	0	0	2	3	0.23	0.422
dairy	2	0	4	1	0	0	0	1	4	1	0	13	**0.13**	**0.43**
goat	0	0	1	0	0	0	0	0	0	0	0	1	1	0.36
saliva	0	1	0	0	0	0	0	0	0	0	0	1	1	0.36
unknown	1	0	5	0	1	0	0	0	0	1	0	8	**0.024**	**0.50**
uretra	0	0	1	0	0	0	0	0	0	0	0	1	1	0.36
vegetable	0	0	1	1	0	0	0	0	0	0	0	2	1	0.34
Sum	5	3	23	3	19	4	3	1	4	2	2		Total for the dataset	
Monte Carlo p (p<0.05)	0.24	0.31	**0.04**	1	**1E-04**	**8E-04**	**0.011**	1	**8E-04**	1	**0.11**			
Cramer's V (0 to 1)	0.34	0.36	**0.47**	0.26	**0.75**	**0.54**	**0.51**	0.31	**0.54**	0.28	**0.41**		**0.012**	**0.53**

Supplementary Table 2. Statistical analysis on the frequencies of strains for each origin type and AFLP group. The significance of association between the nominal variables was quantified using two methods. The Monte Carlo randomisation 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. Cramer's V is a chi-square-based measure of variable independence, giving a value between 0 (statistical independence) and 1 (associated variables). Both calculations were performed in the PAST software.

Strain Similarity	Lr064	Lr009	Lr010	Lr026	Lr032	Lr040	Lr044	Lr047	Lr053	Lr055	Lr071	Lr073	Lr074	Lr075	Lr108	Lr110	Lr134	Lr135	Lr136	Lr137	Lr138	Lr139	Lr140	Lr141	Lr142
Lr064		0.92	0.86	0.89	0.90	0.82	0.92	0.91	0.95	0.91	0.85	0.93	0.94	0.87	0.90	0.91	0.89	0.96	0.94	0.90	0.87	0.94	0.90	0.88	0.85
Lr009	0.92		0.91	0.94	0.96	0.87	0.89	0.95	0.92	0.94	0.94	0.93	0.89	0.88	0.90	0.96	0.85	0.89	0.94	0.92	0.90	0.91	0.87	0.91	0.87
Lr010	0.86	0.91		0.92	0.94	0.90	0.87	0.93	0.88	0.94	0.93	0.91	0.85	0.88	0.91	0.94	0.86	0.87	0.92	0.93	0.91	0.90	0.82	0.92	0.90
Lr026	0.89	0.94	0.92		0.94	0.86	0.90	0.94	0.89	0.95	0.93	0.92	0.85	0.92	0.91	0.95	0.88	0.90	0.93	0.94	0.91	0.93	0.88	0.92	0.91
Lr032	0.90	0.96	0.94	0.94		0.90	0.87	0.96	0.91	0.97	0.96	0.94	0.89	0.89	0.94	0.97	0.86	0.87	0.95	0.94	0.93	0.95	0.84	0.95	0.90
Lr040	0.82	0.87	0.90	0.86	0.90		0.79	0.90	0.84	0.90	0.93	0.88	0.83	0.85	0.89	0.90	0.80	0.79	0.88	0.89	0.92	0.88	0.78	0.93	0.84
Lr044	0.92	0.89	0.87	0.90	0.87	0.79		0.88	0.93	0.89	0.83	0.91	0.90	0.89	0.88	0.89	0.89	0.95	0.91	0.89	0.82	0.90	0.91	0.85	0.88
Lr047	0.91	0.95	0.93	0.94	0.96	0.90	0.88		0.92	0.98	0.94	0.98	0.90	0.90	0.92	0.99	0.84	0.88	0.97	0.92	0.92	0.94	0.84	0.93	0.88
Lr053	0.95	0.92	0.88	0.89	0.91	0.84	0.93	0.92		0.91	0.87	0.95	0.94	0.87	0.90	0.93	0.87	0.94	0.95	0.87	0.87	0.92	0.89	0.88	0.84
Lr055	0.91	0.94	0.94	0.95	0.97	0.90	0.89	0.98	0.91		0.95	0.96	0.89	0.91	0.93	0.98	0.87	0.89	0.96	0.94	0.93	0.96	0.85	0.95	0.90
Lr071	0.85	0.94	0.93	0.93	0.96	0.93	0.83	0.94	0.87	0.95		0.92	0.84	0.90	0.90	0.95	0.82	0.83	0.92	0.90	0.94	0.92	0.82	0.93	0.87
Lr073	0.93	0.93	0.91	0.92	0.94	0.88	0.91	0.98	0.95	0.96	0.92		0.91	0.91	0.93	0.97	0.85	0.91	0.99	0.90	0.90	0.96	0.85	0.91	0.86
Lr074	0.94	0.89	0.85	0.85	0.89	0.83	0.90	0.90	0.94	0.89	0.84	0.91		0.85	0.89	0.90	0.86	0.91	0.91	0.87	0.84	0.90	0.89	0.85	0.82
Lr075	0.87	0.88	0.88	0.92	0.89	0.85	0.89	0.90	0.87	0.91	0.90	0.91	0.85		0.90	0.91	0.87	0.85	0.92	0.90	0.88	0.92	0.86	0.91	0.87
Lr108	0.90	0.90	0.91	0.91	0.94	0.89	0.88	0.92	0.90	0.93	0.90	0.93	0.89	0.90		0.93	0.91	0.88	0.94	0.95	0.94	0.95	0.89	0.94	0.91
Lr110	0.91	0.96	0.94	0.95	0.97	0.90	0.89	0.99	0.93	0.98	0.95	0.97	0.90	0.91	0.93		0.85	0.89	0.98	0.93	0.93	0.95	0.85	0.94	0.89
Lr134	0.89	0.85	0.86	0.88	0.86	0.80	0.89	0.84	0.87	0.87	0.82	0.85	0.86	0.87	0.91	0.85		0.89	0.86	0.92	0.85	0.89	0.91	0.86	0.87
Lr135	0.96	0.89	0.87	0.90	0.87	0.79	0.95	0.88	0.94	0.89	0.83	0.91	0.91	0.85	0.88	0.89	0.89		0.91	0.89	0.82	0.90	0.89	0.85	0.84
Lr136	0.94	0.94	0.92	0.93	0.95	0.88	0.91	0.97	0.95	0.96	0.92	0.99	0.91	0.92	0.94	0.98	0.86	0.91		0.91	0.91	0.97	0.86	0.91	0.86
Lr137	0.90	0.92	0.93	0.94	0.94	0.89	0.89	0.92	0.87	0.94	0.90	0.90	0.87	0.90	0.95	0.93	0.92	0.89	0.91		0.92	0.94	0.86	0.94	0.92
Lr138	0.87	0.90	0.91	0.91	0.93	0.92	0.82	0.92	0.87	0.93	0.94	0.90	0.84	0.88	0.94	0.93	0.85	0.82	0.91	0.92		0.92	0.85	0.94	0.92
Lr139	0.94	0.91	0.90	0.93	0.95	0.88	0.90	0.94	0.92	0.96	0.92	0.96	0.90	0.92	0.95	0.95	0.89	0.90	0.97	0.94	0.92		0.87	0.93	0.89
Lr140	0.90	0.87	0.82	0.88	0.84	0.78	0.91	0.84	0.89	0.85	0.82	0.85	0.89	0.86	0.89	0.85	0.91	0.89	0.86	0.86	0.85	0.87		0.84	0.89
Lr141	0.88	0.91	0.92	0.92	0.95	0.93	0.85	0.93	0.88	0.95	0.93	0.91	0.85	0.91	0.94	0.94	0.86	0.85	0.91	0.94	0.94	0.93	0.84		0.90
Lr142	0.85	0.87	0.90	0.91	0.90	0.84	0.88	0.88	0.84	0.90	0.87	0.86	0.82	0.87	0.91	0.89	0.87	0.84	0.86	0.92	0.92	0.89	0.89	0.90	

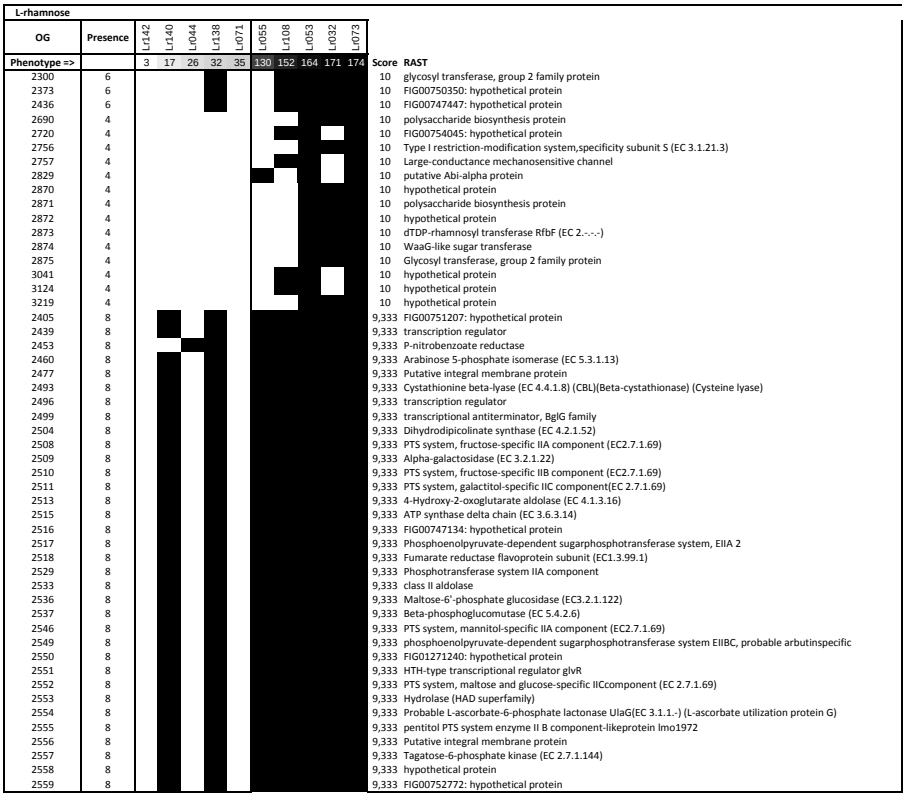
Supplementary Table 3. Gene trait matching candidates list for L-Sorbose. Black represents presence and white absence of the OG from the same row from the strain in the same column.

L-Sorbose													
OG	Presence	Lr138	Lr108	Lr142	Lr071	Lr140	Lr055	Lr044	Lr032	Lr073	Lr053		
Phenotype =>		0	0	2	3.5	118	147	148	152	156	158	Score	RAST
2298	6					■	■		■	■	■	17.5	FIG00751110: hypothetical protein
137	6					■		■	■	■	■	17.5	FIG00754355: hypothetical protein
2455	**6**					■		■	■	■	■	**17.5**	**benzoate MFS transporter BenK**
2506	**6**					■		■	■	■	■	**17.5**	**FIG00750163: hypothetical protein**
2541	**6**					■		■	■	■	■	**17.5**	**Sugar diacid utilization regulator**
2542	**6**					■		■	■	■	■	**17.5**	**Uridine phosphorylase (EC 2.4.2.3)**
2543	**6**					■		■	■	■	■	**17.5**	**Phosphopentomutase (EC 5.4.2.7)**
2568	**6**					■		■	■	■	■	**17.5**	**benzoate MFS transporter BenK**
2808	6					■			■	■	■	17.5	hypothetical protein
2290	8				■	■	■	■	■	■	■	16	hypothetical protein

Supplementary Table 4. Gene trait matching candidates list for α-Methyl-D-Glucoside. Black represents presence and white absence of the OG from the same row from the strain in the same column.

α-Methyl-D-Glucoside													
OG	Presence	Lr142	Lr138	Lr032	Lr055	Lr071	Lr108	Lr044	Lr140	Lr073	Lr053		
Phenotype =>		0	0	0	0	0	**0,5**	**0,5**	47	**138**	**161**	Score	RAST
2730	**4**								■	■	■	**40,00**	**family 43 glycoside hydrolase**
2751	**4**								■	■	■	**40,00**	**hypothetical protein**
2752	**4**								■	■	■	**40,00**	**oligogalacturonide transporter**
2747	4								■	■	■	40,00	FIG00753430: hypothetical protein
2748	4								■	■	■	40,00	Cellobiose phosphotransferase system YdjC-likeprotein
2749	4								■	■	■	40,00	PTS system, maltose and glucose-specific IICcomponent (EC 2.7.1.69)
2750	4								■	■	■	40,00	6-phospho-beta-glucosidase (EC 3.2.1.86)
2492	4								■	■	■	40,00	FIG00752024: hypothetical protein
2560	4								■	■	■	40,00	transcriptional regulator, Cro
2588	4								■	■	■	40,00	Maltodextrin glucosidase (EC 3.2.1.20)
2606	4								■	■	■	40,00	PTS system, beta-glucoside-specific IIBcomponent (EC 2.7.1.69)
2612	4								■	■	■	40,00	Oligo-1,6-glucosidase (EC 3.2.1.10)
2613	4								■	■	■	40,00	hypothetical protein
2616	4								■	■	■	40,00	PTS system transporter subunit IID
2617	4								■	■	■	40,00	PTS system transporter subunit IIC
2618	4								■	■	■	40,00	PTS system transporter subunit IIB
2619	4								■	■	■	40,00	Ribose operon repressor
2686	4								■	■	■	40,00	modification methylase (Cytosine-specificmethyltransferase
2688	4								■	■	■	40,00	hypothetical protein
2717	4								■	■	■	40,00	Glycosyltransferase
2753	4								■	■	■	40,00	abortive infection protein AbiGI
2754	4								■	■	■	40,00	Predicted outer membrane protein
2755	4								■	■	■	40,00	putative glycosyltransferase
2798	4								■	■	■	40,00	FIG00754353: hypothetical protein
2799	4								■	■	■	40,00	ATP-dependent nuclease, subunit B
2824	4								■	■	■	40,00	site-specific recombinase, phage integrasefamily
2826	4								■	■	■	40,00	hypothetical protein
2827	4								■	■	■	40,00	FIG045374: Type II restriction enzyme,methylase subunit YeeA
2867	4								■	■	■	40,00	hypothetical protein
2929	4								■	■	■	40,00	glycosyl hydrolase, group 13
2981	4								■	■	■	40,00	hypothetical protein
2984	4								■	■	■	40,00	hypothetical protein
2988	4								■	■	■	40,00	hypothetical protein
3046	4								■	■	■	40,00	hypothetical protein

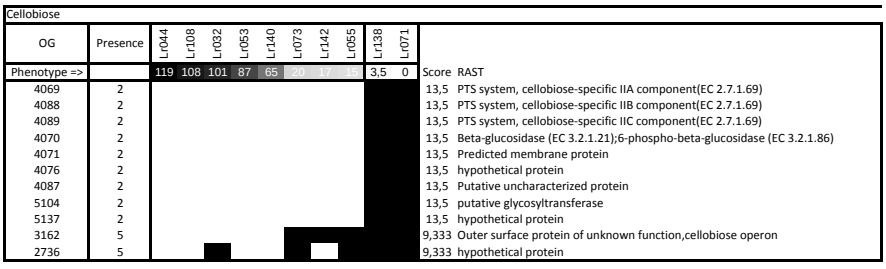
Supplementary Table 5. Gene trait matching candidates list for L-rhamnose. Black represents presence and white absence of the OG from the same row from the strain in the same column.



L-rhamnose

OG	Presence	Lr142	Lr140	Lr044	Lr138	Lr071	Lr055	Lr108	Lr053	Lr032	Lr073	Score	RAST
Phenotype =>		3	17	26	32	35	130	152	164	171	174		
2300	6				■			■	■	■		10	glycosyl transferase, group 2 family protein
2373	6				■			■	■	■		10	FIG00750350: hypothetical protein
2436	6				■			■	■	■		10	FIG00747447: hypothetical protein
2690	4							■	■	■		10	polysaccharide biosynthesis protein
2720	4							■	■		■	10	FIG00754045: hypothetical protein
2756	4							■	■	■		10	Type I restriction-modification system,specificity subunit S (EC 3.1.21.3)
2757	4							■	■		■	10	Large-conductance mechanosensitive channel
2829	4						■	■	■	■		10	putative Abi-alpha protein
2870	4								■	■	■	10	hypothetical protein
2871	4								■	■	■	10	polysaccharide biosynthesis protein
2872	4								■	■	■	10	hypothetical protein
2873	4								■	■	■	10	dTDP-rhamnosyl transferase RfbF (EC 2.-.-.-)
2874	4								■	■	■	10	WaaG-like sugar transferase
2875	4								■	■	■	10	Glycosyl transferase, group 2 family protein
3041	4							■	■		■	10	hypothetical protein
3124	4							■	■	■	■	10	hypothetical protein
3219	4							■	■	■	■	10	hypothetical protein
2405	8	■			■		■	■	■	■	■	9,333	FIG00751207: hypothetical protein
2439	8	■			■		■	■	■	■	■	9,333	transcription regulator
2453	8		■		■		■	■	■	■	■	9,333	P-nitrobenzoate reductase
2460	8	■			■		■	■	■	■	■	9,333	Arabinose 5-phosphate isomerase (EC 5.3.1.13)
2477	8	■			■		■	■	■	■	■	9,333	Putative integral membrane protein
2493	8	■			■		■	■	■	■	■	9,333	Cystathionine beta-lyase (EC 4.4.1.8) (CBL)(Beta-cystathionase) (Cysteine lyase)
2496	8	■			■		■	■	■	■	■	9,333	transcription regulator
2499	8	■			■		■	■	■	■	■	9,333	transcriptional antiterminator, BglG family
2504	8	■			■		■	■	■	■	■	9,333	Dihydrodipicolinate synthase (EC 4.2.1.52)
2508	8	■			■		■	■	■	■	■	9,333	PTS system, fructose-specific IIA component (EC2.7.1.69)
2509	8	■			■		■	■	■	■	■	9,333	Alpha-galactosidase (EC 3.2.1.22)
2510	8	■			■		■	■	■	■	■	9,333	PTS system, fructose-specific IIB component (EC2.7.1.69)
2511	8	■			■		■	■	■	■	■	9,333	PTS system, galactitol-specific IIC component(EC 2.7.1.69)
2513	8	■			■		■	■	■	■	■	9,333	4-Hydroxy-2-oxoglutarate aldolase (EC 4.1.3.16)
2515	8	■			■		■	■	■	■	■	9,333	ATP synthase delta chain (EC 3.6.3.14)
2516	8	■			■		■	■	■	■	■	9,333	FIG00747134: hypothetical protein
2517	8	■			■		■	■	■	■	■	9,333	Phosphoenolpyruvate-dependent sugarphosphotransferase system, EIIA 2
2518	8	■			■		■	■	■	■	■	9,333	Fumarate reductase flavoprotein subunit (EC1.3.99.1)
2529	8	■			■		■	■	■	■	■	9,333	Phosphotransferase system IIA component
2533	8	■			■		■	■	■	■	■	9,333	class II aldolase
2536	8	■			■		■	■	■	■	■	9,333	Maltose-6'-phosphate glucosidase (EC3.2.1.122)
2537	8	■			■		■	■	■	■	■	9,333	Beta-phosphoglucomutase (EC 5.4.2.6)
2546	8	■			■		■	■	■	■	■	9,333	PTS system, mannitol-specific IIA component (EC2.7.1.69)
2549	8	■			■		■	■	■	■	■	9,333	phosphoenolpyruvate-dependent sugarphosphotransferase system EIIBC, probable arbutinspecific
2550	8	■			■		■	■	■	■	■	9,333	FIG01271240: hypothetical protein
2551	8	■			■		■	■	■	■	■	9,333	HTH-type transcriptional regulator glvR
2552	8	■			■		■	■	■	■	■	9,333	PTS system, maltose and glucose-specific IICcomponent (EC 2.7.1.69)
2553	8	■			■		■	■	■	■	■	9,333	Hydrolase (HAD superfamily)
2554	8	■			■		■	■	■	■	■	9,333	Probable L-ascorbate-6-phosphate lactonase UlaG(EC 3.1.1.-) (L-ascorbate utilization protein G)
2555	8	■			■		■	■	■	■	■	9,333	pentitol PTS system enzyme II B component-likeprotein lmo1972
2556	8	■			■		■	■	■	■	■	9,333	Putative integral membrane protein
2557	8	■			■		■	■	■	■	■	9,333	Tagatose-6-phosphate kinase (EC 2.7.1.144)
2558	8	■			■		■	■	■	■	■	9,333	hypothetical protein
2559	8	■			■		■	■	■	■	■	9,333	FIG00752772: hypothetical protein

Supplementary Table 6. Gene trait matching candidates list for cellobiose. Black represents presence and white absence of the OG from the same row from the strain in the same column.



Cellobiose

OG	Presence	Lr044	Lr108	Lr032	Lr053	Lr140	Lr073	Lr142	Lr055	Lr138	Lr071	Score	RAST
Phenotype =>		119	108	101	87	65	29	17	15	3,5	0		
4069	2									■	■	13,5	PTS system, cellobiose-specific IIA component(EC 2.7.1.69)
4088	2									■	■	13,5	PTS system, cellobiose-specific IIB component(EC 2.7.1.69)
4089	2									■	■	13,5	PTS system, cellobiose-specific IIC component(EC 2.7.1.69)
4070	2									■	■	13,5	Beta-glucosidase (EC 3.2.1.21);6-phospho-beta-glucosidase (EC 3.2.1.86)
4071	2									■	■	13,5	Predicted membrane protein
4076	2									■	■	13,5	hypothetical protein
4087	2									■	■	13,5	Putative uncharacterized protein
5104	2									■	■	13,5	putative glycosyltransferase
5137	2									■	■	13,5	hypothetical protein
3162	5						■	■	■	■	■	9,333	Outer surface protein of unknown function,cellobiose operon
2736	5			■			■		■	■	■	9,333	hypothetical protein

CHAPTER 4

L-fucose metabolization in *Lactobacillus rhamnosus*: identification and analysis of the operon

Submitted manuscript.

Authors: Corina Ceapa, Roger S. Bongers, Loo Wee Chia, Jolanda M. Lambert, Jan Knol, Michiel Kleerebezem

Abstract

L-fucose utilization capacity can contribute to bacterial persistence in the mammalian gastrointestinal tract, where this carbohydrate is significantly present as a mucus component. By phenotype - genotype correlation analysis for 25 strains of *L. rhamnosus*, an operon containing 6 genes was identified that is predicted to be involved in fucose metabolism in this species. Deletion of the gene encoding the predicted fucose kinase within the operon in strain *L. rhamnosus* GG, resulted in a strain that scored negative for L-fucose fermentation in an API analysis. However, in a batch culture, the predicted fucosyl kinase deficient derivative only displayed a reduced growth rate on L-fucose containing media compared to the parental strain. This data confirms the involvement of the identified operon in fucose metabolism, although at least for the fucosyl kinase function there appears to be a certain degree of redundancy in this species. Mining of the *L. rhamnosus* genome did not identify the candidate gene encoding this redundant function among 80 genes associated with carbohydrate utilization, and therefore the gene(s) involved in fucose phosphorylation in the fucose kinase mutant strain remains to be identified. Nevertheless, it would be of interest to evaluate the intestinal persistence of the mutant strain that exhibits reduced fucose utilization capacity.

Introduction

Lactobacillus rhamnosus is one of the most frequently applied and widely investigated probiotic species (1–6). It is a phylogenetically versatile bacterium encompassing strains isolated from dairy products (7, 8), plants (9) and mucosal surfaces of mammalian hosts, where they are encountered in the oral (10), gastrointestinal (GI) (11) and vaginal (12, 13) cavities in various host organisms. The main economic importance of the species originates from their use in the manufacturing of fermented foods and as probiotic organisms that are provided to the consumer as an ingredient in dairy or other foods, or as a dietary supplement (14). The probiotic functionalities of *L. rhamnosus* have been extensively documented, and includes the modulation of immune cell functions (15, 16), and the alleviation of abdominal pain in gastrointestinal disorders, particularly among children with irritable bowel syndrome (IBS) (17). In addition, based on its long history of safe use as a probiotic organism (4, 18, 19) and its ability to survive transit through the GI tract, the probiotic strain *L. rhamnosus* GG has been considered an ideal vehicle for mucosal-targeted delivery of vaccine antigens (20, 21). Consequently, extensive studies are ongoing to elucidate *in vivo* mechanisms involved in the intestinal adaptation and host–microbe interactions of *L. rhamnosus*,

aiming to enhance its delivery and fitness in the gut environment. Although metabolic capacities have traditionally not been considered as an important factor to contribute to probiotic function, the ability to survive and persist in the intestinal environment is undoubtedly influenced by these characteristics, whereby the opportunity for the microbe to interact with the host's mucosal tissues can be significantly modulated (14, 22, 23). Consequently, metabolic profiling of candidate probiotic strains can contribute to the selection of strains with the intention to predict their persistence *in situ*, which can be assumed to amplify their capacity to interact with the host.

The availability of environmental carbohydrates can strongly influence the persistence and growth of *L. rhamnosus*. Intestinal representatives of this species can be encountered in close association with the mucosal layer of the intestine, where binding to intestinal mucins can modulate the production of surface exposed proteins and stimulates the formation of bacterial biofilms (24). L-fucose is a prominent component of eukaryotic cell surface decorations and mucin glycosylation, where it can comprise 4 to 14% of the total glycan conjugation (25). Importantly, utilization of L-fucose by *L. rhamnosus* strains as a carbon and energy substrate has been associated with an increased fitness in the intestinal tract (7). Only few Lactobacilli have been shown to be able to utilize L-fucose as a carbon and energy source, which suggests that this inherent capacity of some *L. rhamnosus* strains could be a discriminating trait that bears relevance for their intestinal persistence and may thereby contribute to their health promoting activities. In addition, the ability to utilize fucose has been identified as a metabolic capacity that can stimulate the colonization persistence of more benign intestinal bacteria like Bifidobacteria (26).

Many of the *Lactobacillus rhamnosus* intestinal isolates, including the probiotic strain GG, are able to ferment L-fucose. While L-fucose utilization genes *L. rhamnosus* strain GG were recently identified (27), information about the presence of this operon in other strains and the diversity of the L-fucose utilization genes is not yet available. In fact, it has been reported that this phenotypic trait is much less frequently encountered in dairy isolates (7). In the present study the diversity of the fucose utilization phenotype was further investigated in strains of the species *L. rhamnosus*, which when correlated with the (partial) genome information of 25 *L. rhamnosus* strains enabled the identification of the fucose utilization operon in this species.

Results

L-fucose supports growth of a distinct set of *L. rhamnosus* strains.

API50 CH tests were performed to identify carbon substrates metabolized by genetically diverse *L. rhamnosus* strains isolated from various environments (Supplementary Table 1 – panel A, B). Of the 49 carbon sources assayed (Table 1), all *L. rhamnosus* strains were able to ferment 15 carbon sources, which matched with the API insert predictions of the *L. rhamnosus* fermentation profiles (Table 1, column 2). In addition, the strains could use an additional group of 20 carbon sources to a variable extent, including L-fucose. Among the 25 strains tested, approximately a third of the strains tested were able to utilize L-fucose (Table 1).

Identification of candidate genes involved in growth of *L. rhamnosus* on L-Fucose by gene-trait matching.

The L-fucose fermentation capacity diversity was used to identify candidate genes by gene-trait matching (GTM), in which the phenotypic data were correlated to presence and absence of genes obtained by genome sequencing of the same strains (Supplementary Table 1).

The correlation applied was based on an equation aimed to identify genes of which the presence and absence in the different strains coincides significantly with the presence and absence of the carbohydrate utilization phenotype (i.e., the API50 L-fucose utilization data). Notably, the four candidate genes with the highest correlation (GTM calculation matrix available as Supplementary Table 4) form an operon that has already been annotated (28) as a putative fucose utilization operon, including the genes encoding a FGGY family carbohydrate kinase (*fucK*), a L-fucose isomerase (*fucI*), a transport associated protein (*fucU*) and a carbohydrate proton symporter of the major facilitator family (*mfs*-1) (Figure 2A).

Table 1. API 50 CH profiling of *L. rhamnosus* strains used in this study and for the mutant strain NutRes287. Fermentation ability is indicated in black for positive, grey for partially positive and white for negative.

Strains	Biomerieux database (% of strains)	Lrh42	Lrh10	Lrh46	Lrh48	Lrh40	Lrh44	Lrh45	Lrh39	Lrh47	Lrh36	Lrh34	Lrh38	Lrh37	Lrh49	Lrh50	Lrh33	Lrh24	Lrh29	Lrh15	Lrh4	Lrh35	Lrh6	Lrh41	Lrh3	LGG	NutRes287
0 CONTROL	0																										
1 GLYcerol	42																										
2 ERYthritol	0																										
3 D-ARAbinose	9																										
4 L ARAbinose	8																										
5 RIBose	100																										
6 D-XYLose	0																										
7 L-XYLose	0																										
8 ADOnitol	0																										
9 β-Methyl-D-Xyloside	0																										
10 GALactose	100																										
11 GLUcose	100																										
12 FRUctose	100																										
13 MaNnosE	100																										
14 SorBosE	92																										
15 RHAmnose	100																										
16 DULcitol	14																										
17 INOsitol	42																										
18 MANnitol	100																										
19 SORbitol	100																										
20 α-Methyl-D-Mannoside	7																										
21 α-Methyl-D-Glucoside	85																										
22 N-Acetyl-Glucosamine	100																										
23 AMYgdalin	99																										
24 ARButin	100																										
25 ESCulin	85																										
26 SALicin	100																										
27 CELlobiose	100																										
28 MALtose	99																										
29 LACtose	100																										
30 MELibiose	9																										
31 Sucrose	71																										
32 TREhalose	99																										
33 INUlin	0																										
34 MeLeZitose	99																										
35 RAFfinose	7																										
36 Starch	0																										
37 GLYcoGen	7																										
38 XyLiTol	0																										
39 GENtiobiose	85																										
40 D-TURanose	92																										
41 D-LYXose	42																										
42 D-TAGatose	99																										
43 D-FUCose	0																										
44 L-FUCose	7																										
45 D-ARabitoL	0																										
46 L-ARabitoL	7																										
47 GlucoNaTe	85																										
48 2-Keto-Gluconate	0																										
49 5-Keto-Gluconate	0																										

Two additional genes with a similar correlation significance with the fucose utilization phenotype were located upstream of this 4-gene operon, but were not located within the same operon. These genes were predicted to encode a DeoR family transcriptional regulator (*fcsR*), and a class II sugar aldolase (*aldF*).

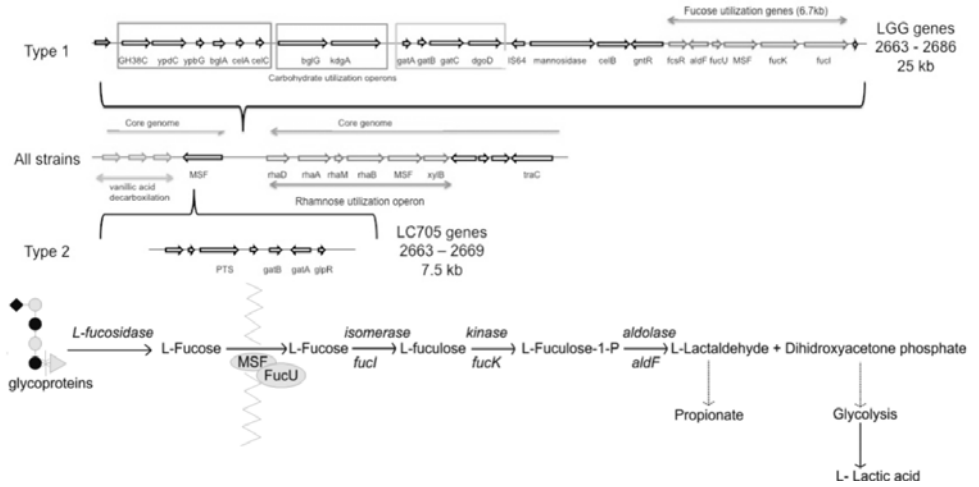


Figure 1. Schematic representations of L-fucose utilization operons and metabolic pathway in L. rhamnosus strains. Panel A: Two genomic conformations are encountered in the strains used: type 1: a large size (25 kb) insert with at its 3'-end the proposed fucose utilization operon; and type 2: smaller sizes (7.5 kb) insert encoding for carbohydrate utilization genes: Glutamyl-tRNA(Gln) amidotransferase (gatA and gatB) and the repressor for the glycerol-3-phosphate regulon (glpR). These inserts are present upstream of the core-rhamnose utilization operon and downstream of the vanillic acid decarboxylation operon of the core genome of L. rhamnosus. B. the proposed pathway of L-fucose utilization by L. rhamnosus encoded in the type 1 locus in panel A.

The genes downstream of the 4-gene operon had lower correlation significance with the fucose utilization phenotype, and may encode auxiliary functions, including a predicted glycosyl hydrolase, carbohydrate transport (PTS), and transcriptional regulation function. Of note, L-fucose is predicted to be imported by the cell through the joint activity of Msf-1 and FucU, and subsequently converted to L-fuculose by the FucI isomerase, followed by its phosphorylation to L-fuculose-1-P by FucK, and the cleavage of L-fuculose-1-P into lactaldehyde and dihydroxyacetone phosphate by the aldolase aldF. The latter metabolic product can be incorporated into the central energy metabolism pathway, enabling the bacterium to utilize fucose as an energy source (Figure 1 Panel C). The *L. rhamnosus* strains, however, do not appear to encode genes that are predicted to function in the utilization of L-lactaldehyde or hydroxypropionaldehyde in central energy metabolism, suggesting that these metabolic intermediates may be excreted into the environment by fucose fermenting strains.

Indeed, the genetic locus identified (Figure 1A; exemplary genetic organization derived from the *L. rhamnosus* GG genome) appears to encode all functions required for

a complete fucose utilization pathway (Figure 2C). Our analysis indicates that these genes localize at the 3'-end of a larger genetic insert that appears to be present in all L-fucose utilizing strains (~ 25 kb, figure 2A, type 1), and which happens to be consistently inserted upstream of the rhamnose utilization operon that is part of the *L. rhamnosus* core genome (7). The genes encoded within this insert appear to be consistently absent in the genomes of the strains that cannot utilize L-fucose, where the insert appears to be replaced by an alternative, smaller (7.5 kb) insert (Figure 1B) that encodes Glutamyl-tRNA(Gln) amidotransferase and the predicted repressor of the glycerol-3-phosphate regulon. In addition, the *L. rhamnosus* genomes were mined for genes that could encode carbohydrate hydrolase functions involved in the liberation of L-fucose from oligosaccharides. This analysis identified three alpha-L-fucosidases in the GG genome (LGG_01874, LGG_02652 and LGG_02741) that appear to be part of the core genome of the *L. rhamnosus* species and therefore present in all strains (7), while additional predicted L-fucosidase encoding genes were identified in some of the other strains, suggesting that this function displays a considerable degree of redundancy in this species.

To evaluate whether similar fucose utilization loci are encoded by other *Lactobacillus* species, the genes encompassed within the proposed fucose utilization operon were used to search for homologues in *Lactobacillus* and other lactic acid bacteria genomes available in the NCBI genome sequence database. Remarkably, no species of the *Lactobacillus* genus harbored the same set of genes, and the *fucK*, *fcsR*, *aldF* and *fucU* genes appeared to be unique for *L. rhamnosus*. However, homologues for the predicted L-Fucose isomerase (*fucI*) were identified in the genomes of several lactobacilli (*L. casei*, *L. zeae*, *L. acidophilus*, *L. buchneri*, *L. plantarum*, *L. pentosus*, *L. crispatus* and *L. shenzhenensis*), streptococci and pediococci, and a homologue of the major facilitator transporter encoding gene (*mfs*-1) was also present in the genomes of *L. casei* and *L. zeae*.

Phenotypic analyses of a fucosyl kinase deficient mutant

To evaluate the hypothesized role of the operon in L- Fucose utilization by *L. rhamnosus*, a gene-specific deletion of the fucosyl kinase was constructed in strain *L. rhamnosus* GG (LGG_02684), yielding a derivative of this strain (designated NutRes287) in which the *fucK* gene is replaced by a chloramphenicol resistance gene-cassette. Unlike transporters and isomerases that are sometimes quite promiscuous in terms of their substrate specificity (29), kinases often display specificity for a certain carbohydrate (30, 31). In addition, no homologs of the fucosyl kinase were found in the *L. rhamnosus* GG genome.

Among the more closely related proteins encoded by *L. rhamnosus* GG is the rhamnokinase that is encoded within the genetically linked rhamnose operon and displays 33% aminoacid identity compared to the FucK protein, which suggests that the predicted rhamnose and fucose kinases are not functionally overlapping. Moreover, previous studies concluded that the rhamnose operon is non-functional in *L. rhamnosus* GG (7). These findings suggest that it is unlikely that the proposed function of the fucosyl kinase gene is redundant in *L. rhamnosus* GG.

The gene replacement of the *fucK* gene by a chloramphenicol resistance cassette was confirmed by PCR (Supplementary Figure 1). The growth of the wild type and mutant on standard MRS agar plates appeared identical, and microscopic examination of the cells did not reveal any apparent differences (data now shown).

Table 2. Growth characteristics of *L. rhamnosus* strains GG and NumRes278 grown in carbohydrate free MRS, carbohydrate free MRS supplemented with 1% glucose (gMRS) or L-fucose (fMRS) at 37^0C after 18 hours of growth.

Linear growth	MRS no carbohydrates		fMRS		gMRS	
	μ (h^{-1})	Max OD600	μ (h^{-1})	Max OD600	μ (h^{-1})	Max OD600
NutRes287	0.011±0.0002	0.28	0.02±0.0016	0.45	0.22±0.0025	1.78
LGG (wt)	0.012±0.0008	0.31	0.06±0.007	0.49	0.33±0.0041	1.79

Importantly, evaluation of the mutant using API50 CH profiling established the inability of the mutant to ferment L-Fucose (Figure 1), confirming the proposed role of the *fucK* gene (and the genetically linked operon) in L-Fucose metabolism.

To further examine the growth of the wild-type strain and its *fucK* derivative on L-fucose, a carbohydrate-free MRS-based medium was prepared that allowed the supplementation with a chosen carbohydrate (1 % W/V) (Table 2). The results of these growth experiments clearly established that for fast growth both strains require the addition of a carbon source in the MRS-based media. The supplementation of glucose to the medium allowed rapid growth of both the wild-type and mutant strain (Table 2). However, the supplementation of the medium with fucose resulted in distinct growth rates of the wild-type and mutant strain, where the wild-type strain displayed a slow but constant growth rate on this carbon source, whereas the *fucK* deletion derivative grew at a rate that was very similar to the rate observed on carbon source free medium (Table 2). Notably, the wild-type strain grows substantially faster on glucose than fucose, which may be due to the difference in energy gain predicted for these two carbon sources. Glucose metabolism generates two molecules of glyceraldehyde-3-phosphate (G3P) available for further glycolytic energy generation, whereas only a single G3P

(dihydroxyacetone phosphate) will be available from L-fucose metabolization (see above). Additionally, the postulated export of the non-utilized L-lactaldehyde moiety derived from L-fucose might require additional energy (32). The lower final optical density reached by the fucose grown cells as compared to those grown on glucose reflects a lower energy generation during L-fucose metabolism (Table 2). Nevertheless, despite its reduced growth rate, the mutant reached the same final optical density as the wild type strain grown of L-fucose after 24 hours of growth.

Discussion

Fucose utilization by bacteria can be a biologically significant ability / activity in the intestine, since it was shown to play a role in microbial persistence in this competitive niche. Moreover, this carbohydrate may also be an important component in the molecular communication between the host and its microbiota.

Fucose is not only highly abundant in the intestine, it has also been shown to play a modulatory role in several mammalian physiological processes, including intestinal system development, immune recognition and intestinal colonization (33). During weaning in rats, the intestine undergoes radical changes that include increased fucose biosynthetic capacity (34), which leads to increased fucosylated glycan expression through the up-regulated fucosyl-transferase activity (35), as a response to the progressive intestinal colonization by the microbiota (36, 37). Analogously, in humans, the faecal oligosaccharide-profiles in the first six months of life display increasing levels of fucosylation (38). Initially these oligosaccharides derive from human milk oligosaccharides (HMO) degradation, while with time they increasingly derive from the degradation of the infant's mucus layer (38). Physiological processes in adulthood can also be modulated by fucose. In adult mice, fucosylated, but not sialylated, human milk oligosaccharides were shown to modulate gut motility in a mouse *ex vivo* colon model (39). In addition, *in vitro* cultured human epithelial cells (Caco2) were shown to launch innate immune responses in response to free fucose, suggesting that the presence of this carbohydrate acts as a danger signal (40), possibly as a reflection of invasion by mucin-hydrolysing bacteria and damage of the mucus barrier.

Bacterial utilization of fucose has also been reported to stimulate the production of fucosylated glycans in the small-intestinal epithelium of mice (41), although only few members of the intestinal microbiota are able to induce this effect (42). Monoassociation of germ-free mice with *Bacteroides thetaiotaomicron*, a prominent member of the intestinal microbiota, restored the epithelial fucosylation activity observed in conventional animals. Importantly, an isogenic strain of *Bacteroides thetaiotaomicron* that lacked the ability to utilize L-fucose as a carbon source did not elicit this response

(42), supporting the role of bacterial colonization and fucose utilization in the modulation of the host's fucosylation activity. *Bacteroides thetaiotaomicron* can produce multiple fucosidases that release fucose from host glycans, which can provide a higher fucose availability in the gut lumen (25), which not only affects epithelial cell functions (see above), but also modulates the resident intestinal microbiota that in its turn can impact host health (43). For example, fucose metabolism in *Escherichia coli* plays a central role in a complex regulatory network that orchestrates major cellular processes, including energy production, carbohydrate and amino acid metabolism, stress response and cell–cell communication (44). Moreover, L-fucose represses the expression of virulence genes in entero-haemorrhagic *Escherichia coli* (EHEC) by its role as a ligand involved in the activation of the FusKR signalling cascade (44, 45).

The constant presence of this carbohydrate in the gut as a component derived from mucus provides a clear competitive advantage for intestinal microbes that can release and use this glycan moiety for growth. The competitive advantage of fucose utilization was evidenced in co-colonization experiments in chicken, where a *fucP* (permease) mutant of *Campilobacter jejuni* was significantly outcompeted by the wild-type strain (49). Analogously, genes involved in L-fucose utilization can be found in representatives of many bacterial phyla that are abundant in the intestine such as *Proteobacteria*, *Firmicutes*, *Actinobacteria*, and *Verucomicrobia*. Notably, similar to *L. rhamnosus*, many of these organisms (e.g., *Roseburia inulinivorans* (50)) encode several hydrolases to liberate L-fucose from their environment, and some are predicted to encode redundant L-fucose utilization operons, supporting the significance of this trait in some environments.

Fucose utilization in our group of *L. rhamnosus* strains appears to be higher as compared to the value provided by the identification database from Biomerieux Apiweb (7%), but lower than the approximately 50 % that was previously found by profiling 100 *L. rhamnosus* strains (7). Although the latter study concluded that dairy strains are unable to utilize L-fucose (7), our profiling experiment identified L-fucose-utilizing strains both among intestinal and dairy isolates.

Gene-trait matching approaches have enabled the identification of genes that are predicted to be involved in the utilization of various carbohydrates in *L. casei* (54), *L. plantarum* (55), *L. lactis* (56). The genes identified in this way were predicted to encode sugar metabolism functions related to growth on lactose, saccharose, galactitol, mannose, mannitol, cellobiose, ribose, sorbitol and sorbose. However, these predicted functions were not verified by experimental (e.g., mutagenesis) approaches.

Nevertheless, these studies support the use of comparative genomics and the relationship with strain-specific phenotypes to deliver gene-function assignments.

The gene-trait matching strategy is bound to be more successful with more 'simple' and stably detected phenotypes that are linked to a limited number of genes. This is apparent from the successful gene-trait matching studies that were published and led to the identification of *L. plantarum* mannose specific adhesin encoding gene (*msa*), that is of potential interest for its role in binding to intestinal epithelial tissues (55, 57). More complex, potentially multi-gene phenotypes provide more challenging gene-trait possibilities, and may fail to pinpoint a single gene or genetic locus that is responsible for the phenotype. Nevertheless, immune modulatory profiles of *L. plantarum* strains enabled the identification of several genes in this organism that contribute to its capacity to stimulate immune cells in a specific manner (58, 59). Notably, some of the more complex phenotypes may be more reliably be assigned to differences in levels of specific transcripts. This recently described transcriptome trait matching approach illustrated how the expression of several conserved genes in *L. plantarum* was associated to survival in an *in vitro* GI-tract assay, and subsequent mutagenesis of these genes confirmed their proposed role in this complex phenotype (60, 61). In our study, differential use of L-Fucose by *L. rhamnosus* strains was readily observed in the API tests as a color change. Genomic correlation analysis was successful in identifying genes of which the presence co-occurred with this phenotype, and which were annotated as a putative L-fucose utilization operon. The correlation of these genes with the phenotype was substantiated by creating a deletion mutant of the fucosyl kinase gene in *L. rhamnosus* GG, leading to a negative API score and almost negligible growth rates on media with L-fucose as the sole supplemented carbon source. These findings support the strength of gene-trait matching approaches that harness genomic diversity for the identification and verification of gene-function relationships.

This study identified the fucose utilization operon of *L. rhamnosus* by a gene-trait matching method. The operon appears to be localized in a genomic region displaying two predominant conformations among the strains of this species, where one of these conformations (type 1; Figure 1) encompasses the fucose operon. The putative L-fucose metabolic pathway in *L. rhamnosus* is encoded by a 6.7 kb gene cluster consisting of six genes of which four appear to be species-specific. There is some indication of horizontal transfer of the L-fucose operon, since the *L. rhamnosus* fucose isomerase and the MFS-1 transporter genes are also present in other *Lactobacillus* genomes. In contrast, none of the closely related *L. casei* strains appeared to encode a genetic locus similar to the *L. rhamnosus fuc* locus. Besides a strongly reduced, if not negligible, growth rate, the *fucK*

mutant of *L. rhamnosus* GG displayed no clear phenotype in L-fucose supplemented rich laboratory medium, illustrated by the similar overnight culture density reached by the mutant and its parental strain. This observation illustrates that growth experiments in a rich laboratory medium are less discriminative as compared to the simple acidification assay that underlies the API test, and that an alternative enzyme may in part take over the role of fucose kinase in a low efficiency L-fucose utilization pathway (Supplementary Figure 2). Moreover, extended incubation of the *fucK* mutant in L-fucose supplemented media, led to an acceleration of the growth rate of the mutant strain in this medium, which may imply that the strain can adapt further to the use of this carbohydrate source. Functional redundancy is common in carbohydrate utilization pathways and can ensure the utilization of the most common energy sources even in case of disruption of the dedicated pathway (7, 62).

While the present study investigated the genetic basis for utilization of L-fucose, this molecule generally needs to be released from niche-specific polysaccharide moieties like mucins. The release is dependent on excreted or cell-surface associated fucosidases, which have been studied in various bacteria, including *L. casei* (63). *L. rhamnosus* encodes fucosidases that share a high level of identity with those identified in *L. casei*. Notably, the expression of these fucosidases in *L. casei* strain BL23 was shown to be subject to repression, since a mutant strain that lacks the fucosidase regulator gene displayed a higher growth rate as compared to its parent strain on fucose-containing polysaccharide substrates (63). However, growth on L-fucose was not shown yet for strains of *L. casei* (64), and analogously, it is also reported negative for this species in the APIweb database. This may suggest that *L. casei* utilizes other carbon sources of which the release from complex oligo- or polysaccharides may depend on fucosidase activity. Intriguingly, in *Roseburia intestinalis*, the fucose operon regulator encoded by *fucR* functions as a sensor that controls gene expression of functions associated with the production of fucosylated glycans that interact with the host's immune system (65), as well as those associated with L-fucose catabolism (50).

L-Fucose utilization among lactobacilli and other LAB appears to be functionally important and relatively rare, thereby increasing the potential interest in further unraveling the potential role of fucose utilization in intestinal persistence and survival. The use of fucose could provide a competitive advantage to strains that are able to utilize this mucus component. Except for its metabolic utilization, controlling the quantity of free fucose could also provide a mechanism involved in host modulation through the physiological role of this carbohydrate in regulation of the bowel movement and innate immunity (39), which in turn would impact on the time of residence and

physicochemical challenge encountered by the same bacteria in the intestine. Therefore, it would be of interest to examine the transcriptional control of the L-fucose utilization genes in *L. rhamnosus*, including the fucosidases, under laboratory conditions, but especially *in situ* in the intestinal tract. Such studies could provide support for the role of this phenotype in intestinal persistence and growth both *in vitro* and *in situ* in the intestine, which is a subject of further studies.

Materials and Methods

Bacterial strains and growth conditions

The bacterial strains and plasmids used in this study are listed in Supplementary Table 1. *L. rhamnosus* strains were routinely propagated in MRS broth (Difco Laboratories, Detroit, MI, USA) or on MRS agar (aMRS, 1.5% (w/v)) at 37°C. Recombinant strains were selected in the presence of 5 μg ml−1 of erythromycin (Sigma-Aldrich, St. Louis, MO, USA) and/or 10 μg ml−1of chloramphenicol (Sigma-Aldrich, St. Louis, MO, USA) when appropriate. For fucose growth studies, cells were grown in carbon source-free MRS supplemented with 1% L-fucose (fucMRS) (66). As controls, carbon source-free MRS or supplemented with 1% glucose (gluMRS) were prepared in the laboratory. Overnight cultures were inoculated at 1% (v/v) into 96-well microplate wells (Corning Costar, Corning, NY) in triplicate, each containing 250 μl of MRS supplemented with 1% of the carbohydrate. Microplates were sealed with clear film (BioRad, Hercules, CA, USA), incubated at 37°C in a ATI Unicam UV/Vis spectrophotometer (ATI Unicam Ltd., Cambridge, UK) microplate reader and optical density of cells was monitored at 600 nm (OD600) for 24 h. The average specific growth rate (μ) from the growth experiments was calculated for the linear interval of logarithmic growth for each strain.

Escherichia coli was used as an intermediate cloning host, and was routinely grown in tryptic soy broth (TSB) (Oxoid) at 37°C under aerobic conditions (250 rpm shaking). One Shot TOP 10 competent *E. coli* were transformed as recommended by the manufacturer (Invitrogen). When appropriate erythromycin or chloramphenicol were added to the media at a final concentration of 150 and 10 μg*ml−1, respectively.

Genome sequencing

Based on AFLP clustering and origin information, we selected 25 strains from the Nutricia and Palaiseau Danone Research Collections (see detailed sequencing information in **Chapter 3**). A complete list of the selected *L. rhamnosus* strains and their origin can be found in Supplementary Table 1. For DNA preparation, 2 ml of overnight

culture was pelleted, washed and resuspended in 2 mL Tris-Cl (10 mM, pH 7.6), EDTA (1 mM), sodium dodecyl sulfate (SDS) (0.5%, w/v) (TES) buffer containing 100 ug/ml RNAse (Sigma). Cell lysis was performed with lysozyme (360 mg/ml) and mutanolysin (140 U /ml) during 2 h at 37°C, then 300 ul water was added followed by addition of 80 ul of 20% SDS solution. Afterwards RNAse treatment was performed for 1 hour at 37°C. The DNA extraction was performed using phenol/chloroform (3x), and precipitation from the watery phase with isopropanol. DNA pellets obtained were washed with 70% ethanol and dried.

Carbohydrate profiling by API50

The API 50CH fermentation system enables strain characterization in terms of its metabolic capacity to ferment a wide range of individual substrates, and can be employed in the taxonomic identification of lactobacilli (67). Carbohydrate acidification for 25 *L. rhamnosus* strains was scored after 24 h and 48 h incubation at 37^{o}C and compared with reference API 50CH profiles for *L. rhamnosus* (BIOMÉRIEUX SA, La Balme les Grottes, 38390 Montalieu Vercieu, France).

Gene trait matching

Candidate genes potentially involved in carbon source utilization were identified by *in silico* gene trait matching using genomic information from 25 *L. rhamnosus* strains and the API50 L-fucose growth. The API50 raw data were transformed into a binary matrix. To assess the significant co-occurrence of *L. rhamnosus* OGs with the growth dataset, we employed a previously established correlation analysis strategy (Chapter 3) that identifies genes of which the presence absence pattern among strains most consistently coincides with the absence / presence of the phenotype of interest, including carbohydrate utilization.

A list of the highest scoring candidate genes and their RAST-based annotation is provided in the supplementary information (Supplementary Table 4). Genes predicted to be involved in the utilization of carbohydrate sources, were further analyzed manually, giving preference to genes organized in operons that have the same phenotype correlation score.

DNA manipulations and transformation

DNA isolations from *E. coli* were performed using Jetstar columns as recommended by the manufacturer (Genomed GmbH, Bad Oberhausen, Germany). For DNA manipulations in *E. coli*, standard procedures were used (68). PCR primers (Table S1) were synthesized by Biolegio, Nijmegen. For cloning and DNA sequencing, PCR

amplicons were generated using KOD DNA polymerase (Toyobo, Japan) using the manufacturer's instructions. Colony PCR was performed using 2x PCR mix polymerase (Promega, Madison, USA) using primer pairs as listed in Supplementary Table 3. Both vector-specific primers annealing to the *cat* region (reverse primer pNZ5319_RC128 and forward primer pNZ5319_RC128) as well as insert-specific primers (Genomic_up with pNZ5319_RC128 and Genomic_dn with pNZ5319_RC129) were used for each colony.

For transformation of *L. rhamnosus* strains, a preculture in MRS broth was cultured in MRS broth containing 2% glycine, growing cells to an OD600 of 0.3. Cell suspensions were cooled on ice for 10 minutes and pelleted by centrifugation for 10 minutes at 3500 rpm. Cell pellets were resuspended in ice cold 30% PEG-1450 and kept on ice for 10 minutes, and again pelleted by centrifugation for 10 minutes at 3500 rpm and resuspended at a 100-fold concentration in ice cold 30% PEG-1450. Subsequently, 40 µl of cell suspension and up to 5 µl of plasmid DNA solution were mixed and used in electroporation, using a GenePulser Xcell electroporator (Biorad, Veenendaal, The Netherlands) in cuvettes with a 2 mm electroporation gap at 1,5 kV, 25 µF capacitance and 400 Ω parallel resistance.

Construction of deletion mutant of the fucosyl kinase encoding gene

To obtain a mutant in *L. rhamnosus* LGG in which the *fucK* gene had been replaced by the chloramphenicol resistance gene a pNZ5319-based (69) integration plasmid has been constructed. To this end, the upstream flanking region of *fucK* was amplified with primers FK_up_F and FK_up_R (see table 3) using chromosomal DNA of *L. rhamnosus* LGG as a template and was cloned in SwaI digested pNZ5319, resulting in the intermediate construct, in which the downstream flanking region of *fucK* amplified with primers FK_do_F and FK_do_R was cloned into the Ecl136II site. The resulting plasmid contains the *fucK* flanking regions separated by the *cat* resistance cassette of pNZ5319.

Ligation mixtures were transformed to *E. coli* TOP10 cells and from chloramphenicol resistant colonies the orientation of the blunt-cloned inserts was checked by PCR using primers 66upFCS120 with Fk_up_R, and FK_up_F with 66doRCS85 for the upstream insert, while primers 71upFCS87 with FK_do_R, and FK_do_F with eryfor20 were used for the downstream insert (data not shown). A single colony isolate harbouring a vector in which both fragments were inserted in the required orientation that mimics their positioning relative to the *fucK* gene in the LGG genome was selected and the plasmid was designated pNR001. The DNA sequence of the inserted regions of pNR001 was confirmed by sequence analysis (data not shown).

Plasmid pNR001 was transformed to *L. rhamnosus* GG. Integrants were selected on MRS plates supplemented with 10 ug/ml chloramphenicol and subsequently replica-plated to MRS containing 30 ug/ml erythromycin. Colonies displaying a chloramphenicol-resistant and erythromycin-sensitive phenotype were selected as double cross-over gene-replacement (Δ*fucK*::*cat*) candidates. Homologous recombination in both the upstream and downstream flanking regions was confirmed by PCR using primers FK_up_genomic with pNZ5319_RC128 (upstream) and primers pNZ5319_RC128 with FK_do_genomic (downstream), and absence of the *ery* cassette was confirmed by PCR using primers CC1 and CC4 ((Supplementary Table 3, Supplementary Figure 1). A single colony in which the Δ*fucK*::*cat* genotype was confirmed was designated NutRes287.

References:

1. Salminen MK. et al. 2002. *Lactobacillus bacteremia during a rapid increase in probiotic use of Lactobacillus rhamnosus GG in Finland.* Clin. Infect. Dis. 35:1155–1160.

2. Alander M. et al. 2001. *Persistence of colonization of human colonic mucosa by a probiotic strain, Lactobacillus rhamnosus GG, after oral consumption.* VTT Publ. 65:79–82.

3. Tannock GW. et al. 2000. *Analysis of the fecal microflora of human subjects consuming a probiotic product containing Lactobacillus rhamnosus DR20.* Appl. Environ. Microbiol. 66:2578–2588.

4. Vancanneyt M. et al. 2006. *Intraspecific genotypic characterization of Lactobacillus rhamnosus strains intended for probiotic use and isolates of human origin.* Appl. Environ. Microbiol. 72:5376–5383.

5. Kankainen M. et al. 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.

6. Lahtinen SJ. et al. 2012. *Probiotic cheese containing Lactobacillus rhamnosus HN001 and Lactobacillus acidophilus NCFM® modifies subpopulations of fecal lactobacilli and Clostridium difficile in the elderly.* Age (Omaha).

7. Douillard FP. et al. 2013. *Comparative genomic and functional analysis of 100 Lactobacillus rhamnosus strains and their comparison with strain GG.* PLoS Genet. 9:e1003683.

8. Salva S, Villena J, Alvarez S. 2010. *Immunomodulatory activity of Lactobacillus rhamnosus strains isolated from goat milk: Impact on intestinal and respiratory infections.* Int. J. Food Microbiol. 141:82–89.

9. Buryako IA. et al. 1999. *Nitrate reduction ability of bacteria of the genus Lactobacillus isolated from plant sources.* Prikl. Biokhimiya i Mikrobiol. 35:401.

10. Yli-Knuuttila H. et al. 2006. *Colonization of Lactobacillus rhamnosus GG in the oral cavity*. Oral Microbiol. Immunol. 21:129–131.

11. De Champs C. et al. 2003. *Persistence of colonization of intestinal mucosa by a probiotic strain, Lactobacillus casei subsp. rhamnosus Lcr35, after oral consumption*. J. Clin. Microbiol. 41:1270–1273.

12. Gardiner GE. et al. 2002. *Persistence of Lactobacillus fermentum RC-14 and Lactobacillus rhamnosus GR-1 but not L. rhamnosus GG in the human vagina as demonstrated by randomly amplified polymorphic DNA.* Clin. Diagn. Lab. Immunol. 9:92–96.

13. Pascual LM. et al. 2008. *Lactobacillus rhamnosus L60, a potential probiotic isolated from the human vagina.* J. Gen. Appl. Microbiol. 54:141–148.

14. Ceapa C. et al. 2013. *Influence of fermented milk products, prebiotics and probiotics on microbiota composition and health*. Best Pract. Res. Clin. Gastroenterol. 27:139–55.

15. Lehtoranta L, Pitkäranta A, Korpela R. 2012. *The effects of probiotic combination Lactobacillus rhamnosus GG and Lactobacillus rhamnosus Lc705 in cytokine and chemokine response in human macrophages.* Int. J. Probiotics Prebiotics 7:17–22.

16. Cáceres P. et al. 2010. Effects of *Lactobacillus rhamnosus* HN001 on acute respiratory infections and intestinal secretory IgA in children. J. Pediatr. Infect. Dis. 5:353–362.

17. Horvath a., Dziechciarz P, Szajewska H. 2011. *Meta-analysis: Lactobacillus rhamnosus GG for abdominal pain-related functional gastrointestinal disorders in childhood.* Aliment. Pharmacol. Ther. 33:1302–1310.

18. Saxelin M. 1997. *GG—a human probiotic strain with thorough clinical documentation.* Food Rev. Int.

19. Alander M. et al. 2001. *Persistence of colonization of human colonic mucosa by a probiotic strain, Lactobacillus rhamnosus GG, after oral consumption.* VTT Publ. 79–82.

20. Marcotte H, Pant N, Hammarström L. 2008. *Engineered lactobody-producing lactobacilli: A novel form of therapy against rotavirus infection.* Future Virol. 3:327–341.

21. Pant N. et al. 2007. *Effective prophylaxis against rotavirus diarrhea using a combination of Lactobacillus rhamnosus GG and antibodies*. BMC Microbiol. 7.

22. Marco ML. et al. 2010. *Convergence in probiotic Lactobacillus gut-adaptive responses in humans and mice*. ISME J. 4:1481–4.

23. De Vos MGJ, Poelwijk FJ, Battich N, Ndika JDT, Tans SJ. 2013. *Environmental dependence of genetic constraint*. PLoS Genet. 9.

24. Sanchez B. et al. 2010. *Adhesive properties, extracellular protein production, and metabolism in the Lactobacillus rhamnosus GG strain when grown in the presence of mucin.* J. Microbiol. Biotechnol. 20:978–984.

25. Pacheco AR. et al. 2012. *Fucose sensing regulates bacterial intestinal colonization*. Nature 491:113–117.

26. Yu Z-T. et al. 2013. *The principal fucosylated oligosaccharides of human milk exhibit prebiotic properties on cultured infant microbiota.* Glycobiology 23:169–177.

27. Becerra JE, Yebra MJ, Monedero V. 2015. *An l-Fucose operon in the probiotic Lactobacillus rhamnosus GG is involved in adaptation to gastrointestinal conditions*. Appl. Environ. Microbiol. 81:3880–8.

28. Morita H. et al. 2009. *Complete genome sequence of the probiotic Lactobacillus rhamnosus ATCC 53103*. J. Bacteriol. 191:7630–7631.

29. Usvalampi A. et al. 2012. *Production of l-xylose from l-xylulose using Escherichia coli l-fucose isomerase.* Enzyme Microb. Technol. 50:71–76.

30. Lee T-W. et al. 2013. *Structural-functional studies of Burkholderia cenocepacia D-glycero-β-D-manno-heptose 7-phosphate kinase (HldA) and characterization of inhibitors with antibiotic adjuvant and antivirulence properties.* J. Med. Chem. 56:1405–17.

31. Nocek B.et al. 2011. *Structural studies of ROK fructokinase YdhR from Bacillus subtilis: insights into substrate binding and fructose specificity*. J. Mol. Biol. 406:325–42.

32. Scartovi V. 1960. *Rhamnose metabolism by L. casei var. rhamnosus and Pediococcus spp.* Ann. Microbiol. 10.

33. Becker DJ, Lowe JB. 2003. *Fucose: biosynthesis and biological function in mammals.* Glycobiology 13:41R–53R.

34. Ruggiero-Lopez D, Biol MC, Louisot P, Martin A. 1991. *Participation of an endogenous inhibitor of fucosyltransferase activities in the developmental regulation of intestinal fucosylation processes.* Biochem. J. 279 (Pt 3:801–6.

35. Biol MC, Martin A, Richard M, Louisot P. 1987. *Developmental changes in intestinal glycosyl-transferase activities.* Pediatr. Res. 22:250–6.

36. Torres-Pinedo R, Mahmood A. 1984. *Postnatal changes in biosynthesis of microvillus membrane glycans of rat small intestine: I. Evidence of a developmental shift from terminal sialylation to fucosylation.* Biochem. Biophys. Res. Commun. 125:546–53.

37. Biol-N'garagba MC, Louisot P. 2003. *Regulation of the intestinal glycoprotein glycosylation during postnatal development: Role of hormonal and nutritional factors.* Biochimie.

38. Albrecht S. et al. 2011. *Occurrence of oligosaccharides in feces of breast-fed babies in their first six months of life and the corresponding breast milk.* Carbohydr. Res. 346:2540–2550.

39. Bienenstock J. et al. 2013. *Fucosylated but not sialylated milk oligosaccharides diminish colon motor contractions.* PLoS One 8:e76236.

40. Chow WL, Lee YK. 2008. *Free fucose is a danger signal to human intestinal epithelial cells.* Br. J. Nutr. 99:449–54.

41. Hooper L V. et al. 1999. *A molecular sensor that allows a gut commensal to control its nutrient foundation in a competitive ecosystem.* Proc. Natl. Acad. Sci. U. S. A. 96:9833–9838.

42. Bry L, Falk PG, Midtvedt T, Gordon JI. 1996. *A model of host-microbial interactions in an open mammalian ecosystem.* Science 273:1380–1383.

43. Hooper L V, Bry L, Falk PG, Gordon JI. 1998. *Host-microbial symbiosis in the mammalian intestine: exploring an internal ecosystem*. Bioessays 20:336–343.

44. Morán-Zorzano MT. et al. 2007. *Occurrence of more than one important source of ADP-glucose linked to glycogen biosynthesis in Escherichia coli and Salmonella*. FEBS Lett. 581:4423–4429.

45. Eydallin G. et al. 2010. *Genome-wide screening of genes whose enhanced expression affects glycogen accumulation in Escherichia coli.* DNA Res. 17:61–71.

46. Thomas M. et al. 2013. *Bacteroides thetaiotaomicron and Faecalibacterium prausnitzii shape the mucus mroduction and mucin O-Glycosylation in colon epithelium.* Gastroenterology 144:S–59 –.

47. Varyukhina S. et al. 2012. *Glycan-modifying bacteria-derived soluble factors from Bacteroides thetaiotaomicron and Lactobacillus casei inhibit rotavirus infection in human intestinal cells*. Microbes Infect. 14:273–278.

48. Keeney KM, Finlay BB. 2013. *Microbiology: EHEC downregulates virulence in response to intestinal fucose.* Curr. Biol. 23:R108–R110.

49. Muraoka WT, Zhang Q. 2011. *Phenotypic and genotypic evidence for L-fucose utilization by Campylobacter jejuni.* J. Bacteriol. 193:1065–1075.

50. Scott KP. et al. 2006. *Whole-genome transcription profiling reveals genes up-regulated by growth on fucose in the human gut bacterium "Roseburia inulinivorans"*. J. Bacteriol. 188:4340–9.

51. Niñonuevo MR. et al. 2008. *Daily variations in oligosaccharides of human milk determined by microfluidic chips and mass spectrometry*. J. Agric. Food Chem. 56:618–26.

52. Schwab C, Gänzle M. 2011. *Lactic acid bacteria fermentation of human milk oligosaccharide components, human milk oligosaccharides and galactooligosaccharides.* FEMS Microbiol. Lett. 315:141–8.

53. Aldredge DL. et al. 2013. *Annotation and structural elucidation of bovine milk oligosaccharides and determination of novel fucosylated structures.* Glycobiology 23:664–676.

54. Smokvina T. et al. 2013. Lactobacillus paracasei comparative genomics: towards species pan-genome definition and exploitation of diversity. PLoS One 8:e68731.

55. Pretzer G. et al. 2005. *Biodiversity-based identification and functional characterization of the mannose-specific adhesin of Lactobacillus plantarum.* J. Bacteriol. 187:6128–6136.

56. Bayjanov JR. et al. 2013. *Genotype-phenotype matching analysis of 38 Lactococcus lactis strains using random forest methods.* BMC Microbiol. 13:68.

57. Gross G. et al. 2008. *Mannose-specific interaction of Lactobacillus plantarum with porcine jejunal epithelium.* FEMS Immunol. Med. Microbiol. 54:215–223.

58. Meijerink M, Hemert S Van, Taverne N. 2010. *Identification of genetic loci in Lactobacillus plantarum that modulate the immune response of dendritic cells using comparative genome hybridization.* PLoS One 5.

59. Van Hemert S. et al. 2010. *Identification of Lactobacillus plantarum genes modulating the cytokine response of human peripheral blood mononuclear cells*. BMC Microbiol. 10:293.

60. Van Bokhorst-van de Veen et al. 2012. *Modulation of Lactobacillus plantarum gastrointestinal robustness by fermentation conditions enables identification of bacterial robustness markers.* PLoS One 7:e39053.

61. Bron PA. et al. 2012. *Transcriptomes reveal genetic signatures underlying physiological variations imposed by different fermentation conditions in Lactobacillus plantarum*. PLoS One 7:e38720.

62. Bjursell MK, Martens EC, Gordon JI. 2006. *Functional genomic and metabolic studies of the adaptations of a prominent adult human gut symbiont, Bacteroides thetaiotaomicron, to the suckling period.* J. Biol. Chem. 281:36269–36279.

63. Rodríguez-Díaz J, Rubio-del-Campo A, Yebra MJ. 2012. Lactobacillus casei ferments the N-Acetylglucosamine moiety of fucosyl-α-1,3-N-acetylglucosamine and excretes L-fucose. Appl. Environ. Microbiol. 78:4613–9.

64. Nigatu A. 2000. *Evaluation of numerical analyses of RAPD and API 50 CH patterns to differentiate Lactobacillus plantarum, Lact. fermentum, Lact. rhamnosus, Lact. sake, Lact. parabuchneri, Lact. gallinarum, Lact. casei, Weissella minor and related taxa isolated from kocho.* J. Appl. Microbiol. 89:969–78.

65. Van Liempt E. et al. 2006. *Specificity of DC-SIGN for mannose- and fucose-containing glycans.* FEBS Lett. 580:6123–6131.

66. De Man JD, Sharpe ME, Rogosa M, Sharpe ME. 1960. *A medium for the cultivation of Lactobacilli.* J. Appl. Bacteriol. 23:130–135.

67. Vandamme P. et al. 1996. *Survey of enterococcal susceptibility patterns in Belgium.* J. Clin. Microbiol. 34:2572–6.

68. Sambrook J, Fritsch EF, Maniatis T. 1989. *Molecular Cloning: A Laboratory Manual.* Cold Spring Harbor laboratory pressNew York.

69. Lambert JM, Bongers RS, Kleerebezem M. 2007. *Cre-lox-based system for multiple gene deletions and selectable-marker removal in Lactobacillus plantarum.* Appl. Environ. Microbiol. 73:1126–35.

CHAPTER 5

Genetically diverse strains of *Lactobacillus rhamnosus* display fermentation dependent strain-specific immune modulation effects

Submitted manuscript. Authors: Corina Ceapa, Jeroen van Bergenhenegouwen, Jolanda Lambert, Michiel Kleerebezem, Jan Knol

Abstract:

Studying immune recognition of bacteria and in particular probiotic strains is of relevance to predict their potential application in prophylactic and therapeutic approaches to diseases. Co-culturing bacteria *in vitro* with primary human immune cells is commonly used as a first evaluation of a strain's pro- or anti-inflammatory potential. However, the methodology for (co-) culturing microorganisms and immune cells is not yet standardized and conditions vary greatly between laboratories. Lack of standardization prevents reliable comparative analysis of the results obtained. In addition to previously recognized variables like bacterial genetics, growth stage and nutrients, we establish here that controlling environmental factors like pH during bacterial culturing are also elements to take into account when evaluating the immune responses to probiotic strains *in vitro*. Initial studies determined the cytokine pattern (IL-10, IL-12, IFNγ, IP-10) induced by 4 *L. rhamnosus* strains in peripheral blood mononuclear cells (PBMCs). A subset of the *L. rhamnosus* strains tested, including Lrh15 and Lrh29, consistently elicited elevated levels of IL-12, TNFα and IFN-γ, suggesting that these *L. rhamnosus* strains could promote a Th1-skewing response. Approximately half of the *L. rhamnosus* strains tested elicited only limited cytokine production in PBMCs, but their spent culture supernatants were able to induce significant amounts of IL-10, suggesting the presence of secreted compounds that contribute to anti-inflammatory responses. These *in vitro* results support the stratification of *L. rhamnosus* strains in two major subtypes, one subtype that is more prone to induce a Th1-like response, and a second subtype that can activate IL-10 production via molecules in the culture supernatant.

Introduction

An emerging area of research refers to the use of specific members of the commensal microbiota for the prevention and therapy of disease [1]. These organisms are labelled probiotics or "live microorganisms that, when administered in adequate amounts, confer a health benefit to the host" (2001, World Health Organization) [2]. Most probiotics belong to the *Lactobacillus* and *Bifidobacterium* genera, but other bacterial genera like *Escherichia* and *Streptococcus* are also reported to provide health benefits [3–5]. Accumulating data indicate that probiotic effects are strain dependent, with strains of the same species eliciting various responses from the host [6, 7].

At a molecular level, bacteria provoke immunological responses in the host by activating pattern recognition receptors (PRRs). PRRs distinguish conserved motifs of the microbiota (microbial associated molecular patterns or MAMPs) and start a signalling cascade leading to transcription of genes associated with cytokine production and adaptive immune response [8]. PRRs are expressed on practically every cell of the body with the largest complement expressed by antigen-presenting cells, such as macrophages or dendritic cells (DC). Some of the PRRs involved in the interaction are the trans-membrane Toll-like receptors (TLRs), which scan the extracellular space, whereas the Nod-like receptors (NLRs) guard the intracellular cytoplasmic compartment. Several TLRs and NLRs are described to interact with components of bacteria. For example, TLR2 heterodimerizes with TLR1 or TLR6 to recognize various cell wall MAMPs like LTA, lipopeptides and peptidoglycan. In addition, TLR9 recognizes bacterial DNA and the NLR NOD2 has been shown to interact with cell wall-derived peptidoglycan moieties [8].

In vitro co-culture assays with bacteria and different types of immune cells, such as human peripheral blood mononuclear cells (PBMCs), human monocyte derived DCs, and mouse bone marrow derived DCs have often been used to assess the immuno-modulatory potential of different species and strains [9–11]. Although *in vitro* tools can ascertain the capabilities of bacteria to initiate or direct immune responses, these types of experiments suffer from poor translational value. Only in a very few published papers *in vitro* immuno-modulatory properties translated into *in vivo* efficacy [10-12]. Since MAMP-PRR interactions drive immunological responses, factors that infer changes in, or levels of, bacterial MAMP expression are important parameters to consider.

The main factors explored so far include bacterial genetics, growth phase and fermentation conditions. Gene-trait-matching analysis identified several genetic loci responsible for immune modulation in *L. plantarum* [12], including a N-acetyl-glucosamine/galactosamine transporter, the LamBDCA quorum sensing system, and

components of the plantaricin (bacteriocin) biosynthesis and transport operons. The bacterial growth phase can modulate immune responses as well, which has been shown *in vivo* in healthy volunteers by van Baarlen et al. [13]. The study showed stationary bacteria to induce changes in the gene expression profiles which correlated with transcription factor Nuclear Factor Kappa-B (NF-κB) modulation, whereas mid logarithmic cells of the same strain did not elicit these changes. In addition, immune-modulatory capacities of bacteria (probiotics) may be influenced by the bacterial metabolic activity profile, like the organic acid production, which are clearly shown to depend on specific growth conditions, including pH, temperature, and the availability of molecular oxygen [14]. In this study, we address these differences in fermentation conditions and show that fermentation conditions affect the way host-cells interact with bacteria.

Lactobacillus rhamnosus strains are commensal microorganisms usually present in the human gut [15], where they may interact with immune cells [16]. Here, we studied the behavior of immune cells stimulated with specific strains of *L. rhamnosus* isolated from different conditions of growth. In order to achieve controlled conditions we utilized pH-controlled batch cultivation and compared the immunomodulation capacity of these cells with those derived from uncontrolled, acidifying batch fermentations.

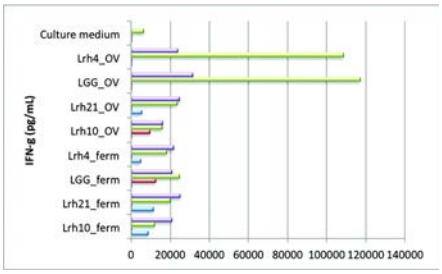
Results

Strain selection and differential activation of PBMCs.

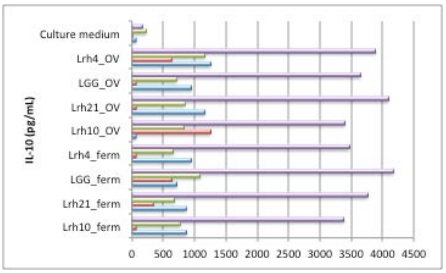
The genomic diversity of forty *L. rhamnosus* strains in terms of presence and absence profile of individual genes is presented in this thesis (Chapter 6), and the genome-diversity mapping allowed the classification of the strains in eight genomic-clades. The four strains selected for the present study (LRH4, LRH21, LGG, and LRH10) classify in different genomic clusters, intending to reflect the species diversity. The immunomodulatory profiles elicited by the four *L. rhamnosus* strains in peripheral blood monocyte cells (PBMCs) were investigated by co-incubation (ratio bacteria:PBMC of 10:1) for 24 hours, followed by determination of the cytokine profile in the culture supernatants. No detectable cytokine production was observed in unstimulated PBMCs (Figure 1), whereas each *L. rhamnosus* strain induced the production of significant amounts of at least some cytokines. Bacterial cells derived from uncontrolled batch cultures induced strain-specific levels of IFNγ, IP-10 and IL-12, whereas the level of Il-10 induced by these cells appeared highly similar (Figure 1). In particular, strains LGG and LRH21 appeared to induce up to six fold higher levels of the

pro-inflammatory cytokines (IFNγ, IP-10, IL-12) as compared to strains LRH10 and LRH15.

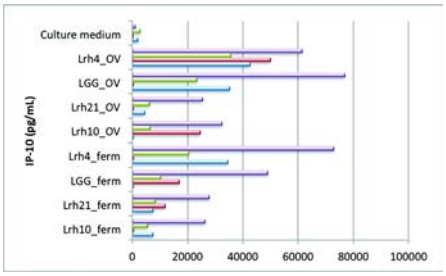
A.



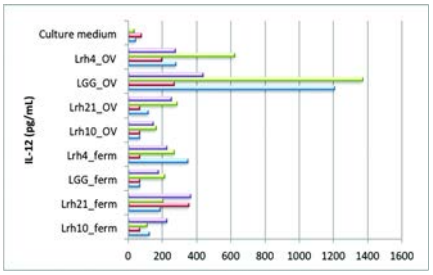
B.



C.



D.



E.

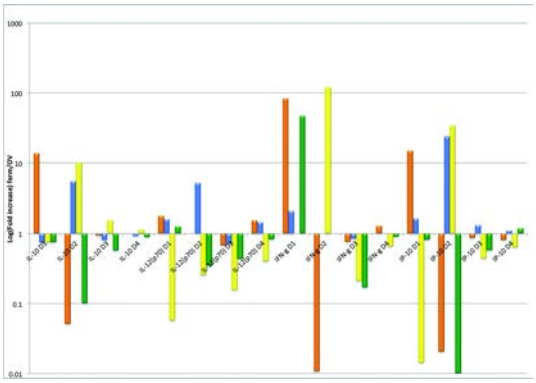


Figure 1. Cytokine profiles produced by peripheral blood mononuclear cells (PBMC) co-cultured for 24h in the presence of four *L. rhamnosus* strains (ratio bacteria: PBMC 10:1): A. IFN-γ ; B. IL-10; C. IL-12(p70); D. IP-10; Tested donors are: Donor 1 (blue), Donor 2 (red), Donor 3 (green), Donor 4 (purple). E. Log(fold increase) between values from the same donor for the fermented versus the overnight grown cells. The control sample corresponds to PBMC growing in RPMI medium. Strains are colour coded: LRH10 (orange), LRH21 (blue), LRH4 (yellow) and LGG (ATCC 53103) (green).

Bacterial cells derived from pH controlled batch cultures induced relatively similar levels of all cytokines. Only in the case of IP-10, there is an increase in production for LRH4 (up to 4 fold) and LGG (up to 2 fold) compared to the other strains. To minimise the effects of the donor variation, a log of the fold increase between values from the same donor for the fermented versus the overnight grown cells was calculated, allowing for the strain effects to be evident (Figure 1, E). While some strains induce the production of dramatically different quantities of cytokines in different donors (strain LRH10 has opposite trends in donor 1 versus 2 production of IL10, IP-10 and IFNγ while inducing average levels of the same cytokines in the other two donors), others show a consistent effect over all our donors (LRH4 induces similar levels of IL-12 in all donors), showing a large donor-to-donor cytokine production between individual strains.

In a second part of the study, when comparing uncontrolled versus pH controlled conditions, LGG and LRH21 cells derived from pH-controlled batch cultures induced approximately 4-fold lower IL-12 and IFNγ production levels compared to cells from uncontrolled batch cultures, leading to production levels of these cytokines that can no longer be discriminated from those induced by strains LRH10 and LRH15 (Figure 1).

This effect of the growth conditions was much less apparent in the capacity of strains LGG and LRH21 to induce IP-10, and also appeared to be completely absent in the levels of IL-10 induced by each of the strains.

These results illustrate that bacterial growth conditions can strongly influence, and potentially mask or modulate the relative immuno-modulatory capacity of individual strains of a species. In particular, more controlled growth conditions (pH control) appeared to suppress some of the strain-specific pro-inflammatory capacity, but whether this is valid for all strains of the species requires the evaluation of a larger panel of strains.

L. rhamnosus strains grown in pH-controlled batch cultures vary in their ability to induce pro-inflammatory responses in PBMCs

To investigate strain-specific stimulation of pro-inflammatory PBMC responses, pH-controlled batch cultures were prepared for 26 *L. rhamnosus* strains and the capacity of these bacterial cells to stimulate cytokine production in PBMCs was investigated (ratio bacteria: PBMC of 10:1). Besides bacterial cells, also their spent culture supernatants were used as a stimulus in these assays. PBMCs obtained from three healthy donors were used and the production of the cytokines IL-1β, IL-8, IL-10, IL-12, tumour necrosis factor α (TNFα), and IFNγ were determined (Supplementary Table 1). In unstimulated PBMCs the levels of each of the cytokines remained below the detection

limit. As a positive control for inflammatory responses LPS was shown to elicit high levels of the pro-inflammatory cytokines IL-12 and TNFα whereas this stimulus induced only low amounts of the anti-inflammatory cytokine IL-10. The 26 Lactobacilli elicited significant and differential cytokine production patterns (Supplementary Table 1). Similarly, spent culture supernatant elicited distinct immuno-modulatory profiles (Figure 2). To correct for the relatively large differences between the responses observed in individual donors, the cytokine production levels were normalized by calculating the relative fold-inductions by each individual stimulation compared to the average level of each cytokine detected in individual donors. This normalization procedure revealed that despite the variability of absolute cytokine levels per donor, the strains elicited similar cytokine induction patterns in each donor (Figure 2).

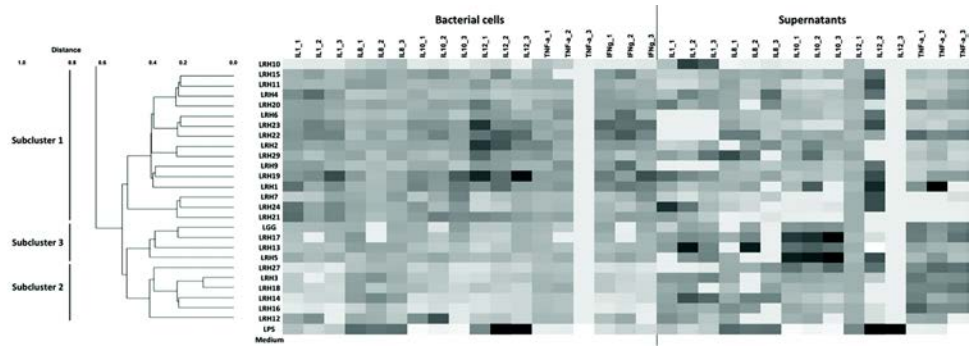


Figure 2. Clustered heatmap of cytokine (IL-1β, IL-8, IL-10, IL-12, TNFα, and IFNγ) production profiles elicited in human peripheral blood mononuclear cells (PBMCs, average? from 3 donors) by stimulation with 26 different *L. rhamnosus* strains or their supernatants. LPS was used as a positive control for the PBMCs stimulation. The continuous grayscale reflects the fold difference of cytokine production relative to the donor average, with white representing no fold induction compared to the donor's average, while black represents more than 2 donor average fold induction.

When using the bacterial cells as stimulus, the most pronounced differences were observed in the induction of IL-1β, IL-12 and IFNγ. While the discrimination is not dramatic, the immuno-stimulatory capacity of individual strains could be largely classified in two subclusters of strains., Strains LRH1, LRH19, LRH22 and LRH23 induced more than a 2.5-fold increase in IL-1β, IL-12, or IFNγ (Subcluster 1, Figure 2), whereas strains LRH3, LRH12, LRH14, LRH16, LRH18, LRH27 induced these cytokines at levels that were more than 2-fold lower as compared to the donor-specific average. (Subcluster 2, Figure 2). Contrary to the bacterial cell stimuli, the cytokines that were discriminatory when PBMCs were stimulated with the culture supernatants

were IL-10 and TNFα, although the supernatant stimulation appeared to be less pronounced and less consistent across donors. Especially the supernatants of strains LRH5, LRH13 and LRH27 elicited a significant and consistent increase of the IL-10 production levels (Subcluster 3, Figure 2), suggesting that these strains secrete some factor in their environment that is able to elicit anti-inflammatory responses in PBMCs.

Discussion

Immunomodulation by intestinal microbiota is recognized to play a prominent role in host health [17, 18]. Elucidating the mechanisms by which microorganisms exert these immuno-modulating capacities can support the use of bacteria that are selected for specific immuno-modulatory potential [12, 19]. In this regard, suitable and reliable cell-based model systems to test immune interaction with bacterial stimuli are essential [20–22]. Varieties of immune cells are found at different sites in the gastrointestinal tract mucosa. Epithelial cells lining the intestinal tract express pattern recognition receptors (PRRs), and they produce mediators to recruit, activate and condition cells of the immune system [23]. Effector immune cells like T and B cells are diffusely distributed throughout the lamina propria and embedded within the epithelium or located in subepithelial lymphoid aggregates (like Peyer's patches) with a role in priming adaptive immune responses [24]. The complexity of the intestinal mucosal immune system hampers *in vivo* screening of bacterial strains for their immuno-modulatory capacities. A comprehensive study of their diversity could lead to understanding in a comprehensive manner that aims at revealing their mechanism of interaction with the host. Therefore, simplified *in vitro* models are frequently used to profile the immune modulating capacities of individual strains, and these models enable the deciphering of the underlying mechanisms of immunomodulation at the molecular level. As a first evaluation towards a strain's pro- or anti-inflammatory potential, culturing bacteria *in vitro* with representative primary human immune cells is a commonly applied approach [1, 25]. Therefore, *in vitro* co-culture assays with bacteria and different types of immune cells, such as human PBMCs, human monocyte derived DCs, or mouse bone marrow derived DCs have been used to assess the immuno-modulatory potential of different species and strains [9, 18, 26]. These efforts have enabled the identification of several probiotic effector molecules, including DC-SIGN binding protein SlpA from *L. acidophilus* that modulates DCs and T cells functions [27] and the role of cell wall-associated polysaccharides in the suppression of macrophages activation by *Lactobacillus casei* strain Shirota [28, 29].

Lactobacillus strains are the most dominant species among the probiotic cultures used to date [30–32]. Despite the large volume of research in the last decade, the immune modulatory properties of *Lactobacillus* probiotics in *in vivo* and/or *in vitro* models are not necessarily consistent across different studies. In the pre-clinical phase one would preferably employ robust and strictly standardized methods that could enable the comparison of results obtained in different laboratories. However, such methodological standardization is currently lacking and as a consequence the *in vitro* immuno-modulatory properties of individual (probiotic) *Lactobacillus* strains is regularly inconsistent between different studies and laboratories [20, 22, 33, 34]. The present study underpins the notion that standardization of the experimental design in these approaches is critical to obtain consistent immune modulation outcomes *in vitro*. To this end, a panel of *L. rhamnosus* strains with an established genomic variation of up to 3% at gene level (Chapter 6) was employed. To reduce variability introduced by culturing methods, these strains were grown in pH-controlled batch cultures rather than uncontrolled batches. Findings demonstrated in this study support this approach by indicating that controlling environmental factors such as pH affects immune responses, thus supporting the rigorous maintenance of such culture condition variables.

The present study is one of several which highlight that different cytokine induction profiles can be measured in *in vitro* stimulation assays using different strains of the same species of *Lactobacillus* [20, 22, 34–36], including *L. rhamnosus* [20, 22, 33, 37–39]. However, to the best of our knowledge, this is the first study indicating the importance of controlled fermentation conditions. To assess the immuno-modulatory effect of the strains, we quantified IL-1β, IL-8, IFNγ, TNFα, IL-12, IP-10 and IL-10 cytokine production by PBMCs stimulated with the different strains. These cytokines play differential roles in skewing the immune system towards more pro-inflammatory responses (IFNγ, TNFα, IL-12) or more anti-inflammatory responses (IL-10). There is a relative consistent pattern in these responses using PBMC cells of different donors, despite the substantial differences in absolute levels of these cytokines produced by PBMCs from individual donors (Figure 1E). These data clearly illustrate that *in vitro* immuno-modulatory capacities of bacterial strains and their spent culture supernatant are strain specific, but that there are also trends shared among strains that allow their classification into more or less pro- or anti-inflammatory subclusters. In this context, it is remarkable that the supernatants of strains LRH6, LRH22 and LRH23 induced virtually no IL-1β, whereas, neither the cells nor the supernatants of strain LRH10 appeared to consistently induce significant production of any of the measured cytokines, illustrating that within such subclusters there are still strain-specific and discriminating

immuno-modulatory properties of cells and their supernatants. Notably, the mechanisms underlying these variations among strains (or strain subclusters) remain unknown, although it seems plausible that certain bacterial effector molecules that may be conserved among subclusters of strains are involved in the interaction with corresponding host cell receptors and steer the cellular responses [37, 40–42].

The work presented could support mechanistic studies that aim to understand the strain specific, immune system associated health effects. Importantly, standardization of bacterial growth conditions appears to play a prominent role in the *in vitro* immunomodulatory effects recorded for individual strains, underpinning the dynamic nature of the immune system. Therefore it is essential that *in vitro* immuno-profiling efforts as an approach to reveal mechanism of interaction and identification of bacterial effector molecules employ highly standardized methods for production of the bacterial stimuli used.

Materials and Methods

Bacterial strains and growth conditions

In vitro testing was performed on panels of 4 and 26 *L. rhamnosus* strains with distinct genotypic profiles (Chapter 3, Chapter 6).

For immune profiling of uncontrolled (37°C, static flask) and pH-controlled batch cultures (37°C, 160 rpm agitation, pH 7; 0.45L DASGIP fermenter system, Eppendorf, Julich, Germany, pH was maintained by automatic titration of 2 M NaOH), the *L. rhamnosus*, strains were grown in Mann-Rogusa Sharpe (MRS) broth (Merck, Darmstadt, Germany) until stationary phase (for batch: 16 h after the start of the culture, OD600 nm = 3.6 ± 0.4; for fermenter: half an hour after acid production has stopped, OD600 nm = 4.2 ± 0.5). Pre-cultures originated from the same colony and were always inoculated at the same time at a 1% concentration (v/v).

Bacterial cells were washed twice in phosphate buffered saline (PBS) at pH 7.4, and resuspended at a density of 2×10^7 Colony forming units (CFU)/ml prior to being introduced in PBMC co-cultures. CFUs were enumerated by plating serial dilutions of the cultures on MRS agar (data not shown).

Cell preparation and *in vitro* culture

Human peripheral blood mononuclear cells (PBMCs) were obtained from standard buffy-coat preparations from routine blood donors, after obtaining informed consent, using Ficoll-Isopaque density gradient centrifugation (Ficoll-Paque, Pharmacia, Upsalla, Sweden). Freshly isolated (< 4 hours) PBMC were subsequently co-cultured with bacteria in a 1:10 ratio (PBMC: CFU) in RPMI 1640 (Gibco, Breda, The Netherlands)

supplemented with 1% l-glutamine (Gibco), 50 I.U./mL Penicillin-Streptomycin (Gibco) to prevent bacterial overgrowth.

Cytokine analysis

Co-culture supernatant samples were analysed by the multiplex suspension array system using Bio-plex Pro beads (Bio-Rad Laboratories), determining the cytokines IL-10, IL-12 (p70), IP-10, IFN-γ, and TNF-α. Supernatant samples were diluted 1:2 with Bio-Plex human serum sample diluent (171-305000), and run according to the manufacturer's protocol. All samples were analysed in duplicate and cytokine levels detected are expressed as picograms per millilitre of supernatant. LPS (L3024 Sigma, *E. coli*) was used as a positive control. Data analysis was done with the Bio-plex Manager 5.0 Standard Software.

While raw data is presented for the cytokine comparison between batch and culture controlled bacteria, in the case of the responses from 26 strains, data was donor average normalized and included in a clustering analysis. Raw data was processed using PAST software [49] and is available as Supplemental Table 1. Cytokine data was incorporated in a multivariate matrix, and underlying structures within and between the groups of strains were explored using clustering analyses in the PAST software suite [49]. Clustering was performed based on Euclidian distances calculated with raw data using unweighted pair group method with arithmetic mean algorithm (UPGMA).

Statistics

Experimental data are expressed as mean + S.E.M. The statistical significance of differences in mean values was determined with the student's t-test in the PAST software. The control sample corresponds to PBMC growing in RPMI medium for 24h. The t-test for 2 paired samples was used to assess differences between the same strains grown in batch or chemostat conditions.

References:

1. Ceapa C. et al. : *Influence of fermented milk products, prebiotics and probiotics on microbiota composition and health.* Best Pract Res Clin Gastroenterol 2013, 27:139–55.

2. Dicks LM, Botes M: *Probiotic lactic acid bacteria in the gastro-intestinal tract: health benefits, safety and mode of action.* Benef Microbes 2010:11–29.

3. Ouwehand AC, Salminen S, Isolauri E: *Probiotics: An overview of beneficial effects. Antonie van Leeuwenhoek*, Int J Gen Mol Microbiol 2002, 82:279–289.

4. Schultz M: *Clinical use of E. coli Nissle 1917 in inflammatory bowel disease.* Inflamm Bowel Dis 2008:1012–1018.

5. Ménard S. et al. : *Lactic acid bacteria secrete metabolites retaining anti-inflammatory properties after intestinal transport.* Gut 2004, 53:821–828.

6. Vissers YM. et al. : *Lactobacillus strains differentially modulate cytokine production by hPBMC from pollen-allergic patients.* FEMS Immunol Med Microbiol 2011, 61:28–40.

7. Snel J. et al. : *Strain-specific immunomodulatory effects of Lactobacillus plantarum strains on birch-pollen-allergic subjects out of season.* Clin Exp Allergy 2011, 41:232–242.

8. Wells JM: *Immunomodulatory mechanisms of lactobacilli.* Microb Cell Fact 2011:S17.

9. De Kivit S, Kraneveld AD, Garssen J, Willemsen LEM: *Glycan recognition at the interface of the intestinal immune system: Target for immune modulation via dietary components.* In Eur J Pharmacol. Volume 668; 2011.

10. McKim JM: *Building a tiered approach to in vitro predictive toxicity screening: a focus on assays with in vivo relevance.* Comb Chem High Throughput Screen 2010, 13:188–206.

11. Punt A. et al. : *Effect of combining in vitro estrogenicity data with kinetic characteristics of estrogenic compounds on the in vivo predictive value.* Toxicol Vitr 2013, 27:44–51.

12. Meijerink M, Wells JM: *Probiotic modulation of dendritic cells and T cell responses in the intestine.* Benef Microbes 2010:317–326.

13. Baarlen P Van. et al. : *Differential NF- kB pathways induction by Lactobacillus plantarum in the duodenum of healthy.*PNAS 2009, 106(7):2371-6.

14. Bron PA. et al. : *Transcriptomes reveal genetic signatures underlying physiological variations imposed by different fermentation conditions in Lactobacillus plantarum.* PLoS One 2012, 7(7):e38720.

15. Heilig HGHJ. et al. : *Molecular diversity of Lactobacillus spp. and other lactic acid bacteria in the human intestine as determined by specific amplification of 16S ribosomal DNA.* Appl Environ Microbiol 2002, 68:114–123.

16. Ley RE, Peterson D a, Gordon JI: *Ecological and evolutionary forces shaping microbial diversity in the human intestine.* Cell 2006, 124:837–48.

17. Van Baarlen P, Wells JM, Kleerebezem M, Baarlen P Van: *Regulation of intestinal homeostasis and immunity with probiotic lactobacilli.* Trends Immunol 2013, 34:208–215.

18. Martín R, Miquel S, Ulmer J, Langella P, Bermúdez-Humarán LG: *Gut ecosystem: How microbes help us*. Benef Microbes 2014:219–233.

19. Bron P a, Tomita S, Mercenier A, Kleerebezem M: *Cell surface-associated compounds of probiotic lactobacilli sustain the strain-specificity dogma.* Curr Opin Microbiol 2013, 16:262–9.

20. Ghadimi D. et al. : *Effects of probiotic bacteria and their genomic DNA on TH1/TH2-cytokine production by peripheral blood mononuclear cells (PBMCs) of healthy and allergic subjects.* Immunobiology 2008, 213:677–92.

21. Troy EB, Kasper DL: *Beneficial effects of Bacteroides fragilis polysaccharides on the immune system.* Front Biosci 2010, 15:25–34.

22. Elmadfa I, Klein P, Meyer AL: *Immune-stimulating effects of lactic acid bacteria in vivo and in vitro.* Proc Nutr Soc 2010, 69:416–420.

23. Michallet MC, Rota G, Maslowski K, Guarda G: *Innate receptors for adaptive immunity*. Curr Opin Microbiol 2013:286–302.

24. Mowat AM, Agace WW: *Regional specialization within the intestinal immune system.* Nat Rev Immunol 2014, 14:667–685.

25. Ottman N, Smidt H, de Vos WM, Belzer C: *The function of our microbiota: who is out there and what do they do?* Front Cell Infect Microbiol 2012.

26. Hooper L V, Gordon JI: *Commensal host-bacterial relationships in the gut.* Science 2001, 292:1115–1118.

27. Konstantinov SR. et al. : *S layer protein A of Lactobacillus acidophilus NCFM regulates immature dendritic cell and T cell functions.* Proc Natl Acad Sci U S A 2008, 105:19474–19479.

28. Yasuda E, Serata M, Sako T: *Suppressive effect on activation of macrophages by Lactobacillus casei strain Shirota genes determining the synthesis of cell wall-associated polysaccharides.* Appl Environ Microbiol 2008, 74:4746–55.

29. Dong H, Rowland I, Yaqoob P: *Comparative effects of six probiotic strains on immune function in vitro.* Br J Nutr 2012, 108:459–70.

30. Klaenhammer TR. et al. : *Functional genomics of probiotic Lactobacilli.* J Clin Gastroenterol 2008, 42 Suppl 3:S160–S162.

31. Marco ML. et al. : *Convergence in probiotic Lactobacillus gut-adaptive responses in humans and mice.* ISME J 2010, 4:1481–4.

32. O'Sullivan O. et al. : *Comparative genomics of lactic acid bacteria reveals a niche-specific gene set.* BMC Microbiol 2009, 9:50.

33. Iliev ID. et al. : *Strong immunostimulation in murine immune cells by Lactobacillus rhamnosus GG DNA containing novel oligodeoxynucleotide pattern.* Cell Microbiol 2005, 7:403–414.

34. Boyle RJ. et al. : *In vivo, in vitro and transplacental immune effects of a probiotic bacterium Lactobacillus rhamnosus GG in humans.* J Allergy Clin Immunol 2009, 123:S200–S200.

35. Lehtoranta L, Pitkäranta A, Korpela R: *The effects of probiotic combination Lactobacillus rhamnosus GG and Lactobacillus rhamnosus Lc705 in cytokine and chemokine response in human macrophages.* Int J Probiotics Prebiotics 2012, 7:17–22.

36. Ibnou-Zekri N, Blum S, Schiffrin EJ, von der Weid T: *Divergent patterns of colonization and immune response elicited from two intestinal Lactobacillus strains that display similar properties in vitro.* Infect Immun 2003, 71:428–436.

37. Plantinga TS. et al. : *Differential Toll-like receptor recognition and induction of cytokine profile by Bifidobacterium breve and Lactobacillus strains of probiotics.* Clin Vaccine Immunol 2011, 18:621–628.

38. Veckman V. et al. : *Streptococcus pyogenes and Lactobacillus rhamnosus differentially induce maturation and production of Th1-type cytokines and chemokines in human monocyte-derived dendritic cells.* J Leukoc Biol 2004, 75:764–771.

39. Villena J. et al. : *Orally administered Lactobacillus rhamnosus modulates the respiratory immune response triggered by the viral pathogen-associated molecular pattern poly(I:C).* BMC Immunol 2012:53.

40. Kleerebezem M, Vaughan EE: *Probiotic and gut lactobacilli and bifidobacteria: molecular approaches to study diversity and activity.* Annu Rev Microbiol 2009, 63:269–90.

41. Vanderpool C, Yan F, Polk DB: *Mechanisms of probiotic action: Implications for therapeutic applications in inflammatory bowel diseases.* Inflamm Bowel Dis 2008, 14:1585–1596.

42. Bron P. et al. : *Lactobacillus plantarum possesses the capability for wall teichoic acid backbone alditol switching.* Microb Cell Fact 2012, 11:123.

CHAPTER 6

Variable regions of *Lactobacillus rhamnosus* genomes suggest strong strains segregation by horizontal gene transfer events

Submitted manuscript. Authors: Corina Ceapa, Mark Davids, Jarmo Ritari, Jolanda Lambert, Michiel Wels, François P. Douillard, Tamara Smokvina, Willem M. de Vos, Jan Knol, Michiel Kleerebezem.

Abstract

Background: *Lactobacillus rhamnosus* is a diverse Gram positive species with strains isolated from different ecological niches. Here, we report the sequencing of 40 diverse strains of *L. rhamnosus* and genomic comparison of these strains, with a focus on the variable genome.

Results: Genomic comparison of 40 *L. rhamnosus* strains discriminated the conserved genes (core genome) and regions of plasticity involving frequent rearrangements and horizontal transfer (variome). The *L. rhamnosus* core genome encompasses 2164 genes, out of 4711 genes in total (the pan genome). The accessory genome is dominated by genes encoding carbohydrate transport and metabolism, EPS biosynthesis, bacteriocin production, pili production, the *cas* system and the associated CRISPR loci, and more than 100 transporter functions and mobile genetic elements like phages, plasmid genes and transposons. Classification of the genomes on basis of core genes aminoacid differences generated a clade distribution that matched with the variome-derived clade distribution. The clades are dominantly determined by horizontal gene transfer (HGT) events, which are paralleled by evolutionary diversification of core gene functions, exemplifying the importance of horizontal transfer in the evolutionary segregation of this species. The CRISPR-Cas system could have contributed to this evolutionary segregation.

Conclusions: *L. rhamnosus* strains contain the genetic and metabolic machinery and strain specific gene functions required to adapt to a large range of environments. A remarkable congruency of the evolutionary relatedness of the strains' core and variome functions, possibly favoring inter-species genetic exchanges, underlines the importance of gene-acquisition and loss within the *L. rhamnosus* strain diversification.

Keywords

Comparative genomics, diversity, niche adaptation, probiotic, core and pan genome.

Background

The genus *Lactobacillus* encompasses a phylogenetically diverse group of bacteria and currently amounts to more than 180 recognized species. *Lactobacillus* strains have an important industrial role, being used throughout centuries for the preparation of fermented foods, such as yoghurt, fermented olives, cheese, beer, wine, cocoa. More recently, several *Lactobacillus* species are also studied for their potential to contribute to a health supporting diet [1]. Already more than a century ago, scientists recognized prospective benefits for human health that could be conferred by the consumption of *Lactobacillus* containing products [2]. Their industrial versatility relies on a large phenotypic diversity, of which the genetic origin has not yet been elucidated in its entirety, although genomics of the lactobacilli support their extensive functional variation [3]. Genomic approaches have provided insight in the evolutionary relationships among the *Lactobacillus* species, and have also revealed some genes and functions that explain their presence in particular ecological niches like the gastrointestinal tract [4], fermented dairy products [5], and plant associated environments [6]. A combination of gene gain and loss plays a prominent role during niche adaptation, where particularly gene loss appears to be associated with the adaptation to the nutrient rich dairy environment [4].

Together with *Lactobacillus plantarum*, *Lactobacillus reuteri* and *Lactobacillus casei*, *L. rhamnosus* has the largest genome among the species of the *Lactobacillus* genus or the lactic acid bacteria (LAB) in general. *L. rhamnosus* is an anaerobic, facultative heterofermentative rod-shaped microorganism that is encountered in the fermentation of food and feed-raw materials [7], but also frequently found in the human and animal intestinal [8], oral and uro-genital system [9].

While originally used for their ability to produce acid during fermentation, which lowers the pH and prevents the growth of spoilage bacteria, *L. rhamnosus* are also employed in cheese and yoghurt manufacturing for their proteolytic and flavor generating capacities [10], and play a key role in the production of fermented milk, cheese, sauerkraut [11] and kimchi [12]. In addition, several *L. rhamnosus* strains fulfill the basic requirements that support their marketing as probiotics, including their capacity to survive passage through the human gastro-intestinal tract (GIT) after ingestion and the transient colonization or persistence in the ileum and colon [13], their safety for human consumption [14] and their efficacy in human intervention studies [15]. However, for many of these features, the underlying molecular mechanisms remain to be determined. Among its most studied strains, *L. rhamnosus* GG (ATCC 53103) stands apart as the most documented probiotic strain (887 scientific articles and

58 patents by 2015 – according to the Scopus Database). Health benefits potentially associated with the human consumption of *L. rhamnosus* GG include rotavirus diarrhea prevention [16] and a reduction of respiratory infections in children [17]. Moreover, the HN001 strain was reported to prevent eczema development [18], whereas strain GR-1 showed potential towards treatment of bacterial vaginosis [19]. In mice, among other potentially probiotic *Lactobacillus rhamnosus* strains, the JB-1 strain was suggested to modulate enteric neuron excitability, which may impact on the gut - brain axis by means of the pneumogastric nerve [20] and is thought to contribute to behavioral changes observed in mice to which this strains was administered [21]. These different health effects were associated with different strains of *L. rhamnosus*, but due to the lack of controlled comparative studies, it remains to be established if and to what extent all these effects are strain-specific [22].

While for some of the health benefits and niche adaptation factors, the mechanism of action of *L. rhamnosus* strains remains unclear, some of the molecular mechanisms that may underlay or contribute to these effects have been studied [23]. For example, expression of mucus-binding pili in the intestine [24] and MBF, an active mucus-specific surface adhesin [25], play a role in the adherent mechanisms that contribute to intestinal colonization by *L. rhamnosus* GG. In addition, glycosylated muramidases p40 and p75 were shown to inhibit apoptosis in mouse colon epithelial cells and cultured mouse colon explants treated with TNF-α, thereby positively influencing epithelial layer integrity and repair [26], which could contribute to reinforcement of the epithelial barrier [27]. Moreover, the large surface protein MabA was shown to play a role in epithelial adhesion of *L. rhamnosus* cells, and its expression correlated with increased levels of bacterial adhesion to intestinal epithelial cells and tissues of the murine GIT [28].

However, these first mechanistic insights do not fully explain the complex interaction *L. rhamnosus* strains develop with their host. In this context, a detailed pangenomic analysis of a substantial number of *L. rhamnosus* strains might potentially provide molecular discriminators that can be associated with functional variations within the species, including their differential capacity to affect the host's (mucosal) physiology. In this context, the previously reported sequence scanning of 100 strains of *L. rhamnosus* [29] allowed the determination of the core genome of the *L. rhamnosus* species. However, the exclusive mapping of the sequence information onto the *L. rhamnosus* GG template genome restricted the full species diversity analysis to functions that are present in *L. rhamnosus* GG.

Since the variable traits of the species, which are absent from *L. rhamnosus* GG

can provide novel insights into the species' diversity, this study presents the *in silico* genomic and metabolic analysis of the diversity of the *L. rhamnosus* species using the genomic sequences of 40 strains of the species, that were isolated from various environmental niches. The main focus of the comparative genome analysis was targeted to the analysis of the distribution and function of the 'variome' of the *L. rhamnosus* species, i.e. the genes that do not belong to the previously reported core genome.

Results and discussion

General features of *L. rhamnosus* genomes

Strains of the species *L. rhamnosus* originate from a variety of ecological niches. In the present study, we describe the sequencing and comparative genomics of 40 *L. rhamnosus* strains isolated from different sources like dairy fermented products, plants, human and animal intestine and human clinical samples (Supplementary Table 1). A high number of isolates were obtained from the mammalian intestinal tract environment, i.e. there are 8 isolated from human feces, 3 from the intestine of healthy individuals and one from goat feces. In addition, *L. rhamnosus* strains of other human origins were also included, i.e. 17 strains from hospitalized patients, including blood (infection) isolates and a vaginal isolate. The group is completed by seven dairy isolates and one beer isolate.

Genome size, G+C content and protein coding sequences (CDS) counts vary among strains, independent of the strain's niche of isolation. The average genome size is 3 Mb ±0.2 Mb, thereby placing the *L. rhamnosus* genomes amongst the largest within the *Lactobacillus* group [30]. The molecular percentage G+C content varied among sequenced genomes from 46.5 to 46.8 % (Supplementary Table 1). The mean percentage G+C content of 40 strains was comparable to the closely related *L. casei* species (46.6% compared to 46.3%) [31] but slightly higher than that of *L. plantarum* (44,5%) [32, 33] or the average for the lactobacilli that is estimated at 42.4% [34].

The newly sequenced genomes had average sequence coverage of approximately 14-fold (ranging from 9- to 22-fold) and could be assembled into 61 to 535 contigs (with an average number of contigs of 178). Due to the large number of strains analyzed, the number of gaps of each genome will have a lower influence on the OG estimation. We considered this level of coverage sufficient for comprehensive and comparative characterization of the pangenome and variome diversity among these strains, and thus providing an adequate reflection of the diversity of the species.

Since previous work has already established the *L. rhamnosus* core genome at high resolution [29], we here employed this core genome definition to define the species' pangenome and focus on the pangenome sub fraction that defines the variable part of the of the *L. rhamnosus* strains analyzed here. This focus explores the species' variome to identify the basis of strain (and clade) variability, as well as potential evolutionary pathways for the species *L. rhamnosus*, while using the characteristics of the core genome only to highlight the relevance of certain predicted functional variations.

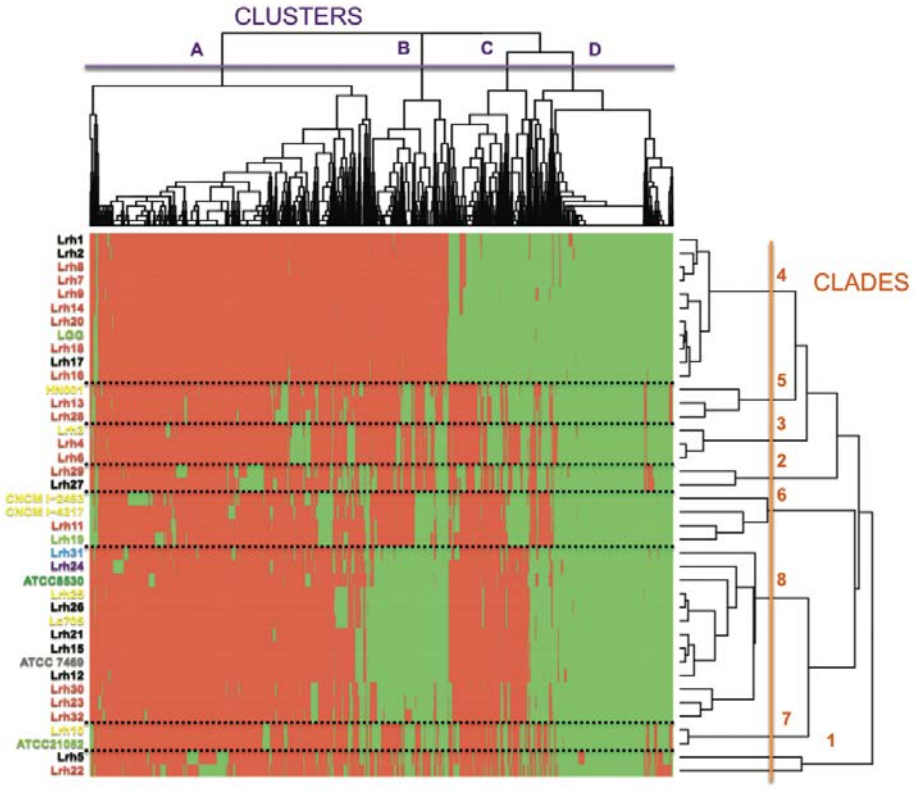


Figure 1. Hierarchical map and variome tree of the 40 *L. rhamnosus* genomes based on presence - absence of genes (green – presence, red – absence of genes). Orthologous groups of genes could be classified in four groups noted cluster A-D. Cluster D includes 10% of core genes to allow a shared-OGs skewed classification of the clusters. At strain level there are eight recognizable clades of strains (1-8). The niche from which the strains were isolated is indicated by the color-coding of the strain's ID-code: yellow: dairy; black – human feces; red: clinical (blood); light green: healthy intestine; blue: vagina; purple: goat feces; dark green: beer; grey: type strain (unknown).

The total predicted genetic content encompassed 4711 orthologous groups (OGs; forming the pan genome) with an average of 2707 (-/+74) OGs per genome, out of which 2164 are present in all genomes [29] (Supplementary Table.2). The size of the pan-genome is therefore higher than the one reported for other LAB species: around 2800 OGs for *Oenococcus oeni* [35], 4200 OGs for *L. casei* [5] and *L. paracasei* [3]. In our assessment, neither the core- nor the pan-genome estimates appear to have reached saturation (Supplementary Figure 1), implying that the total diversity inherent to the species remains still to be determined. Putative biological functions were assigned to 2761 (58%) of the predicted OGs and another 1272 (27%) OGs share sequence conservation with conserved proteins of unknown function in other organisms (Supplementary Table 2). Among the conserved core genome OGs, 67% have their closest relative sequence in *L. casei* and *L. paracasei*, followed by *Lactobacillus zeae* (30%), *Lactobacillus plantarum* (1%), *L. pentosus*, *Oenococcus oeni* and *Streptococcus* species' (Supplementary Table.2), which is in good agreement with previous taxonomic classification of *L. rhamnosus* [31].

In brief, the core genome contains genes needed for replication, transcription, translation, central and cell wall metabolism, biosynthesis of most amino acids and metabolism of nucleotides, fatty acids and phospholipids. Hypothetical proteins having a yet unknown function amount to more than 682 OGs of the core genome (32% of the core genome). The core genome also contains more than 18 complete sugar utilization gene clusters, and a variety of cell-surface components, discussed in detail below.

The variable gene content of the analyzed genomes was investigated using the gene presence/absence matrix and visualized by creating a heatmap using R (Figure 1) and used to construct a variome or ‘‘pan- genome’’ tree based on variable genome content (variome tree; Figure 1). The variome tree indicates that the strains can be separated in 8 main groups: clades 1 to 8. The clade-classification identifies *L. rhamnosus* unique and clade-specific OGs (Supplementary Figure 4). Species-specific orphans OGs (OOGs) that are uniquely present in *L. rhamnosus* and missing in all other bacteria (based on the entries in the NCBI nr database), amounted to 4,9% of the *L. rhamnosus* pangenome. The present study identified a total of 232 OOGs in the *L. rhamnosus* group, which includes a substantial amount of small-sized ORFs (more than 10% have less than 75 aa) and may therefore be somewhat overestimated. Notably, 37 OOGs are shared among all *L. rhamnosus* strains and belong to the core genome of the species, but they consistently lack a function-prediction.

Clade-specific OGs are present in all members of a specific variome-based genetic clade and absent in all other genomes. As can be expected from the variome-based

clade predictions (Figure 1), there were large differences between the numbers of clade-specific OGs. Clade 8 has the lowest number of clade specific OGs (13) followed by clade 4 (48), while clade 7 has the highest, 115 OGs (Supplementary Table 3). Global function scanning among the clade-specific gene sets, indicated some clade specific functional enrichments: clade 1, carbohydrate transport and metabolism, and protein and nucleotide degradation functions; clade 2, phage and protein transport and degradation; clades 3, 6 and 8, glycan degradation and carbohydrate transport; clade 4, amino acid transport, cell wall metabolism and restriction-modification; clade 5 and 7, restriction-modification.

Phage-related genes are present in all strains but in reduced amounts in strains of clades 4 and 5, whereas strains of clades 4 and 7 appear to lack plasmids. Notably, the CRISPR-Cas system appeared to be entirely missing from strains in clades 2 and 8, while this system is consistently present in all other strains of the other clades. Below, we will describe in detail a selection of functional characteristics that appear to be distributed in a clade-specific manner.

The number of strain-specific OGs fluctuates from 0 to 100 per strain and amounted to 706 OGs (15%) in total (Supplementary Table 5), which is most probably a somewhat overestimated number, due to the inclusion of small ORFs (see also above), and the potential over assignment of ORFs in general. Nevertheless, some of the strain-specific genes are predicted to encode functions that may be acquired recently and/ or confer advantages in certain environments or represent specific genetic entities. The strains Lrh22, CNCM I-2493 and Lrh24 appear to encode more than 50 unique genes each, and thereby standout from the group of strains. Notably, in strain Lrh22 among those unique genes is a plasmid replication associated gene, which may imply that some of the other strain-specific genes may also be located on that plasmid. Analogously, some other strains (Lrh3, Lrh23 and Lrh32) also appear to contain possible plasmids, contributing to their diversity. Notably, the only beer isolate, ATCC8530, contains only two strain-specific functions (pediocin pepB, and thioredoxin) that display very high level homology with *L. pentosus* and could play a role in the elevated resistance to hops in the presence of ethanol [79] reported for this strain. Some of the strain-specific genes also appear to reflect gene exchanges with other species, like the presence of an apparently complete prophage and an arsenic resistance operon in strain CNCM I-2493, which appear to be almost identical to genetic entities in *L. casei*. Analogously, strain Lrh24 encodes 5 prophage pi1 genes and a complete polysaccharide cluster (10 genes) with very high identity with genes found in *L. paracasei*.

Genomic prediction of *L. rhamnosus* metabolic diversity

Efficient use of environmental resources, especially carbohydrates, is relevant for the survival of bacteria in different habitats, including in the intestine, which is characterized by highly diet and host dependent nutrient availability. Metabolism genes assigned within the variome are dominated by carbohydrate transport and associated metabolic pathways, supporting the idea that sugar utilization is important for the evolution and adaptation of *L. rhamnosus* as a species [22], which is in agreement with observations made in other *Lactobacillus* species such as *L. reuteri* [36] and *L. casei* [22] (Supplementary Table 7).

Carbohydrate import and utilization

One of the largest functional classes encoded by *L. rhamnosus* genomes - transport proteins (in total 418 OGs) - is dominated by proteins predicted to be involved in carbohydrate import (44% of transporters with a predicted substrate) (Figure 2, Supplementary Table 6). Notably, for 68 transport-associated OGs (62 ABC and 6 PTS OGs) no substrate specificity could be predicted, and these functions could further expand the substrate utilization capacities of *L. rhamnosus* strains. The average number of transport associated genes per genome is 315 ± 12, a number comparable to the number of transport functions encountered in the genome of the versatile *L. plantarum* WCFS1 [32]. A large fraction of these transporters (approximately 40%) have a clear distribution pattern among the different strains and clades, enabling the discrimination of clades on basis of these functions (Supplementary Table 6).

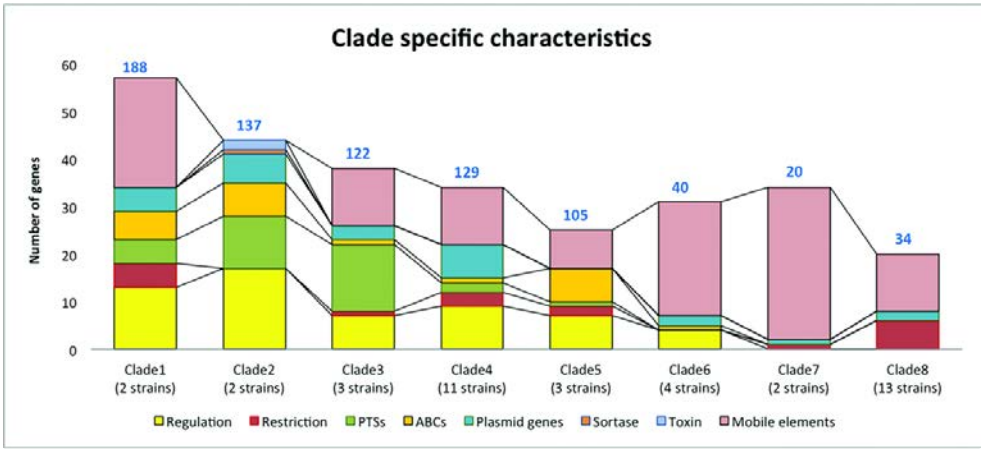


Figure 2. Number of clade specific genes and their functionality. Annotated genes are displayed; the number of clade specific genes with unknown functionality is added on top of each bar.

This observation is congruent with other studies that already identified strain-specific carbohydrate utilization capacities as a means of functional classification of *L. rhamnosus* strains [37]. In the genome of strain GG [38], the genes encoding ATP-binding cassette (ABC) transporters appear to be evenly distributed across the bacterial chromosome, whereas a regional enrichment is observed for phosphotransferase systems (PTSs) in the region at the 5'-side of the origin of replication (Supplementary Figure 3). This functional enrichment in a particular region of the chromosome is analogous to the positioning of the so-called 'sugar-island' of *L. plantarum* WCFS1 [32]. These regions also display a high degree of variability among strains of the species [39], indicating local genome plasticity and implying that PTSs transporters might be more frequently acquired and/or lost by horizontal gene transfer as compared to ABC transporters in both these *Lactobacillus* species. The main driver is probably that PTS systems are always carbohydrate transport systems, while ABC transporters are also known to transport other compounds (AA, metals, small molecules). This may imply also that ABC transport functions have less redundancy and are therefore more valuable as compared to the PTSs that could serve to some extent expendable or redundant functions. Nevertheless, 20 of the 134 PTS associated OGs (20 PTSs), appear to be conserved among all strains and are predicted to transport the monosaccharides glucose, galactose, mannose, fructose, the polyols sorbitol, and mannitol, and the disaccharides cellobiose, trehalose, sucrose, or oligosaccharides (glucosides). The import and utilization of this panel of carbohydrate sources for growth is consistent with the phenotypic analyses of strains of the *L. rhamnosus* species [29, 37]. The remaining PTS encode transporters predicted to import the additional carbohydrates maltose, arbutin, sorbose, N-acetylgalactosamine, galactosamine, and lactose, but include also systems that have a predicted substrate redundancy relative to the core-genome associated PTS functions (Supplementary Table 6). Of the overall 163 ATP-binding cassette (ABC) transporters (the gene number varying between 133 and 148 genes per genome; Supplementary Table 6), 72 % belong to the core genome functions (118 OGs forming between 99 to 108 units). These core genome ABC transport systems are predicted to encode both importers and exporters, with substrate assignments that include carbohydrates, amino acids, lipids, and metals (Supplementary Table 6). The carbohydrate ABC transporters encoded within the variome are predicted to mediate maltose/maltodextrin transport in strains of clade 1 and ribose/ D-xylose importers in all clades except clade 4.

Variation in the transporters encoded by any strains can be explained by strain-specific gene loss in comparison to a common ancestor [31] or horizontal gene transfer

acquisition from bacteria that co-inhabit the same niche [40–42]. In some cases the latter evolutionary process can be recognized in the clade distribution of these functions. For example, the sorbose PTS transport function that encompasses 8 genes, is only present in strains belonging to clades 5 to 8, and shares very high sequence identity (>80%) with the sorbose transporter PTSs identified in *L. casei* and *L. paracasei* species, which could imply that *L. rhamnosus* strains acquired these genes from these close relative species. The selective loss of these OGs in several strains of all three species, is likely to illustrate that this event occurs relatively frequently in representatives of the *L. casei* taxonomic group [31]. Analogously, an operon encoding 5 OGs annotated as a ribose ABC transporter is absent in all strains of clade 4 (except strain Lrh2), whereas a taurine ABC transporter is absent in several strains of clades 6 and 8. The OGs belonging to these functions have close homologues in the genomes of *L. casei* and *L. zeae*, and may therefore also represent genetic traits that are exchanged among these species.

Several of the transporters do not appear to follow a clade-specific distribution and appear to be present in various strains of different clades. An N-Acetyl Galactosamine PTS (GalNAc) transporter is present in all strains, but a functionally redundant and genetically distinct PTS appears to be acquired by horizontal gene transfer by various strains of clades 2, 3, 6, 7 and 8. Gene acquisition by horizontal gene transfer of the arbutin PTS, which is only found in the two strains of clade 1 is also most likely explained by horizontal gene transfer, especially since the best homologues of these genes are found in *Lactobacillus farciminis*.

A recent comparative genomics study reported that *L. rhamnosus* GG (clade 4) has the highest number of predicted transport proteins among the *L. casei* group representatives [31]. However, the current analysis revealed that strains of clades 3 and 4 encode even a larger number of transport genes, making these strains of *L. rhamnosus* the richest transport encoding strains of this taxonomic group (Supplementary Table 6).

The large variety of carbohydrates *L. rhamnosus* strains are adapted to grow on is apparent in the percentage of genes of the pan-genome (23%) and especially the variome (34%) assigned to functions related to carbohydrate digestion, transport and metabolism (Supplementary Table 7). A more detailed view at the carbohydrate utilization patterns of *L. rhamnosus* strains *in vitro* is presented in [37]. In the framework of glycan degradation, all strains were predicted to be able to release glucosyl, galactosyl and mannosyl reducing-end residues of saccharide polymers, as well as cleave 1-6 linked fucose and 1-3 and 2-6 linked N-Acetyl-Glucosamine

(GlucNAc) residues from (host associated cell-surface) glycosides. Strains of clades 2 and 6 have a reduced ability to degrade glycans compared to the other strains and, for example, are predicted to be unable to release protein O-linked glycans.

The differential presence of extracellular glycosidases, transport systems and metabolic pathways for the release and utilization of L-fucose, galactose, N-acetyl galactoseamine, which are particularly present in clades 4, 5, and 6, supports the idea that some *L. rhamnosus* strains could be adapted to grow on either human milk oligosaccharides (HMO) or mucus [43]. Notably, *L. rhamnosus* GG (clade 4), was shown to adhere to mucus *in vitro* [24] and appears to encode the capacity to use mucus-associated saccharides as carbon and energy source for growth [44]. Strains of clade 8, which is also dominated by human derived isolates, lack these extracellular glycosidases, but share the capacity to produce other galactosidases and phospho-glucosidases, which may also reflect their intestinal adaptation.

Fermentative capacity, pyruvate dissipation

Lactobacillus rhamnosus is a facultative heterofermentative organism whose main carbohydrate fermentation pathway is glycolysis through the Embden–Meyerhof–Parnas (EMP) pathway [52]. A partial pentose phosphate pathway also appears to be present (missing enzymes: gluconolactonase, 2-keto-3-deoxygluconate aldolase, 3-hexulose-6-phosphate synthase) and is likely to support biosynthetic routes as well as pentose utilization.

L. rhamnosus grows preferentially on glucose that is degraded via the EMP pathway leading to pyruvate, which is then transformed into D- and L-lactate by the corresponding lactate dehydrogenases, encoded by *ldhL* and *ldhD* genes that are part of the core genome. These functions appear redundant in the *L. rhamnosus* core genome, with three copies of *ldhD* and seven copies of *ldhL*, supporting the critical role of this function in the overall metabolism of this bacterium. Nevertheless, *L. rhamnosus* strains cultivated in cheese-like environments displayed relatively low lactic acid production and produced significant amounts of acetic acid, illustrative of the heterofermentative character of the species [64]. Besides D- and L-lactic acid and acetic acid pathways, additional pyruvate-dissipation pathways are encoded by the genome leading to formic acid, acetaldehyde, ethanol, oxaloacetate, and acetoin (Supplementary Table 7). Production of all these compounds is industrially relevant in the preservation of food raw-materials [45], and cheese ripening [46], but also for substrate production in bio-plastic manufacturing [47].

All *L. rhamnosus* strains contain only a partial TCA cycle [48] and harbor genes

predicted to encode the pathway required for succinate to citrate conversion. Nevertheless, the utilization of molecular oxygen as a terminal electron acceptor appears possible via the production of menaquinones, or vitamin K, which is a capacity that appears unique for *L. rhamnosus* among the *Lactobacillus* genus [49].

Nitrogen metabolism

The niches from which the *L. rhamnosus* strains used in this study were isolated encompass both protein-rich (e.g. dairy) and protein-poor environments (e.g. soil and plants) [29] (Supplementary Table 1). The *L. rhamnosus* core genome encodes import functions (dominated by the ABC-family transporters) for several amino acids, for which the biosynthesis pathway is not encoded by the genome, including alanine, isoleucine, leucine, valine, phenylalanine, threonine, and tyrosine. Notably, within the variome an ABC transporter is predicted to import taurine.

The main source of amino acids and peptides produced by LAB that live in milk, are derived from caseins, the most abundant protein in milk. Hydrolysis of caseins by LAB is initiated by cell-envelope proteinases (CEP) that degrade the protein into oligopeptides. The *L. rhamnosus* core genome encodes two CEP, PrtP and PrtR, that are responsible for hydrolysis of caseins. Other core protein degradation enzymes include protease PepS16, aminopeptidases PepA and MAP, endopeptidases PepS, PepS24, and PepM16 and proline-specific peptidases PIP and PepQ.

While the genomic content of amino acid and peptide transport differs considerably between strains, all strains seem to be equipped with a similar cytoplasmic peptidase repertoires, of which only four out of the 36 predicted peptidases vary among strains (Supplementary Table 4). These encompass 3 metallopeptidases and a serine peptidase. The intracellular metallopeptidase encoded by the *abgB* gene is present in all strains of clusters 2 and 4 (enriched for human isolates) and might provide an advantage to these strains in the intestinal niche since this function is predicted to support aminobenzoyl-glutamate utilization for folic acid recycling [50]. For the additional metallopeptidases (DppA and ZmpB) no information is available about their specific function, although the observation that ZmpB is secreted may imply that it exerts its function in the bacterial cell envelope.

Complementary to this extensive protein degradation machinery, all *L. rhamnosus* genomes encode complete pathways for the biosynthesis of 12 amino acids: arginine, asparagine, aspartic acid, cysteine, glutamine, glutamic acid, glycine, histidine, lysine, methionine, proline, serine, and tryptophan. However, the operon encoding serine acetyltransferase, cystathionine gamma-lyase and cystathionine beta-synthase that is

involved in the transformation of serine to cysteine, appears to be absent from strains in clades 4 and 7. Intriguingly, the sequence of this operon in the other strains displays high-level identity with similar operons encountered in *Lactobacillus brevis/ L. buchneri*, suggesting it has been acquired horizontal gene transfer.

Regulation and environmental interaction

Regulatory functions

Bacteria have developed sophisticated mechanisms for the regulation of both catabolic and anabolic pathways. Generally, bacteria do not synthesize degradative (catabolic) enzymes unless the substrates for these enzymes are present in their environment [51]. Proteins controlling cellular transcription (regulatory proteins) constitute 6% of *L. rhamnosus* pangenome proteins. Among them, 123 OGs are annotated as core regulator genes and 84 OGs display variation among the strains. Two sigma factor encoding genes are core functions in *L. rhamnosus*, including the housekeeping sigma factor 70 encoded by *rpoD*, and a sigma 24-like factor for which the target genes remain to be determined, but could include stress response as has been observed in *E. coli* [52]. Strains belonging to the same clade share a variable number of regulator OGs with different predicted functions. For example, 28 regulator OGs of clade 1 include the predicted regulators of arabinose, rhamnose, fucose, fructose, and tagatose transport as well as the LexA protease, and the MazF and YafQ toxin-antitoxin systems, whereas 32 regulator OGs for clade 2 are predicted to control transport of ribose, galactitol, glycerol, and oligopeptides, as well as a Zn dependent protease, and the 20 regulator OGs for clade 5 that control transport of mannose, tagatose, galactitol, cellobiose, mannitol, sorbitol, sorbose, pentitol, xylulose, and arabinose, illustrating the regulatory variability among the different clades. Notably, although clades 7 and 8 encode 32 and 28 variome-associated regulator OGs, none of these appears to be specific for either of these clades (Figure 2 – Supplementary Table 3 summarizing clade specific predicted protein functions). A few regulators that could impact bacterial adaptation to the environment are shared among several strains. Notably, among these is a Prck including quorum-sensing system that was shown to respond to environmental homoserine lactones in *S. enterica* [53] and is also found in *L. casei* Zhang [54], which is present among most members of clades 1, 5, 6, 7 and 8 and is genetically linked to an EPS biosynthesis operon. In addition, in strains of clades 2, 3 and 4 a LysR family regulator is predicted to control the synthesis of rhamnose-containing polysaccharides [55], suggesting that environmental adaptation could include the adjustment of extracellular exposed glycan-polymers.

Transcription regulation in response to environmental stimuli frequently involves two-component regulatory systems (TCSs) in many bacteria. In *L. rhamnosus* all seven TCSs belong to the core genome and are predicted to control envelope stress signal transduction (BaeS-BaeR, and LiaS-LiaR), cell-wall metabolism and corresponding antibiotic resistance (VicK-VicR, CiaH-CiaR, and VraR-VraS), membrane lipid fluidity in response to low temperature growth (DesK-DesR), secreted protein production (AgrC-AgrA), and citrate/malate metabolism (DcuR-DcuS).

Antimicrobial production and immunity

Bacteriocins and bacterial immunity proteins

Many lactic acid bacteria produce small antimicrobial peptides or bacteriocins that can provide a competitive advantage by inhibiting or killing competing (close relative) bacterial species. Bacteriocins can display both narrow or broad spectrum inhibitory activities [56]. The *L. rhamnosus* genomes include one to three operons predicted to encode bacteriocin production (Supplementary Table 8), which is in agreement with the notion that various *L. rhamnosus* strains were shown produce varying antimicrobial activities [29, 57–59].

The *L. rhamnosus* core genome contains an operon predicted to encode three class II bacteriocin encoding genes that share similarity with bacteriocin genes identified in *L. casei* and *L. paracasei* [60]. The genetic locus also encodes bacteriocin production-associated functions like immunity genes, an ABC-transporter with an accessory protein as well as a three-component quorum-sensing regulatory module. The latter suggests that the production of these bacteriocins could be controlled in a quorum-sensing manner [61].

The *L. rhamnosus* bacteriocin encoding genes in the variome do not have a clear clade distribution. A set of four antimicrobial peptide encoding genes are distributed among the strains in an apparent random manner with at least one member of each clade containing one up to all four peptides. Another antimicrobial peptide resembles a Class II pediocin-like bacteriocin [62], which is only present in the beer isolate ATCC8530 and is closely related to the pediocin-like bacteriocin PA-1 identified in *L. pentosus* [63].

Surface proteins

The capacity to interact with environmental surfaces, including host cells, is an important factor in niche colonization, and persistence. Several molecules that play a role in *L. rhamnosus* cross-talk with host cells have been identified and are among the

core capacities of the species, including the p40/p75 muramidases [26], and lipoteichoic acid modification D-alanylation genes (*dlt* operon) [64] (Supplementary Table 9), which have been proposed to interaction with the host mucosa by binding to mucus, influencing epithelial and immune functions, respectively [65]. In the core genome of *L. rhamnosus* we identified 37 genes that encode sortase dependent surface proteins (characterized by the conserved LPxTG motif) (Supplementary Table 9), including 17 proteins with a domain composition that may be predictive for a role in host interaction, including the recognition of fibronectin- and collagen-binding domains as well as a bacterial Ig-like domain (Figure 3). For instance, the third-largest protein - LGG_02923 (OG1951) contains a signal peptide, an LPxTG anchor, and a domain with four leucine rich repeats that is not frequently found in bacteria, but was proven to play a role in infection by some pathogens [66]. In 21 genomes the copy number per genome is 1, and it varies from 2 to 5 in the others.

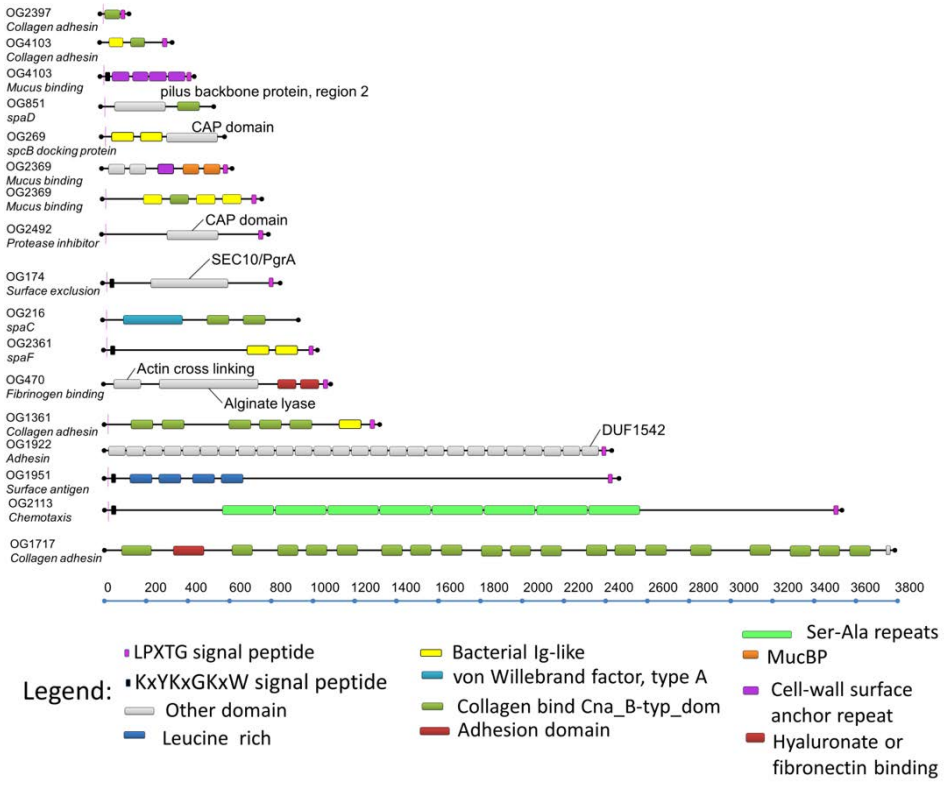


Figure 3. LPxTG proteins with host interaction potential. LPxTG proteins not included in this figure: sortases, hydrolases, lyases, proteases.

Proteins predicted to be exported or surface exposed within the *L. rhamnosus* variome do not appear to follow a clade-specific distribution (Supplementary Table 9), which may illustrate recent and variable gene acquisition or loss of these functions among the *L. rhamnosus* strains. Some of these proteins have been investigated *in vitro* for their possible role in the interaction with human epithelial or immune cells, including the adhesion and biofilm stimulating factor encoded by *mabA* [28], the secreted-docking protein pair encoded by the *spcABCD* [67], and the main surface protease and anti-inflammatory molecule PrtP [68] and the pilin encoding *spaCBA-srtC1* operon [43].

These proteins are all large, multi-domain proteins that contain the sortase dependent LPxTG anchoring motif, as well as several potential adhesion domains.

This observation prompted us to investigate other large multi-domain proteins that are predicted to be secreted or surface exposed. The largest among these genes is encoded by OG1717 (> 3500 amino acid residues), and is shared among 21 of the strains analyzed here, and entirely missing from clades 2, 3, 4 and 5. Notably, the strains that encode this gene appear to have a duplicated copy of this gene that encodes a protein with a signal sequence, one or more adhesion domains and 16 collagen binding domains (Figure 3).

The capacity to adhere to mucus, the extracellular matrix and/or intestinal epithelial cells are interesting properties with regard to probiotic features such as colonization of the gastrointestinal tract and interaction with the host. Among the variome surface exposed proteins, there is only one predicted mucus-binding protein: OG2369 (605aa). The protein contains a cell wall anchor repeat and two MucBP mucus-binding domains and has a close resemblance to *L. paracasei* surface proteins. It is only present in two strains: CNCM I-2493 originating from a fermented dairy product and Lrh24, isolated from animal (goat) feces.

Extracellular polysaccharides (EPS)

Bacterial extracellular polysaccharides (EPS) can play various roles in environmental interactions, including mammalian host interactions, for example by preventing recognition of the bacteria by the host immune system [69]. Six EPS biosynthesis gene clusters were identified in the *L. rhamnosus* pan-genome (Supplementary Table 8), and each strain contains two to four of these without an obvious clade-specific pattern being apparent. Only EPS cluster ‘5’ appears to belong to the core genome [55, 70], and its inactivation was shown to reduce the production of rhamnose-rich EPS [70]. The EPS cluster ‘6’, containing several predicted mannosyl-glycosyltransferases may be responsible for mannosyl-EPS production in *L. rhamnosus* which was previously

proposed on basis of ConA surface-probing of *L. rhamnosus* GG using atomic force microscopy [71].

The remaining EPS do not have clearly predicted specificities for the glycosyltransferases they encompass and clusters '1' and '2' were present in only a single strain Lrh24, and HN001, respectively.

Mobile genetic elements, and defense against foreign DNA

Among the best-described mobile genetic elements in bacteria are conjugative plasmids, IS elements, transposons and (pro-) phages [72]. These genetic elements play a prominent role in genetic mobility within a strain's genome (intracellular mobility), but also between bacterial genomes of different strains (intercellular mobility).

Mobile genetic elements (MGE) can rearrange genomes by the integration and/or deletion of genetic elements, one of the driving forces in the evolution of organisms [73].

Plasmids

Three plasmids of *L. rhamnosus* have previously been described, pLC001 (strain LC705), pLR001 and pLR002 (strain HN001) [38]. Next to the search for these known plasmids, novel plasmid identification relied on the detection of contigs that contain plasmid-associated genes, particularly focusing on replication functions.

The latter approach failed to identify novel plasmid entities in the genome database created in this study. Notably, strains belonging to clades 4 and 7 did consistently appear to lack plasmids, which was also the case for individual strains in the other clades (Lrh5 from clade 1, Lrh11 from clade 6, and ATCC7469, ATCC8530, Lrh23, Lrh12 and Lrh15 from clade 8; Figure 4). The distribution of the previously identified plasmids among the current strain collection appeared not to follow a clade-specific distribution.

Bacteriophages

Fourteen regions were identified that encode a total of 328 phage related OGs. Seven phage related regions that encompass more than 20 genes each, appear to be part of the core genome, and appear to be consistently inserted in the same chromosomal locus (Supplementary Table 8). Additional phage-related loci were variably present in the *L. rhamnosus* clades and strains (Figure 4) and resembled previously recognized phages of *Lactobacillus* origin (e.g., Lc-Nu [74], Lrm1 [75], A2 [76], Lb338-1 [77], phi adh [78], phi AT3 [79]). The gene composition of these phage associated loci is congruent with the canonical mosaic-like composition of (pro)phage genomes, which exemplifies

the high evolutionary rate of phage diversification [80].

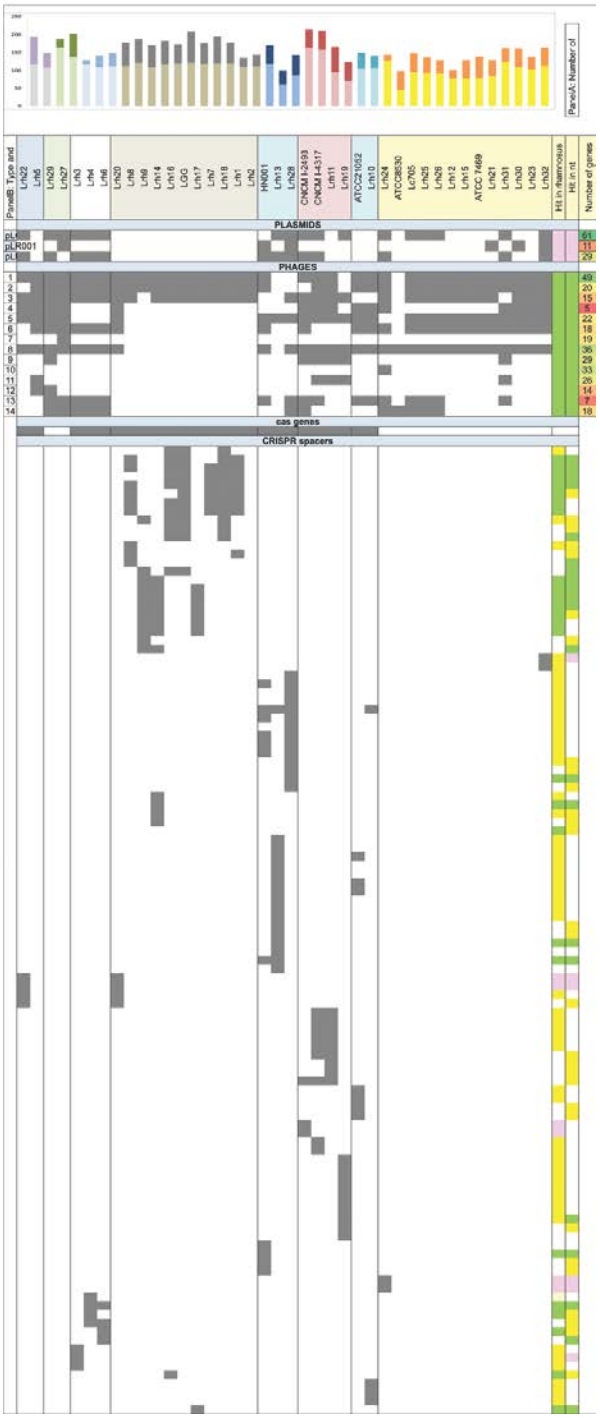


Figure 4. Summary of mobile elements present in *L. rhamnosus* strains. Panel 1: Number of genes annotated as mobile elements in the pan genome, including plasmids, phages, integrases, transposases (lighter colors) and number of genomic islands (dark colors). Panel 2: Type and distribution of mobile elements in each genome: plasmids, phages, cas genes and CRISPR spacers. Grey represents gene presence and white gene absence. For the spacers' analysis, only spacers that present a hit in any of the databases are represented. In the columns Hits in rhamnosus and Hits in nt, the colors represent the type of hit: yellow – unknown; pink – plasmid; green – phage genes, white – not found. Strains are organized by genetic clades separated by vertical lines.

Restriction modification

Restriction modification systems (RMS) are known to be variable among strains of a species [81] and play an important role in protection against foreign DNA, including phage attack. Therefore, RMS system presence and diversity were evaluated in the *L. rhamnosus* genomes (Supplementary Table 2), revealing that there are no core-genome associated RMS, and the RMS within the variome display a large inter-strain variation. Among the most common systems are the Mrr system [82] in strains of clades 1 and 8 (except Lrh31), the Type 1 RMS [83] shared between the majority of clades 2, 3, 7 and 8, and the Type II and Type III RMS [84] characteristic for clade 4 and shared by some strains of clade 5. In addition, some RMS are present in certain strains of which the closest relative lacks them. Notably, all strain-specific RMS appeared to be intact, which was also seen in other organisms like *H. pylori* [85], whereas the majority of the shared RMS (> 7 RMS) are predicted to be incomplete or inactivated by mutation(s). This could relate to positive selection for recently acquired strain-specific RMS, a process similar to *H. pylori* [85], where through constant acquisition of new RMS and the selective inactivation or removal of the older ones, the defense against foreign DNA, including phages, is constantly updated.

CRISPR-Cas system

Horizontal gene transfer (HGT) can be beneficial for bacteria by bringing in new functionalities as long as it does not disrupt fitness in the ecosystem. This balance is partly maintained by the CRISPR-Cas adaptive immunity system, which uses clustered regularly interspaced short palindromic repeats (CRISPR) and *cas* (CRISPR-associated) genes. The CRISPR loci are partially palindromic repeats separated by short stretches of spacer DNA that are acquired from invasive bacteriophage or plasmid DNA. These spacer sequences allow cells to recognize and cleave invasive DNA identical to the included sequence [86].

The CRISPR locus in the *L. rhamnosus* genomes is of type II-A (Lsal1 family) with a CRISPR repeat sequence of 5'-GTCTCAGGTAGATGTCAGATCAATCAGTTCAAG AGC -3'. Lsal1-type CRISPR loci were identified in 25 out of the 40 strains (Figure 4), and could play a prominent role in *L. rhamnosus* evolution by preventing intra-species HGT. Notably, the presence of the CRISPR-Cas system follows the clade distribution in several ways, which is clearly illustrated by the observation that all clade 2 and 8 strains lack a CRISPR-Cas system. Both the sequence of the *cas* genes of in *L. rhamnosus* and their genetic organization resembles that of *L. casei* [3], suggesting its inheritance from a common

ancestor. Notably, the presence of the CRISPR-Cas system does not seem to correlate with the number of genomic islands recognized in these different strains.

Noteworthy, groups of spacer sequences within the repeat locus are specific for each of the clades 3, 4, 5 and 6 (Figure 4, Supplementary Figure 2) and their predicted targets (on basis of 100% sequence identity to genes in the NCBI nr database) belong to *Lactobacillus* phages (Lc-Nu, A2, P2 and Lrm1) and plasmids (*Lactobacillus* casei str. Zhang plasmid plca36, *Lactobacillus rennini* plasmid pREN, *Lactobacillus plantarum* 16 plasmid Lp16C, plasmid pC30il, cryptic plasmid pLJ42, *Lactobacillus* helveticus H10 plasmid pH10 and *Lactobacillus salivarius* UCC118 plasmid pSF118-20) (Supplementary Table 2, Figure 4). Remarkably, numerous spacers show sequence identity to *Lactobacillus rhamnosus* phages (Lc-Nu, Pi2, A2) and plasmids (pLR002) present in their own genome and/or clade, and several spacer sequences are complementary to phages and hypothetical proteins of other *L. rhamnosus* strains and/or clades. For example, 14 different spacer sequences in clade 2 are identical to regions of prophage pi2 genes that are found in the same strains whereas 7 are identical to *L. rhamnosus* genes (some of them annotated as integrases) from strains in other clades. Overall, only 23 out of 112 of the spacer sequences targeted genetic entities from other species (the NCBI nr database) and were not found in the *L. rhamnosus* pan-genome, suggesting that the CRISPR system in *L. rhamnosus* serves as a protection against the entry of mobile genetic elements from the same species. Thereby, the *L. rhamnosus* CRISPR system could restrict the intra strain and intra clade genetic exchange events, which could prominently influence clade evolution by favoring the acquisition of genetic material from more distant species rather than promote the genomic convergence of the *L. rhamnosus* clades. This concept is also supported by the many apparent acquisitions of transporter encoding genes from more distant species, like *Lactobacillus farciminis*, *Enterococcus faecalis*, *Leuconostoc kimchii* etc. (Supplementary Table 2).

Core genome evolutionary drift and variome distribution

The core proteome phylogenetic tree of a species relies on single aminoacid polymorphisms (SAPs) of the conserved proteins, which are accumulated over time due to spontaneous mutations. Recent whole-genome sequencing has revealed that single nucleotide polymorphisms (SNPs) are the most prevalent form of genetic diversity among different strains of the same bacterial species [87]. The randomness of their occurrence implies that the number of sequence variations between strains of a species (SNPs and/or SAPs) can be interpreted as a measure of evolutionary time relative to a common ancestor. However, some sequence variations (e.g. SNPs and

SAPs) can play a role in niche adaptation and fitness and can therefore be selected by the environment and may predominate in specific, niche-adapted clades of a bacterial species. For example, the presence of certain SAPs in bacteria has been associated with changes in virulence [88], antibiotic resistance [89], metabolism [90] and persistence in the host intestine [91]. Patterns of SAPs within core proteome functions were used to determine the phylogenetic relationship of the *L. rhamnosus* strains (Figure 5A), by alignment of 1008 orthologous proteins that are encoded in a single copy in the core genome of the *L. rhamnosus* strains.

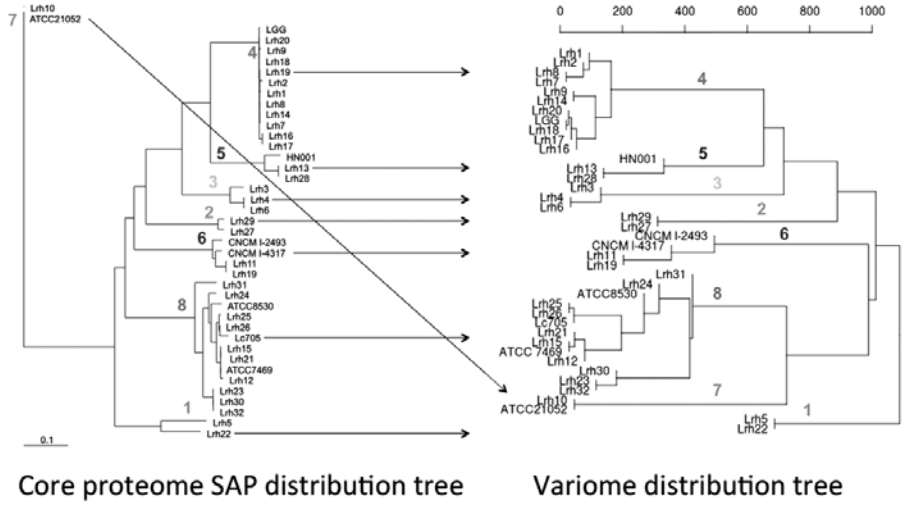


Figure 5. *L. rhamnosus* genome based phylogenetic relatedness, (Panel A) based on core proteome SAP distribution and (panel B) based on hierarchical clustering of variome gene distribution. The scale for the presence-absence tree represents the number of variable OGs the clustering is based on. Arrows connecting the same strains of both trees aims at highlighting the common groups of strains, which are also marked by the variome based clade numbering (panel B).

This analysis identified a total of 5,127 SAPs within the (single copy) core proteome of at least one of the *L. rhamnosus* strains and allowed the phylogenetic clustering of the strains on basis of evolutionary drift of the core proteome. This analysis revealed a substantial evolutionary distance between the strains in clade 7 (Lrh10 and ATCC21052; 5A) and the other 38 strains of *L. rhamnosus*, whereas strains of clades 4 and 8 that are all human isolates are quite closely related.

This phylogenetic grouping is in agreement with genomic BLAST analysis of public *L. rhamnosus* genomes performed by NCBI, which assigned strain ATCC21052

to a subgroup of the species with a 6% variation compared to the other genomes available. In contrast, this analysis appears to contradict a recent phylogenetic analysis of the *L. casei* group that included several *L. rhamnosus* strains [31], and concluded that strain ATCC21052 clustered closely together with strain ATCC 53103 (GG).

However, this latter study employed the sequences of 'only' 34 ribosomal proteins for the phylogenetic analysis and thus provides substantially lower resolution as our analysis that includes 1008 core proteome sequences. Notably, the Human Microbiome Project has taken the ATCC21052 strain as a reference for the *L. rhamnosus* species, but our findings imply that this strain is not the best representative of this species.

Based on the current analysis it would be recommendable to choose representative strains of the different clades as a reference gene-set for the *L. rhamnosus* species, to ensure the appropriate representation of its genetic diversity.

Remarkably, despite the different evolutionary processes that drive the phylogenetic discrimination of *L. rhamnosus* strains on basis of either the core proteome SAPs (random mutations without selective advantage) and the variome distribution (gene acquisition and loss) (Figure 1), the two phylogenetic trees present a highly congruent strain distribution. The variome based tree allowed the recognition of 8 clades among the 40 sequenced strains, and exactly the same grouping into the same 8 clades can also be recognized from the core-proteome based tree (Figure 5). The congruent evolution of core and variome relationships between strains has implications for the evolution of the *L. rhamnosus* species. The parallel evolution of variome and core diversification implies that clade determining gene acquisition and loss events have the same evolutionary age as the core-genome discriminating SAPs. This is in apparent contradiction with the concept that gene acquisition and loss are strong drivers of niche adaptation and commonly represent more recent genome diversification events. The possible role of the CRISPR-cas system described above may have played a prominent role in this restricted variome diversification, especially within the most congruent clades. Notably, this remarkable evolutionary path is distinct from the recognizable niche-associated evolution of the taxonomically related species *L. casei*. Similar analyses in this species allowed the recognition of dairy environment associated gene loss events that is typical for genome decay in the rich environment of milk [5], and which appeared to have occurred independent of the evolutionary drift of the core proteome.

Conclusions

The availability of forty genome sequences of *Lactobacillus rhamnosus* has enabled us to obtain a better understanding of the functional diversity and evolutionary

relatedness of this species, which encompasses many industrially relevant strains. The current analysis that focused on the variome associated genes and their distribution among clades of strains within the *L. rhamnosus* species, complements the previously presented core genome analysis [29].

The species *L. rhamnosus* is closely related to *L. casei* and *L. zeae*, and encompasses a genetically diverse group of strains, with a high frequency of discriminative polymorphisms in its core genome (SNPs and SAPs) and an impressive accessory genome or variome distribution. The comprehensive *in silico* examination of the variome associated functions and their distribution in terms of metabolic and regulatory diversity further illustrates the evolutionary diversity of this versatile *Lactobacillus* species that has evolved to grow and persist in a variety of ecological niches, including the intestinal tract of humans and animals. Notably, most of the strains appear to encode a extensive protein and sugar transport and catabolism capacity, which is congruent with their environmental versatility. Nevertheless, several genes and operons seem to have been acquired by HGT in some strains or clades that encode carbohydrate transport and catabolism functions, but also several other industrially important phenotypic traits, like polysaccharide biosynthesis (EPS), bacteriocin production, restriction–modification systems and/or bacterial defense systems (CRISP-Cas). In addition, other aspects of genomic diversity that may reflect niche adaptations can also be recognized, including the diversity of extracellular functions putatively involved in host interactions by cell adhesion or modulation of the host immune system.

The application of comparative genome sequencing to determine core genome and variome functions provided us with an unprecedented view of genome dynamics and adaptive evolution of the *L. rhamnosus* species. The recognition of the highly congruent phylogenetic relatedness of the core proteome and variome distribution patterns among the different strains included in this study indicates that HGT played an important role in the species' evolution. The CRISPR-Cas system may have contributed to this evolutionary path, by limiting intra-species mobility of genetic elements that could counteract evolutionary convergence of the 8 clades recognized by the variome distribution patterns. Finally, complementing these *in silico* analyses with phenotypic profiling of the strains of this species can expand our understanding of gene-function relationships and targeted genetic engineering strategies can subsequently establish the role of specific genes and functions in the adaptation to particular niches, where the molecular mechanisms underlying the cross-talk of this bacterium with the host intestine mucosa is of particular interest in the context of its

use as health-promoting diet ingredient, i.e. a probiotic.

Materials and methods

Strain selection, genome sequencing and annotation

36 different *L. rhamnosus* strains from the Helsinki University and the Danone Nutricia Research Collection were selected based on origin information or AFLP classification. The strain selection aimed to cover as much as possible the diversity inherent to the species. The type strain ATCC7649 is part of the group of newly sequenced strains. A complete list of the selected *L. rhamnosus* strains and their origin can be found in Supplementary Table 1. Strains with a CNCM code have been deposited in the Institute Pasteur CNCM public library. In addition, four public strains genomes (strains GG, HN001, Lc705, ATCC 21052) were downloaded and processed with the same bioinformatics pipeline as the newly sequenced genomes.

For DNA preparation, 2 ml of overnight culture was pelleted, washed and resuspended in 20mM Tris-HCl, pH 8.0, 1mM EDTA, pH 8.0, 8% sucrose, 50mM sodium chloride (TES) buffer. Cell lysis was performed with lysozyme (360 mg/ml) and mutanolysin (140 U /ml) during 2 h at 37°C, then 300 ul water was added and 80 ul of 20% sodium dodecyl sulfate (SDS) solution. DNA was extracted using phenol/chloroform (1:1) (3x). The DNA was precipitated with isopropanol and washed with 70% ethanol. RNAse treatment was performed using 100 ug/ml RNAse (Sigma) during 1 hour at 37°C.

Draft genome sequences of 36 *L. rhamnosus* strains were obtained (GATC Biotech, Germany) using 454 GS FLX sequencing at a sequence coverage ranging from 9-22x (see complete sequencing statistics in Supplementary Table 1). Genomes assembly was performed using Newbler 2.6 and 2.8 with standard settings. Contig sequences of all strains were annotated using the RAST pipeline [92]. Genome sequences of *L. rhamnosus* strains taken from public databases were re-analyzed (gene assignments and annotation) using the same procedures to ensure consistency of the gene-function predictions for each genome used.

Orthology estimation and genome comparisons

The orthology prediction was performed for all genomes (5 public and 35 sequenced by 454 GS FLX in the present study) using sequence clustering and search program Usearch v. 6.0.307 [93] . To define orthologous groups (OGs), all the predicted open reading frame (ORF) amino acid sequences of all the strains were first ordered by length using the 'sortbylength' command, and then clustered with the

'cluster_smallmem' command using identity threshold of 0.5. Thus, the members of each resulting ORF sequence cluster shared at least 50% amino acid identity over their entire sequence lengths based on the longest sequence. The longest sequence in each cluster was chosen as the representative of the orthologous group. The pangenome was thus defined as the set of representative sequences of the orthologous groups. The ortholog data was managed using R v. 2.15.3 with Biostrings package and in-house scripts.

Since most of the genomes are in a draft state, ORF prediction might have missed protein-coding sequences (CDSs). To detect the eventually missing genes, CDSs were aligned against each of the genomes using tblastn. CDSs that had a hit with 95% identity and query length or had the best alignment on the edge of a contig were classified as being present in the corresponding genome.

Search for *L. rhamnosus*-specific genes (orphan OGs)

To identify *L. rhamnosus*-specific genes we created a database of the protein sequences of all completely sequenced bacterial genomes present in the NCBI nr database. All known *Lactobacillus rhamnosus* genomes were excluded from this set. The final database comprised of 3,505,217 proteins from 1047 genomes. We compared all the genes belonging to *Lactobacillus rhamnosus* strains to this database using BlastP with an initial E-value cut-off of 1e−30, the same value used to create the OG matrix. Genes that had no blast hit against any of the proteins in the database were considered to be *L. rhamnosus* specific. The genes specific for the species' core and pan gene sets were analyzed using the same approach.

Variome classification based on presence-absence of genes

Genomes and OGs were clustered in R [94] using the complete method based on the Manhattan distance of presence - absence matrix. A subset of 10% of core genes was added to the variable genome in order to identify the least variable OG clusters. The output was visualized with heatmap.2 from the gplots package [95].

BRIGS mapping on the strain GG genome

To identify the variable regions, all genomes were aligned against the strain GG genome and visualized using BRIG [96]. Rings were color coded according to the variome classification of the strains. ABC and PTS transporters were visualized by aligning the genes against the genome and taking the best hit. Contigs mapping to multiple regions were all shown in the ring. Repetitive regions in the ring overlap on the reference and therefore do not display the repeat.

CRISPR identification and characterization

CRISPR loci and associated *cas* genes were identified using a combination of homology to previously identified *cas* genes and their corresponding CRISPR repeats [86] and by *de novo* identification using the CRISPR Recognition Tool (CRT) [97]. Spacer homologies to foreign genetic elements were assessed using BLASTN [98] on two databases: one created with all *L. rhamnosus* genomes in this study and the NCBI complete nucleotide collection nr [99]. Nucleotide conservation between CRISPR spacers and corresponding proto-spacers in phages, plasmids and chromosomal sequences were visualized using Ugene [100].

Phylogenetic tree

We define the phylogenetic relationship of the various strains from the patterns of single-amino acid substitutions of the proteins that belong to the core genome of the species and were present in a single copy in all strains (Supplementary Table.2). The protein sequences of 1,008 OGs with a single member in each *L. rhamnosus* genome were aligned using MUSCLE [101], which delivered 5,127 positions in the core protein sequences with altered amino acid content. These alignments were concatenated after which a maximum-likelihood tree was constructed using PHYML [102].

Metabolic mapping – KEGG

The draft metabolic network and associated gene–reaction relationships of the pan, core and clades' genomes was constructed using a KEGG automatic annotation server (KAAS) [103], according to the KEGG database [104] using the standard settings and mapping against 40 typical prokaryotes genomes. Then, manual checking of the metabolic maps obtained from the KAAS annotation identified the relevant metabolic diversity and gaps.

When analyzing the *L. rhamnosus* genes encoding transport functions, the substrate specificity of PTSs was predicted based on homology to annotated PTS genes (the best hit in the NCBI nr database) and based on their genomic context, focusing on genetically linked genes with predicted functions related to enzyme and regulatory functions in carbohydrate metabolism.

Domain annotation

Identification of proteins containing LPXTG signals (for sortase dependent genes), mucus-, fibrinogen- or fibrin- binding, collagen- and Ig-like- domains was performed using Interproscan [105] analysis for all pan genome proteins.

Bacteriocin detection

BAGEL3 is an automated pipeline for identifying genes encoding class I and II bacteriocins using an identification approach that combines direct mining for the encoding gene and indirect mining of genetic context for accessory functions related to the production of these bacteriocins and related peptides [106].

References:

1. Ceapa C. et al. : *Influence of fermented milk products, prebiotics and probiotics on microbiota composition and health.* Best Pract Res Clin Gastroenterol 2013, 27:139–55.

2. Vasiljevic T, Shah NP: *Probiotics-From Metchnikoff to bioactives*. Int Dairy J 2008:714–728.

3. Smokvina T. et al. : *Lactobacillus paracasei comparative genomics: towards species pan-genome definition and exploitation of diversity*. PLoS One 2013, 8:e68731.

4. Lukjancenko O, Ussery DW, Wassenaar TM: *Comparative genomics of Bifidobacterium, Lactobacillus and related probiotic genera.* Microb Ecol 2012, 63:651–73.

5. Broadbent JR. et al. : *Analysis of the Lactobacillus casei supragenome and its influence in species evolution and lifestyle adaptation*. BMC Genomics 2012, 13:533.

6. Mahony J. et al. : *Phages of lactic acid bacteria: The role of genetics in understanding phage-host interactions and their co-evolutionary processes*. Virology 2012:143–150.

7. Bergey DH: *Bergey's Manual of Systematic Bacteriology - Vol 3: The Firmicutes*. 2009.

8. Heilig HGHJ. et al. : *Molecular diversity of Lactobacillus spp. and other lactic acid bacteria in the human intestine as determined by specific amplification of 16S ribosomal DNA.* Appl Environ Microbiol 2002, 68:114–123.

9. Hummelen R. et al. : *Deep sequencing of the vaginal microbiota of women with HIV*. PLoS One 2010, 5.

10. Smit G, Smit BA, Engels WJM: *Flavour formation by lactic acid bacteria and biochemical flavour profiling of cheese products.* FEMS Microbiol Rev 2005:591–610.

11. Yang Z. et al. : *Capsular and slime-polysaccharide production by Lactobacillus rhamnosus JAAS8 isolated from Chinese sauerkraut: Potential application in fermented milk products.* J Biosci Bioeng 2010, 110:53–57.

12. Lee H. et al. : *Functional properties of Lactobacillus strains isolated from kimchi.* Int J Food Microbiol 2011, 145:155–161.

13. Klaenhammer TR. et al. : *Functional genomics of probiotic Lactobacilli*. J Clin Gastroenterol 2008, 42 Suppl 3:S160–S162.

14. Wind RD. et al. : *Tolerance and safety of the potentially probiotic strain Lactobacillus rhamnosus PRSF L477 : a randomised , double blind placebo controlled trial in healthy volunteers Tolerance and safety of the potentially probiotic strain Lactobacillus rhamnosus*. Br J Nutr 2010, 104:1806–16.

15. Toh ZQ. et al. : *Probiotic therapy as a novel approach for allergic disease*. Front Pharmacol 2012, 3 SEP.

16. Szajewska H. et al. : *Meta-analysis: Lactobacillus GG for treating acute diarrhoea in children.* Aliment Pharmacol Ther 2007, 25:871–881.

17. Liu S, Hu P, Du X, Zhou T, Pei X: *Lactobacillus rhamnosus GG supplementation for preventing respiratory infections in children: a meta-analysis of randomized, placebo-controlled*

trials. Indian Pediatr 2013:377–381.

18. Wickens K. et al. : *A protective effect of Lactobacillus rhamnosus HN001 against eczema in the first 2 years of life persists to age 4 years.* Clin Exp Allergy 2012, 42:1071–1079.

19. Hummelen R. et al. : *Lactobacillus rhamnosus GR-1 and L. reuteri RC-14 to prevent or cure bacterial vaginosis among women with HIV*. Int J Gynecol Obstet 2010, 111:245–248.

20. Perez-Burgos A. et al. : *Psychoactive bacteria Lactobacillus rhamnosus (JB-1) elicits rapid frequency facilitation in vagal afferents.* Am J Physiol Gastrointest liver Physiol 2013, 304:G211–20.

21. McLean PG. et al. : *Targeting the microbiota-gut-brain axis to modulate behavior: Which bacterial strain will translate best to humans?* Proc Natl Acad Sci 2012:E174–E174.

22. Douillard FP. et al.: *Comparative genomic and functional analysis of Lactobacillus casei and Lactobacillus rhamnosus strains marketed as probiotics*. Appl Environ Microbiol 2013, 79:1923–33.

23. Van Bergenhenegouwen J. et al. : *Extracellular vesicles modulate host-microbe responses by altering TLR2 activity and phagocytosis.* PLoS One 2014, 9:e89121.

24. Kankainen M. et al. : *Comparative genomic analysis of Lactobacillus rhamnosus GG reveals pili containing a human- mucus binding protein.* Proc Natl Acad Sci U S A 2009, 106:17193–17198.

25. Von Ossowski I. et al. : *Functional characterization of a mucus-specific LPXTG surface adhesin from probiotic Lactobacillus rhamnosus GG.* Appl Environ Microbiol 2011, 77:4465–4472.

26. Bäuerl C. et al. : *Functional analysis of the p40 and p75 proteins from Lactobacillus casei BL23*. J Mol Microbiol Biotechnol 2010, 19:231–241.

27. Nermes M. et al. : *Interaction of orally administered Lactobacillus rhamnosus GG with skin and gut microbiota and humoral immunity in infants with atopic dermatitis*. Clin Exp Allergy 2011, 41:370–377.

28. Vélez MP. et al. : *Characterization of MabA, a modulator of Lactobacillus rhamnosus GG adhesion and biofilm formation*. FEMS Immunol Med Microbiol 2010, 59:386–398.

29. Douillard FP.et al. : *Comparative genomic and functional analysis of 100 Lactobacillus rhamnosus strains and their comparison with strain GG*. PLoS Genet 2013, 9:e1003683.

30. Canchaya C. et al. : *Diversity of the genus Lactobacillus revealed by comparative genomics of five species*. Microbiology 2006, 152:3185–3196.

31. Toh H. et al. : *Genomic adaptation of the Lactobacillus casei group*. PLoS One 2013, 8:e75073.

32. Kleerebezem M. et al. : *Complete genome sequence of Lactobacillus plantarum WCFS1*. Proc Natl Acad Sci U S A 2003, 100:1990–1995.

33. Siezen RJ. et al. : *Complete resequencing and reannotation of the Lactobacillus plantarum WCFS1 genome*. J Bacteriol 2012:195–196.

34. Kant R, Blom J, Palva A, Siezen RJ, de Vos WM: Comparative genomics of Lactobacillus. Microb Biotechnol 2011, 4:323–32.

35. Borneman AR, McCarthy JM, Chambers PJ, Bartowsky EJ: *Comparative analysis of the Oenococcus oeni pan genome reveals genetic diversity in industrially-relevant pathways.* BMC Genomics 2012, 13:373.

36. Hüfner E, Britton RA, Roos S, Jonsson H, Hertel C: *Global transcriptional response of Lactobacillus reuteri to the sourdough environment.* Syst Appl Microbiol 2008, 31:323–338.

37. Ceapa C. et al. : *Phenotype profiling links genotype to carbohydrate utilization signatures in the species Lactobacillus rhamnosus.* Appl Environ Microbiol 2015:AEM.00851–15–.

38. Morita H. et al. : *Complete genome sequence of the probiotic Lactobacillus rhamnosus ATCC 53103*. J Bacteriol 2009, 191:7630–7631.

39. Siezen RRJ. et al. : *Phenotypic and genomic diversity of Lactobacillus plantarum strains isolated from various environmental niches.* Environ Microbiol 2010, 12:758–773.

40. Levin BR, Cornejo OE: *The population and evolutionary dynamics of homologous gene recombination in bacteria.* PLoS Genet 2009, 5.

41. Richards VP. et al. : *Comparative genomics and the role of lateral gene transfer in the evolution of bovine adapted Streptococcus agalactiae*. Infect Genet Evol 2011, 11:1263–75.

42. Barraclough T, Balbi K, Ellis R: *Evolving concepts of bacterial species*. Evol Biol 2012:148–157.

43. Reunanen J. et al. : *Characterization of the SpaCBA pilus fibers in the probiotic Lactobacillus rhamnosus GG.* Appl Environ Microbiol 2012:2337–2344.

44. Sánchez B. et al. : *Adhesive properties, extracellular protein production, and metabolism in the Lactobacillus rhamnosus GG strain when grown in the presence of mucin*. J Microbiol Biotechnol 2010, 20:978–984.

45. Bove CG. et al. : *Metabolic and proteomic adaptation of Lactobacillus rhamnosus strains during growth under cheese-like environmental conditions compared to de Man, Rogosa, and Sharpe medium*. Proteomics 2012, 12:3206–3218.

46. Bove CG. et al. : *Evaluation of genetic polymorphism among Lactobacillus rhamnosus non-starter Parmigiano Reggiano cheese strains.* Int J Food Microbiol 2011, 144:569–572.

47. Flieger M. et al. : *Biodegradable plastics from renewable sources.* Folia Microbiol (Praha) 2003, 48:27–44.

48. Golbach JL. et al. : *Growth response of lactobacillus rhamnosus to microaerophilic incubation and addition of reductants in a riboflavin microtiter plate assay.* J Rapid Methods Autom Microbiol 2007, 15:320–331.

49. Hess A, Holländer R, Mannheim W: *Lipoquinones of some spore-forming rods, lactic-acid bacteria and actinomycetes*. J Gen Microbiol 1979, 115:247–252.

50. De Crécy-Lagard V. et al. : *Comparative genomics of bacterial and plant folate synthesis and salvage: predictions and validations.* BMC Genomics 2007, 8:245.

51. Deutscher J: *The mechanisms of carbon catabolite repression in bacteria.* Curr Opin Microbiol 2008:87–93.

52. Haldenwang WG: *The sigma factors of Bacillus subtilis.* Microbiol Rev 1995, 59:1–30.

53. Michael B, Smith JN, Swift S, Heffron F, Ahmer BMM: *SdiA of Salmonella enterica is a LuxR homolog that detects mixed microbial communities*. J Bacteriol 2001, 183:5733–5742.

54. Zhang W. et al. : *Complete genome sequence of Lactobacillus casei Zhang, a new probiotic strain isolated from traditional homemade koumiss in inner Mongolia, China*. J Bacteriol 2010, 192:5268–5269.

55. Péant B. et al. : *Comparative analysis of the exopolysaccharide biosynthesis gene clusters from four strains of Lactobacillus rhamnosus.* Microbiology 2005, 151(Pt 6):1839–51.

56. Riley MA, Wertz JE: *Bacteriocin diversity: Ecological and evolutionary perspectives.* Biochimie 2002:357–364.

57. De Keersmaecker SCJ. et al. : *Strong antimicrobial activity of Lactobacillus rhamnosus GG against Salmonella typhimurium is due to accumulation of lactic acid.* FEMS Microbiol Lett 2006, 259:89–96.

58. Vong L, Lorentz RJ, Assa A, Glogauer M, Sherman PM: *Probiotic Lactobacillus rhamnosus inhibits the formation of neutrophil extracellular traps*. J Immunol 2014, 192:1870–7.

59. Pithva S, Shekh S, Dave J, Vyas BRM: *Probiotic attributes of autochthonous Lactobacillus rhamnosus strains of human origin.* Appl Biochem Biotechnol 2014, 173:259–277.

60. Kuo YC.et al. : Characterization of putative class II bacteriocins identified from a non-bacteriocin-producing strain Lactobacillus casei ATCC 334. Appl Microbiol Biotechnol 2013, 97:237–246.

61. Kleerebezem M.et al. : *Quorum sensing by peptide pheromones and two-component signal-transduction systems in Gram-positive bacteria.* Mol Microbiol 1997, 24:895–904.

62. Motlagh AM. et al. : *Nucleotide and amino acid sequence of pap-gene (pediocin AcH production) in Pediococcus acidilactici* H. Lett Appl Microbiol 1992, 15:45–48.

63. Henderson JT. et al. : *Purification and primary structure of pediocin PA-1 produced by Pediococcus acidilactici PAC-1.0.* Arch Biochem Biophys 1992, 295:5–12.

64. Abi Khattar Z. et al. : *The dlt operon of Bacillus cereus is required for resistance to cationic antimicrobial peptides and for virulence in insects.* J Bacteriol 2009, 191:7063–7073.

65. Lee I-C, Tomita S, Kleerebezem M, Bron P a: *The quest for probiotic effector molecules--unraveling strain specificity at the molecular level*. Pharmacol Res 2013, 69:61–74.

66. Courtemanche N, Barrick D: *Folding thermodynamics and kinetics of the leucine-rich repeat domain of the virulence factor Internalin B.* Protein Sci 2008, 17:43–53.

67. Gagic D, Wen W, Collett M a, Rakonjac J: *Unique secreted-surface protein complex of Lactobacillus rhamnosus, identified by phage display*. Microbiologyopen 2013, 2:1–17.

68. Habimana O. et al. : *Positive role of cell wall anchored proteinase PrtP in adhesion of lactococci.* BMC Microbiol 2007, 7:36.

69. Fanning S. et al. : *Bifidobacterial surface-exopolysaccharide facilitates commensal-host interaction through immune modulation and pathogen protection*. Proc Natl Acad Sci 2012:2108–2113.

70. Lebeer S. et al. : *Identification of a gene cluster for the biosynthesis of a long, galactose-rich exopolysaccharide in Lactobacillus rhamnosus GG and functional analysis of the priming glycosyltransferase*. Appl Environ Microbiol 2009, 75:3554–3563.

71. Francius G. et al. : *Detection, localization, and conformational analysis of single polysaccharide molecules on live bacteria.* ACS Nano 2008, 2:1921–1929.

72. Frost LS, Leplae R, Summers AO, Toussaint A: Mobile genetic elements: the agents of open source evolution. Nat Rev Microbiol 2005, 3:722–732.

73. Liu M, Siezen RJ, Nauta A: *In silico prediction of horizontal gene transfer events in Lactobacillus bulgaricus and Streptococcus thermophilus reveals protocooperation in yogurt manufacturing.* Appl Environ Microbiol 2009, 75:4120–9.

74. Tuohimaa A, Riipinen KA, Brandt K, Alatossava T: *The genome of the virulent phage Lc-Nu of probiotic Lactobacillus rhamnosus, and comparative genomics with Lactobacillus casei phages.* Arch Virol 2006, 151:947–965.

75. Durmaz E. et al. : *Genome sequence and characteristics of Lrm1, a prophage from industrial Lactobacillus rhamnosus strain M1*. In Appl Environ Microbiol. Volume 74; 2008:4601–4609.

76. Alvarez MA, Rodríguez A, Suárez JE: *Stable expression of the Lactobacillus casei bacteriophage A2 repressor blocks phage propagation during milk fermentation.* J Appl Microbiol 1999, 86:812–816.

77. Alemayehu D. et al. : *Genome of a virulent bacteriophage Lb338-1 that lyses the probiotic Lactobacillus paracasei cheese strain.* Gene 2009, 448:29–39.

78. Altermann E, Klein JR, Henrich B: *Primary structure and features of the genome of the Lactobacillus gasseri temperate bacteriophage φadh*. Gene 1999, 236:333–346.

79. Lo TC, Shih TC, Lin CF, Chen HW, Lin TH: *Complete genomic sequence of the temperate bacteriophage AT3 isolated from Lactobacillus casei ATCC 393*. Virology 2005, 339:42–55.

80. Abedon ST: Phage evolution and ecology. Adv Appl Microbiol 2009:1–45.

81. Xu Q, Morgan RD, Roberts RJ, Blaser MJ: *Identification of type II restriction and modification systems in Helicobacter pylori reveals their substantial diversity among strains.* Proc Natl Acad Sci U S A 2000, 97:9671–9676.

82. Waite-Rees PA. et al. : *Characterization and expression of the Escherichia coli Mrr restriction system.* J Bacteriol 1991, 173:5207–19.

83. Loenen WAM, Dryden DTF, Raleigh EA, Wilson *GG: Type i restriction enzymes and their relatives.* Nucleic Acids Res 2014, 42:20–44.

84. Loenen WAM, Dryden DTF, Raleigh EA, Wilson GG, Murrayy NE: *Highlights of the DNA cutters: A short history of the restriction enzymes.* Nucleic Acids Res 2014, 42:3–19.

85. Lin LF, Posfai J, Roberts RJ, Kong H: *Comparative genomics of the restriction-modification systems in Helicobacter pylori.* Proc Natl Acad Sci U S A 2001, 98:2740–2745.

86. Brouns SJJ. et al. : *Small CRISPR RNAs guide antiviral defense in prokaryotes*. Science 2008, 321:960–964.

87. Harris SR. et al. : *Evolution of MRSA during hospital transmission and intercontinental spread.* Science 2010, 327:469–474.

88. Carroll RK. et al. : *Naturally occurring single amino acid replacements in a regulatory protein alter streptococcal gene expression and virulence in mice.* J Clin Invest 2011, 121:1956–1968.

89. Chewapreecha C. et al. : *Comprehensive identification of single nucleotide polymorphisms associated with beta-lactam resistance within pneumococcal mosaic genes.* PLoS Genet 2014, 10:e1004547.

90. Tolentino R, Chastain C, Burnell J: *Identification of the amino acid involved in the regulation of bacterial pyruvate, orthophosphate dikinase and phosphoenolpyruvate synthetase*. Adv Biol Chem 2013, 03:12–21.

91. Van Bokhorst-van de Veen H. et al. : *Genotypic adaptations associated with prolonged persistence of Lactobacillus plantarum in the murine digestive tract.* Biotechnol J 2013, 8:895–904.

92. Overbeek R. et al. *: The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST).* Nucleic Acids Res 2014, 42.

93. Edgar RC: *Search and clustering orders of magnitude faster than BLAST*. Bioinformatics 2010, 26:2460–2461.

94. R Development Core Team RFFSC: *R: A Language and Environment for Statistical Computing.* R Found Stat Comput 2008:2673.

95. Warnes GR: *Gplots: Various R programming tools for plotting data.* J Phycol 2012:569–575.

96. Alikhan N-F, Petty NK, Ben Zakour NL, Beatson S a: *BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons.* BMC Genomics 2011, 12:402.

97. Bland C. et al. : *CRISPR recognition tool (CRT): a tool for automatic detection of clustered regularly interspaced palindromic repeats.* BMC Bioinformatics 2007, 8:209.

98. Altschul SF. et al. : *Gapped BLAST and PSI-BLAST: A new generation of protein database search programs*. Nucleic Acids Res 1997:3389–3402.

99. Pruitt KD, Tatusova T, Maglott DR: *NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins*. Nucleic Acids Res 2007, 35:D61–D65.

100. Okonechnikov K. et al. : *Unipro UGENE: A unified bioinformatics toolkit.* Bioinformatics 2012:1166–1167.

101. Edgar RC: *MUSCLE: a multiple sequence alignment method with reduced time and space complexity*. BMC Bioinformatics 2004, 5:113.

102. Guindon S, Lethiec F, Duroux P, Gascuel O: *PHYML Online - A web server for fast maximum likelihood-based phylogenetic inference*. Nucleic Acids Res 2005, 33.

103. Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M: *KAAS: An automatic genome annotation and pathway reconstruction server*. Nucleic Acids Res 2007, 35.

104. Kanehisa M, Goto S, Furumichi M, Tanabe M, Hirakawa M: *KEGG for representation and analysis of molecular networks involving diseases and drugs.* Nucleic Acids Res 2009, 38.

105. Zdobnov EM, Apweiler R: I*nterProScan--an integration platform for the signature-recognition methods in InterPro*. Bioinformatics 2001, 17:847–848.

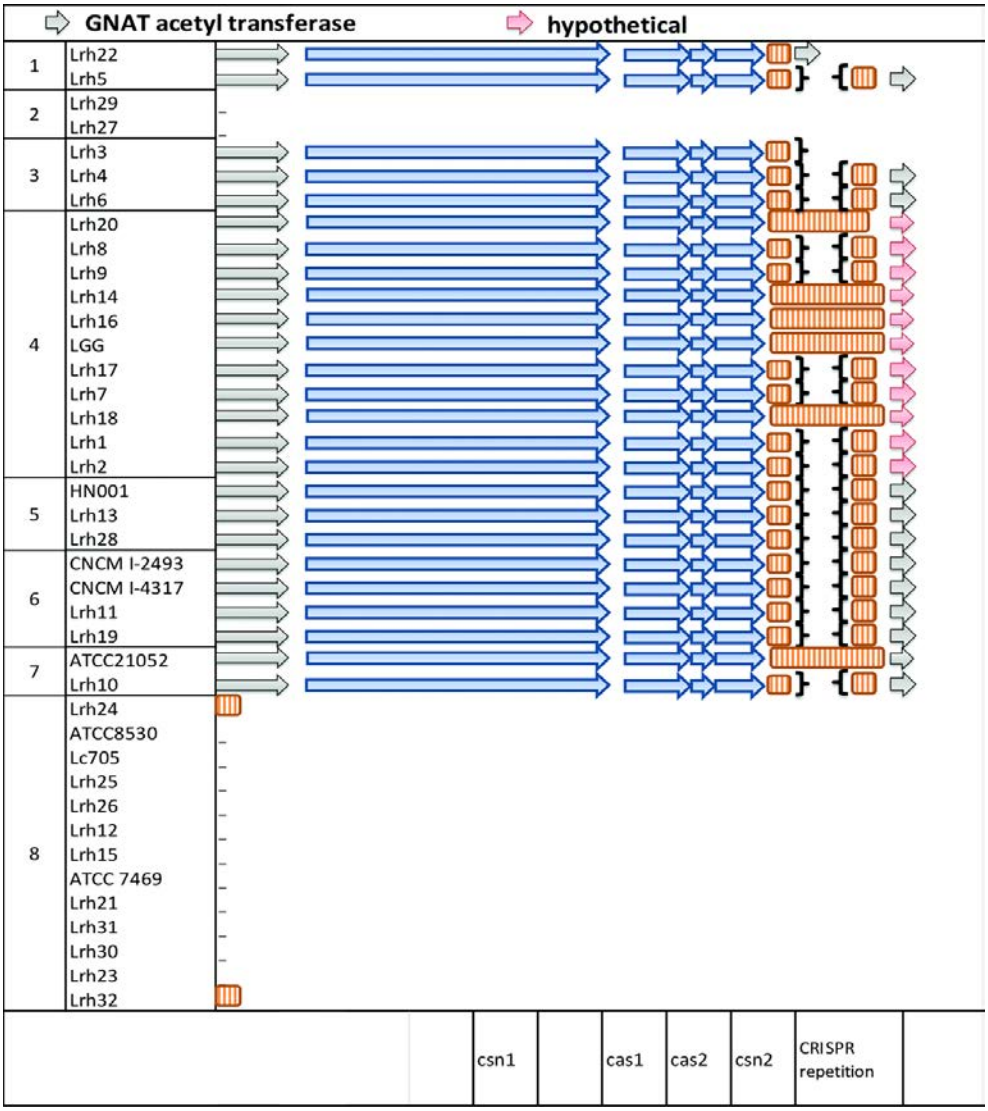
106. Van Heel AJ. et al. : *BAGEL3: Automated identification of genes encoding bacteriocins and (non-)bactericidal posttranslationally modified peptides.* Nucleic Acids Res 2013, 41.

Appendix 2. Supporting data:

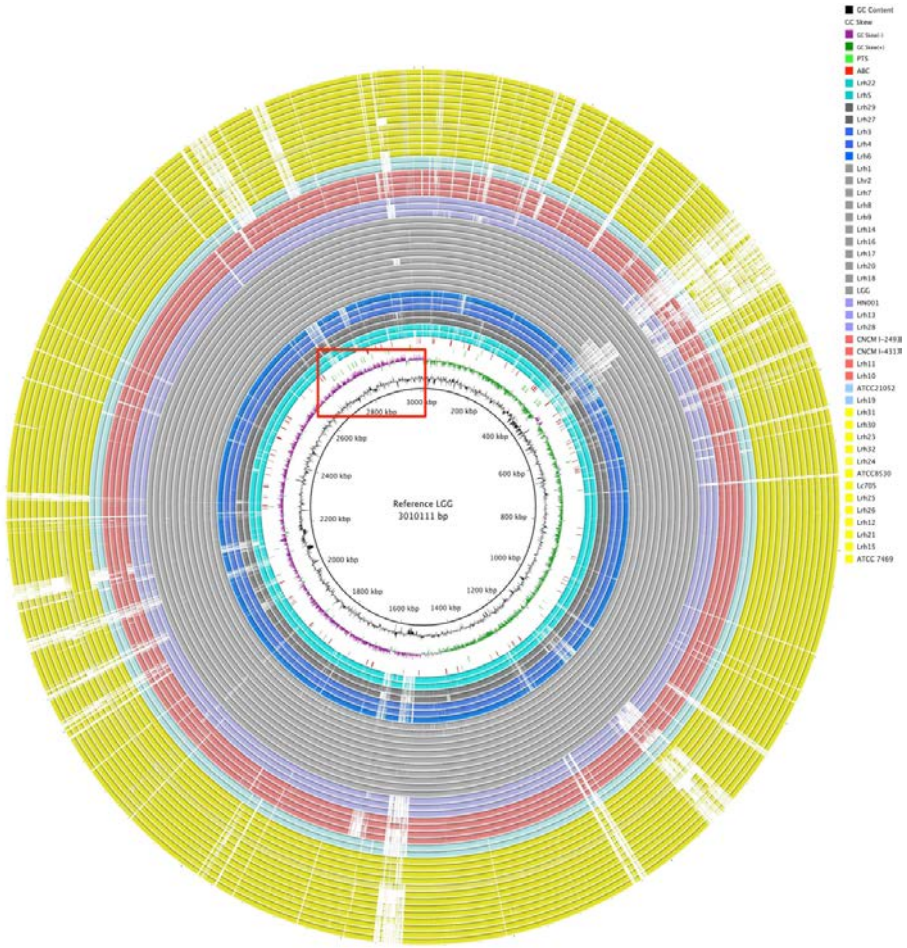
Supplementary Figure 1. Core/ pangenome OG saturation.

Strain	Length (Mb)	G+C (%)	ORFs	Isolated from	Contigs	Coverage	LPXTG genes	Reference	NCBI accesion number	Bioproject ID	Biosample ID
ATCC21052	2,87	46,7	2909	healthy intestine	1	20	19	Reference genome for the Human Microbiome Project			
ATCC7469	2,93	46,7	2934	clinical, type strain	81	17	21	this work			
ATCC8530	2,96	46,8	2933	beer	1	20	21	[35]			
CNCM I-2493	3,12	46,4	3163	fermented dairy product	261	13	21	this work	JTIX00000000	PRJNA253894	SAMN03196620
CNCM I-4317	3	46,5	3006	fermented dairy product	136	13	22	this work	JTIW00000000	PRJNA253894	SAMN03196621
HN001	2,91	46,6	2973	fermented dairy product	96	36	17				
Lc705	3,03	46,6	2881	fermented dairy product	1	25	23	[21]			
LGG	3,01	46,7	2926	healthy intestine	1	25	22	[21, 36]			
Lrh1	2,79	46,6	2779	human feces	535	9	19	this work	JTIV00000000	PRJNA253894	SAMN03196625
Lrh10	2,88	46,7	2898	fermented dairy product	91	15	19	this work	JTIU00000000	PRJNA253894	SAMN03196618
Lrh11	2,89	46,6	2923	blood	112	10	20	this work	JTIT00000000	PRJNA253894	SAMN03196637
Lrh12	2,89	46,6	2884	human feces	384	10	21	this work	JTIS00000000	PRJNA253894	SAMN03196628
Lrh13	2,9	46,7	2891	blood	209	9	22	this work	JTIR00000000	PRJNA253894	SAMN03196647
Lrh14	2,9	46,6	2883	blood	260	9	17	this work	JTIQ00000000	PRJNA253894	SAMN03196644
Lrh15	2,92	46,6	2923	blood	147	12	21	this work	JTIP00000000	PRJNA253894	SAMN03196633
Lrh16	2,92	46,6	2901	clinical	145	15	23	this work	JTIO00000000	PRJNA253894	SAMN03196622
Lrh17	2,92	46,6	2895	human feces	127	16	23	this work	JTIN00000000	PRJNA253894	SAMN03196627
Lrh18	2,93	46,6	2918	blood	222	9	22	this work	JTIM00000000	PRJNA253894	SAMN03196638
Lrh19	2,93	46,6	2933	healthy intestine	99	12	25	this work	JTIL00000000	PRJNA253894	SAMN03196650
Lrh2	2,8	46,6	2793	human feces	401	9	18	this work	JTIK00000000	PRJNA253894	SAMN03196626
Lrh20	2,93	46,6	2929	blood	102	16	21	this work	JTIJ00000000	PRJNA253894	SAMN03196645
Lrh21	2,95	46,6	2945	human feces	85	22	21	this work	JTII00000000	PRJNA253894	SAMN03196623
Lrh22	2,96	46,6	2988	blood	195	9	19	this work	JTIH00000000	PRJNA253894	SAMN03196648
Lrh23	2,96	46,6	2971	blood	61	15	21	this work	JTIG00000000	PRJNA253894	SAMN03196649
Lrh24	2,98	46,6	3008	animal (goat) feces	345	11	23	this work	JTIF00000000	PRJNA253894	SAMN03196631
Lrh25	2,98	46,6	2994	fermented dairy product	132	15	23	this work	JTIE00000000	PRJNA253894	SAMN03196619
Lrh26	2,99	46,6	3004	human feces	112	16	23	this work	JTID00000000	PRJNA253894	SAMN03196629
Lrh27	3	46,5	3061	human feces	240	14	20	this work	JTIC00000000	PRJNA253894	SAMN03196624
Lrh28	3	46,6	2982	blood	112	19	21	this work	JTIB00000000	PRJNA253894	SAMN03196643
Lrh29	3,01	46,5	3082	blood	437	11	20	this work	JTIA00000000	PRJNA253894	SAMN03196632
Lrh3	2,82	46,7	2856	fermented dairy product	362	10	17	this work	JTHZ00000000	PRJNA253894	SAMN03196636
Lrh30	3,01	46,6	3010	blood	96	12	21	this work	JTHY00000000	PRJNA253894	SAMN03196639
Lrh31	3,02	46,5	3061	vagina	124	10	25	this work	JTHX00000000	PRJNA253894	SAMN03196651
Lrh32	2,95	46,6	2980	blood	64	18	21	this work	JTHW00000000	PRJNA253894	SAMN03196642
Lrh4	2,83	46,8	2852	blood	287	11	17	this work	JTHV00000000	PRJNA253894	SAMN03196634
Lrh5	2,84	46,7	2845	human feces	161	12	18	this work	JTHU00000000	PRJNA253894	SAMN03196630
Lrh6	2,85	46,8	2876	blood	173	14	17	this work	JTHT00000000	PRJNA253894	SAMN03196635
Lrh7	2,86	46,7	2851	blood	86	17	17	this work	JTHS00000000	PRJNA253894	SAMN03196646
Lrh8	2,86	46,7	2875	blood	86	17	18	this work	JTHR00000000	PRJNA253894	SAMN03196640
Lrh9	2,88	46,7	2864	blood	112	13	18	this work	JTHQ00000000	PRJNA253894	SAMN03196641

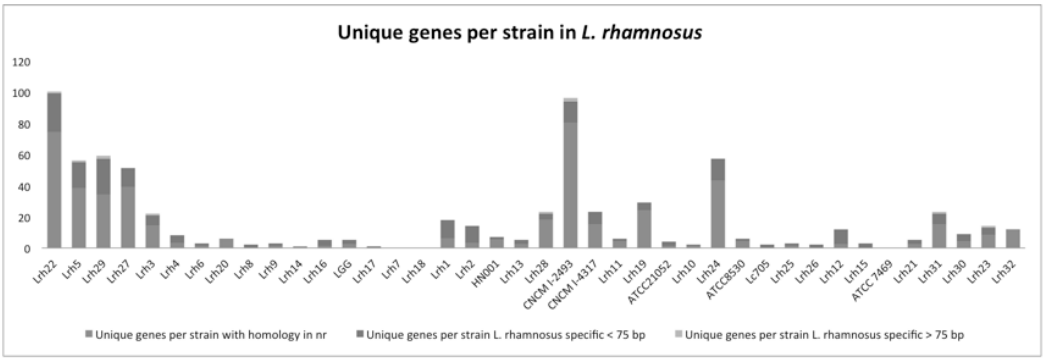
Supplementary Figure 2. Cas genes/ CRISPR architecture in *L. rhamnosus* strains.

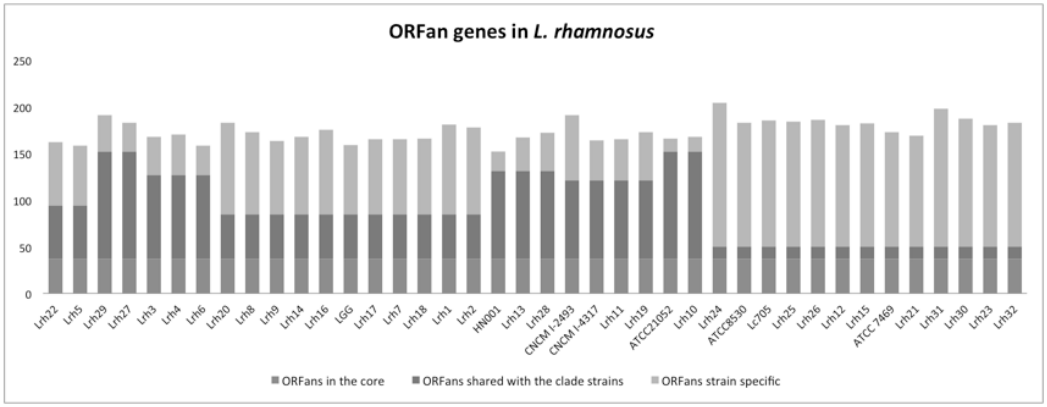


Supplementary Figure 3. Genomes alignment to strain GG. Highlighted region in red contains an enrichment of carbohydrate transport PTS genes.



Supplementary Figure 4. Unique and ORF-an genes of the *L. rhamnosus* strains.





Supplementary Table.1. A general overview of the origin and genome statistics of the 40 *Lactobacillus rhamnosus* genomes used in the study: size, GC content, number of genes, niche of isolation, sequencing parameters and previous literature. The gene locuses names used for GG genome are based on the original locus tag published by Kankainen et al. and not the newly assigned tags by NCBI. The original names are still there within the Genbank files and are called old locus tag. Table is available for download at the Wageningen University Library at: http://library.wur.nl.

Supplementary Table.2. Presence absence map for all orthologous groups (OGs) present in the *L. rhamnosus* strains. Fields representing number of genomes the protein is present in size, closest hit in the NCBI database, ATCC 53103 gene number from the public databases and whether the genes contains an LPXTG motif are added to the table. Genomes order is based on the genetic clades. Table is available for download at the Wageningen University Library at: http://library.wur.nl.

Supplementary Table 3. Clade specific OGs.

	Clade1	Clade2	Clade3	Clade4	Clade5	Clade6	Clade7	Clade8
Regulation	13	17	7	9	7	4	0	0
Restriction	5	0	1	3	2	0	1	6
PTSs	5	11	14	2	1	0	0	0
ABCs	6	7	1	1	7	1	0	0
Plasmid genes	5	6	3	7	0	2	1	2
Sortase	0	1	0	0	0	0	0	0
Toxin	0	2	0	0	0	0	0	0
Mobile elements	23	0	12	12	8	24	32	12
Hypotheticals	188	137	122	129	105	40	20	34
Total	283	212	225	199	148	65	28	49

Supplementary Table 4. Predicted peptidase OGs of the *L. rhamnosus* strains.

KEGG	Gene name	Protease type	Annotation	OG	Presence
K03101	lspA	Aspartic Peptidases	signal peptidase II	orth347	core
K00764	purF	Cysteine Peptidases	amidophosphoribosyltransferase	orth1865	core
K00820	glmS	Cysteine Peptidases	glucosamine--fructose-6-phosphate aminotransferase	orth2970	core
K01304	pcp	Cysteine Peptidases	pyroglutamyl-peptidase	orth3100	core
K01372	pepC	Cysteine Peptidases	bleomycin hydrolase	orth569	core
K01951	guaA	Cysteine Peptidases	GMP synthase (glutamine-hydrolysing)	orth1928	core
K01953	asnB	Cysteine Peptidases	asparagine synthase (glutamine-hydrolysing)	orth2202	core
K05520	pfpI	Cysteine Peptidases	protease I	orth1891	core
K07010		Cysteine Peptidases	putative glutamine amidotransferase	orth2534, orth4421	core
K07284	srtA	Cysteine Peptidases	sortase A	orth1655, orth3244	core
K08659	pepDA, pepDB	Cysteine Peptidases	dipeptidase	orth1130, orth2096, orth2237	core
K01256	pepM	Metallo Peptidases	aminopeptidase N	orth2997	core
K01258	pepT	Metallo Peptidases	tripeptide aminopeptidase	orth122	core
K01262	pepP	Metallo Peptidases	Xaa-Pro aminopeptidase	orth3107	core
K01265	map	Metallo Peptidases	methionyl aminopeptidase	orth3605	core
K01271	pepQ	Metallo Peptidases	Xaa-Pro dipeptidase	orth2552	core
K01409	gcp	Metallo Peptidases	O-sialoglycoprotein endopeptidase	orth2555	core
K01436		Metallo Peptidases	amidohydrolase	orth1222	core
K03798	ftsH	Metallo Peptidases	cell division protease	orth255	core
K05823		Metallo Peptidases	N-acetyldiaminopimelate deacetylase	orth1901	core
K07386	pepO	Metallo Peptidases	putative endopeptidase	orth1148, orth2704	core
K08602	pepF, pepB	Metallo Peptidases	oligoendopeptidase F	orth2583	core
K08643	zmpB	Metallo Peptidases	zinc metalloprotease	orth470	variable
K11749	rseP	Metallo Peptidases	regulator of sigma E protease	orth2271	core
K12941	abgB	Metallo Peptidases	aminobenzoyl-glutamate utilization protein B	orth614	variable
K16203	dppA	Metallo Peptidases	D-amino peptidase	orth1368	variable
K01259	pip	Serine endopeptidases	proline iminopeptidase	orth858, orth3239	core
K01281	pepXP	Serine endopeptidases	X-Pro dipeptidyl-peptidase	orth3514	core
K01338	lon	Serine endopeptidases	ATP-dependent Lon protease	orth2471	variable
K01356	lexA	Serine endopeptidases	repressor LexA	orth978	core
K01358	clpP	Serine endopeptidases	ATP-dependent Clp protease, protease subunit	orth117	core
K01361		Serine endopeptidases	lactocepin	orth1104	core
K03100	lepB	Serine endopeptidases	signal peptidase I	orth2680	core
K03797	ctpA	Serine endopeptidases	carboxyl-terminal processing protease	orth292	core
K07258	dacA, dacB, dac(	Serine endopeptidases	D-alanyl-D-alanine carboxypeptidase (penicillin-bind	orth2197	core
K01419	hslV, clpQ	Threonine Peptidases	ATP-dependent HslUV protease	orth1201	core

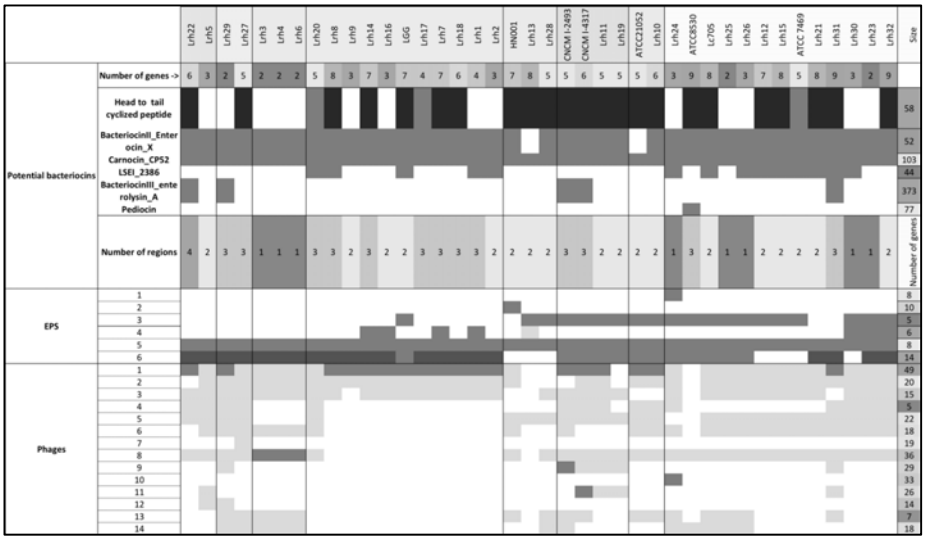
Supplementary Table.5. Specific genes for *L. rhamnosus* strains – listing of OG numbers and RAST annotation. Table is available for download at the Wageningen University Library at: http://library.wur.nl.

Supplementary Table.6. All identified transporter OGs. Table is available for download at the Wageningen University Library at: http://library.wur.nl

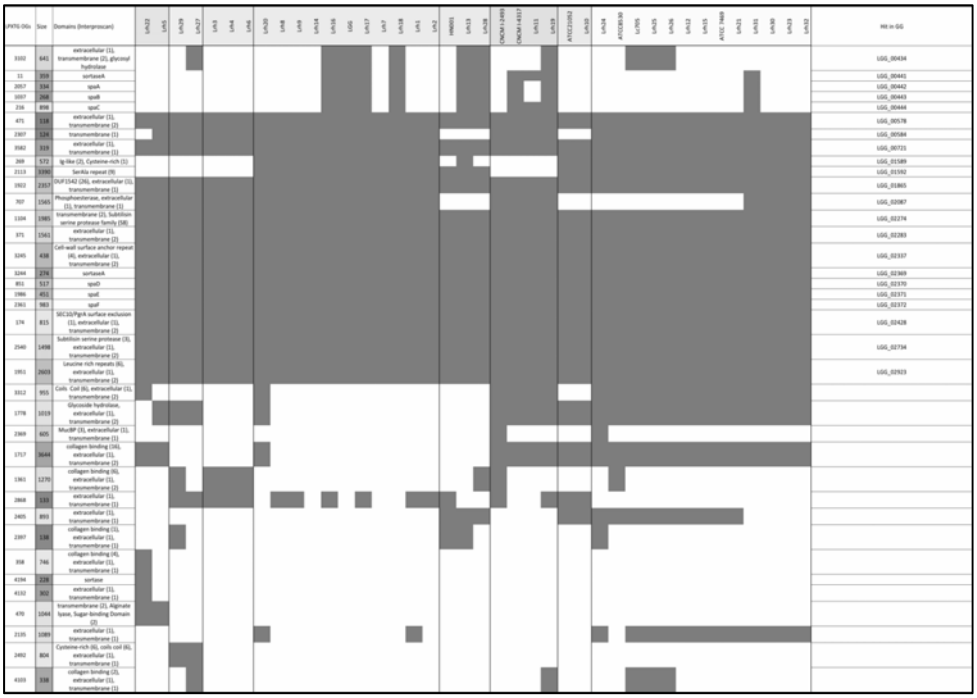
Supplementary Table.7. KEGG - KAAS created metabolic maps summary.

KEGG MAP	Metabolism overview	pan	core	clade1	clade2	clade3	clade4	clade5	clade6	clade7	clade8
10	Glycolysis / Gluconeogenesis	26	24	24	25	25	24	25	24	25	25
20	Citrate cycle (TCA cycle)	11	10	11	11	11	10	11	10	11	11
30	Pentose phosphate pathway	23	17	19	20	20	21	20	21	20	20
40	Pentose and glucuronate interconversions	22	10	16	15	15	12	15	15	15	15
51	Fructose and mannose metabolism	33	20	27	29	29	28	29	29	29	29
52	Galactose metabolism	31	18	21	28	30	23	28	31	28	30
53	Ascorbate and aldarate metabolism	5	4	4	5	5	4	5	5	5	5
61	Fatty acid biosynthesis	11	11	11	11	11	11	11	11	11	11
71	Fatty acid metabolism	2	2	2	2	2	2	2	2	2	2
72	Synthesis and degradation of ketone bodies	3	2	2	2	2	3	2	2	2	2
2010	ABC transporters	74	57	68	63	62	65	63	62	63	62
2020	Two-component system	27	24	27	26	26	25	26	25	26	26
2030	Bacterial chemotaxis	1	0	1	1	1	1	1	1	1	1
2060	Phosphotransferase system (PTS)	43	26	30	42	43	30	42	38	42	42
3010	Ribosome	52	52	52	52	52	52	52	52	52	52
3013	RNA transport	2	2	2	2	2	2	2	2	2	2
3018	RNA degradation	8	8	8	8	8	8	8	8	8	8
3020	RNA polymerase	5	5	5	5	5	5	5	5	5	5
3030	DNA replication	14	14	14	14	14	14	14	14	14	14
3060	Protein export	10	10	10	10	10	10	10	10	10	10
3070	Bacterial secretion system	9	8	9	8	9	8	8	9	8	9
3410	Base excision repair	11	11	11	11	11	11	11	11	11	11
3420	Nucleotide excision repair	7	7	7	7	7	7	7	7	7	7
3430	Mismatch repair	17	16	16	17	16	16	17	16	17	16
3440	Homologous recombination	19	19	19	19	19	19	19	19	19	19
130	Ubiquinone and other terpenoid-quinone biosynthesis	6	1	3	3	3	4	3	2	3	3
190	Oxidative phosphorylation	13	12	13	13	13	12	13	12	13	13
195	Photosynthesis	8	8	8	8	8	8	8	8	8	8
230	Purine metabolism	47	47	47	47	47	46	46	46	47	46
240	Pyrimidine metabolism	41	41	40	42	42	40	41	41	42	41
250	Alanine, aspartate and glutamate metabolism	18	18	18	18	18	18	18	18	18	18
253	Tetracycline biosynthesis	4	4	4	4	4	4	4	4	4	4
260	Glycine, serine and threonine metabolism	16	12	14	15	16	13	15	15	15	16
270	Cysteine and methionine metabolism	16	14	16	14	16	14	14	16	14	16
280	Valine, leucine and isoleucine degradation	9	6	6	6	6	6	6	6	6	6
290	Valine, leucine and isoleucine biosynthesis	2	2	2	2	2	2	2	2	2	2
300	Lysine biosynthesis	13	13	13	13	13	13	13	13	13	13
310	Lysine degradation	1	1	1	1	1	1	1	1	1	1
311	Penicillin and cephalosporin biosynthesis	1	1	1	1	1	1	1	1	1	1
312	beta-Lactam resistance	1	1	1	1	1	1	1	1	1	1
330	Arginine and proline metabolism	14	12	13	13	13	13	13	14	13	13
340	Histidine metabolism	11	11	11	11	11	11	11	11	11	11
350	Tyrosine metabolism	5	5	5	5	5	5	5	5	5	5
360	Phenylalanine metabolism	3	3	3	3	3	3	3	3	3	3
361	Chlorocyclohexane and chlorobenzene degradation	1	1	1	1	1	1	1	1	1	1
362	Benzoate degradation	4	4	4	4	4	4	4	4	4	4
363	Bisphenol degradation	1	1	1	1	1	1	1	1	1	1
380	Tryptophan metabolism	2	2	2	2	2	2	2	2	2	2
400	Phenylalanine, tyrosine and tryptophan biosynthesis	8	7	7	7	7	8	7	7	7	7
401	Novobiocin biosynthesis	1	1	1	1	1	1	1	1	1	1
410	beta-Alanine metabolism	1	1	1	1	1	1	1	1	1	1
430	Taurine and hypotaurine metabolism	2	2	2	2	2	2	2	2	2	2
450	Selenocompound metabolism	7	7	7	7	7	7	7	7	7	7
460	Cyanoamino acid metabolism	3	3	3	3	3	3	3	3	3	3
471	D-Glutamine and D-glutamate metabolism	2	2	2	2	2	2	2	2	2	2
473	D-Alanine metabolism	4	4	4	4	4	4	4	4	4	4
480	Glutathione metabolism	7	7	7	7	7	7	7	7	7	7
500	Starch and sucrose metabolism	23	21	22	21	21	22	21	21	21	21
510	N-Glycan biosynthesis	1	1	1	1	1	1	1	1	1	1
511	Other glycan degradation	4	1	2	2	3	2	2	4	2	3
520	Amino sugar and nucleotide sugar metabolism	28	21	27	23	25	26	23	25	23	26
521	Streptomycin biosynthesis	8	4	5	8	8	8	8	8	8	8
523	Polyketide sugar unit biosynthesis	4	0	1	4	4	4	4	4	4	4
524	Butirosin and neomycin biosynthesis	1	1	1	1	1	1	1	1	1	1
531	Glycosaminoglycan degradation	1	1	1	1	1	1	1	1	1	1
540	Lipopolysaccharide biosynthesis	2	0	2	0	2	0	0	2	0	2
550	Peptidoglycan biosynthesis	15	15	15	15	15	15	15	15	15	15
561	Glycerolipid metabolism	9	9	9	9	9	9	9	9	9	9
562	Inositol phosphate metabolism	10	10	10	10	10	10	10	10	10	10
564	Glycerophospholipid metabolism	9	9	9	9	9	9	9	9	9	9
590	Arachidonic acid metabolism	1	1	1	1	1	1	1	1	1	1
591	Linoleic acid metabolism	1	1	1	1	1	1	1	1	1	1
600	Sphingolipid metabolism	3	1	2	2	3	1	2	3	2	3
603	Glycosphingolipid biosynthesis - globo series	1	1	1	1	1	1	1	1	1	1
620	Pyruvate metabolism	25	24	24	24	24	24	24	25	24	24
621	Dioxin degradation	2	2	2	2	2	2	2	2	2	2
622	Xylene degradation	3	3	3	3	3	3	3	3	3	3
623	Toluene degradation	2	1	2	2	2	1	2	1	2	2
624	Polycyclic aromatic hydrocarbon degradation	1	1	1	1	1	1	1	1	1	1
625	Chloroalkane and chloroalkene degradation	3	3	3	3	3	3	3	3	3	3
626	Naphthalene degradation	1	1	1	1	1	1	1	1	1	1
627	Aminobenzoate degradation	3	3	3	3	3	3	3	3	3	3
630	Glyoxylate and dicarboxylate metabolism	7	5	7	7	7	6	7	6	7	7
640	Propanoate metabolism	11	10	10	10	10	11	10	10	10	10
643	Styrene degradation	1	1	1	1	1	1	1	1	1	1
650	Butanoate metabolism	13	10	11	11	11	10	11	10	11	11
660	C5-Branched dibasic acid metabolism	2	2	2	2	2	2	2	2	2	2
670	One carbon pool by folate	10	10	10	10	10	10	10	10	10	10
680	Methane metabolism	14	9	10	12	12	9	12	12	12	12
710	Carbon fixation in photosynthetic organisms	12	10	10	10	10	11	10	10	10	10
720	Carbon fixation pathways in prokaryotes	13	12	13	13	13	12	13	12	13	13
730	Thiamine metabolism	8	6	8	8	8	8	8	8	8	8
740	Riboflavin metabolism	1	1	1	1	1	1	1	1	1	1
750	Vitamin B6 metabolism	3	2	3	2	2	2	2	2	2	2
760	Nicotinate and nicotinamide metabolism	7	6	7	6	6	6	6	6	6	6
770	Pantothenate and CoA biosynthesis	8	7	7	7	8	7	7	7	7	8
780	Biotin metabolism	4	4	4	4	4	4	4	4	4	4
785	Lipoic acid metabolism	1	1	1	1	1	1	1	1	1	1
790	Folate biosynthesis	2	2	2	2	2	2	2	2	2	2
860	Porphyrin and chlorophyll metabolism	3	3	3	3	3	3	3	3	3	3
900	Terpenoid backbone biosynthesis	12	11	11	11	11	11	11	12	11	11
908	Zeatin biosynthesis	1	1	1	1	1	1	1	1	1	1
910	Nitrogen metabolism	8	5	5	5	5	5	5	5	5	5
920	Sulfur metabolism	5	5	9	8	6	8	8	6	6	6
930	Caprolactam degradation	1	1	1	1	1	1	1	1	1	1
940	Phenylpropanoid biosynthesis	1	1	1	1	1	1	1	1	1	1
944	Flavone and flavonol biosynthesis	1	1	1	1	1	1	1	1	1	1
960	Tropane, piperidine and pyridine alkaloid biosynthesis	1	1	1	1	1	1	1	1	1	1
970	Aminoacyl-tRNA biosynthesis	26	26	26	26	26	26	26	26	26	26
983	Drug metabolism - other enzymes	8	7	7	8	8	7	8	8	8	8
1040	Biosynthesis of unsaturated fatty acids	2	2	2	2	2	2	2	2	2	2
1051	Biosynthesis of ansamycins	1	0	0	0	0	1	0	0	0	0
1054	Nonribosomal peptide structures	1	1	1	1	1	1	1	1	1	1
1055	Biosynthesis of vancomycin group antibiotics	1	0	1	1	1	1	1	1	1	1

Supplementary Table.8. Bacteriocins, EPS and phages. A. Presence and number of bacteriocins identified using BAGEL. Legend: grey = one copy of the gene /strain and red >1 gene/ strain. B. Presence and number of EPS clusters and phages identified. Strains are organized by genetic clade. Legend: light grey = one copy of the gene /strain, grey = between 1 and 10 genes/ strain and dark grey >10 genes/ strain. C.



Supplementary Table.9. Selected *L. rhamnosus* genes and gene clusters associated with an effect on human cells by *in vitro* studies and proteins containing an LPxTG domain identified in the *L. rhamnosus* pan-genome and their domain composition.



CHAPTER 7

General discussion

An astonishing microbial diversity is encountered in the intestinal environment and its functions are indispensable for the activity of not only the intestine, but also the entire human organism (1). Intestinal bacteria can have beneficial effects for the host such as the digestion of complex carbohydrates, colonization resistance against invading pathogens, maturation of the adaptive mucosal immune system, and the production of metabolites, including vitamins, SCFA and essential amino acids (2). Therefore, gathering information about the cross-talk between microorganisms and their environment is needed to understand our own physiology (3). Among the common inhabitants of the human intestine, lactobacilli have been recognized as potential health beneficial microorganisms for the human (mammalian) host (4, 5). Strengthening of the intestinal barrier, modulation of the immune response, and antagonism of pathogens either by the production of antimicrobial compounds or through competition for nutrients and/or mucosal binding sites, are only some of the suggested ways lactobacilli can benefit their host. Although new techniques continue to improve our understanding of the human intestinal interactions, the precise molecular mechanisms underlying most of these beneficial activities remain poorly understood. Remarkably, while studies have linked the species *Lactobacillus rhamnosus* to all of the health benefits mentioned above (6–13), no comprehensive study established whether the effects are strain specific or conserved for all strains of the species.

This thesis is intended to provide further understanding of the effects of (single) bacterial strains on human health, an ambition that principally would require the analysis of bacterial health benefits *in situ* (14). However, the complexity of the intestinal environment and the lack of appropriate tools to investigate interactions at a microenvironment level hamper such *in situ* approaches, and stimulate the fundamental and mechanistic work that can be done *in vitro*. As shown in Chapter 2, studies performed with single strains are convenient mainly because controlled settings allow for targeted research and are more probable to generate unambiguous results. Notably, such experiments have already given very valuable insights in the molecular responses of bacteria to various environmental conditions such as nutrients or immune factors from the host (15, 16). Therefore, *in vitro* research can facilitate the selection of effective strains or strain-combinations and design studies to evaluate their health effects in animal models and eventually in humans.

In this context, the line of experimentation followed in this thesis provided a better understanding of the genetic and phenotype diversity of *L. rhamnosus* strains and unravelled molecular mechanism of host responses towards strains of this species.

Strains of *L. rhamnosus* can be isolated from various environments, e.g., GI-tract, reproductive system, dairy and plant (fermentations). The majority of the phenotyping studies presented in this thesis were performed with a set of strains shown to have a different genetic profile by AFLP (Chapter 3). Different strategies were employed to better understand the adaptation capacity of strains of this species to their environment through reviewing their phenotypic diversity, including the profiling of the capacity to utilize different carbon sources, exert differential immune-cell stimulation, pathogens inhibition and surface property characteristics.

The use of a broad array of carbohydrate-phenotype arrays (Biolog) (17) can be a suitable approach to classify bacterial strains of the same species (18, 19). We show in Chapter 3 that phenotyping arrays are providing a valuable method for cataloguing bacterial strains, which was shown to partially match their genetic classification by AFLP, and provided more direct metabolic and phenotypic information. Use of Biolog phenotyping was previously evaluated as a good method for species (20), but not strain identification (21, 22). In this thesis, we argue (Chapter 3) that the number of tested carbon sources is relevant, therefore an advanced understanding on a species' metabolic capacity repertoire would allow for the development of a more concise carbon-utilization-based strain identification method, which has a similar discriminatory efficiency as the genetic fingerprinting methods for strain typing and classification (Chapter 3) (23). All *L. rhamnosus* strains have in common the ability to use 14 carbon sources that can be considered to be highly available in the intestine and diet, and encompass mono, di- and oligo- or polysaccharides. Profiling of the 58 differentially used carbon sources led to the classification of strains in three metabolic clusters that are partially coinciding with the AFLP-based genetic grouping, indicating an important role for carbohydrate utilization in the species' evolution. Analogously, genes involved in carbohydrate utilization, especially transporters, dominate the species variome (Chapter 6). Metabolic grouping appears unrelated to the recorded origin of isolation of the strains, showing either that other ecological pressures caused the group separation, or the difficulty to record the real environmental fitness of an environmentally promiscuous bacterial species like *L. rhamnosus*. The second hypothesis could be well supported, considering that previous genetic diversity studies illustrated this species' adaptation patterns to a diversity of environments (24).

Currently marketed probiotic strains of *L. rhamnosus* classify into the three different metabolic groups. A large number of plant-derived carbohydrates are used by close relatives of the type strain (ATCC 7469), while isolates more related to strain HN001 have an increased peptidase repertoire, a characteristic of dairy strains. Isolates

belonging to the third group, that includes strain GG, lack the ability to use several plant-derived carbohydrates but maintain the proteolytic repertoire. Genome reduction by loss of metabolic genes was observed in the related species *L. casei* (25), in isolates that were adapted to the dairy environment (26). A similar phenomenon could possibly be deduced from the genomic content and phenotypic characteristics of GG-related strains of *L. rhamnosus*, which may have lost metabolic capacities due to their long-term use in the dairy environment over the last 20 years (27).

The carbohydrate utilization profiling also provided information on carbon sources that could function as prebiotics for certain *L. rhamnosus* strains (Chapter 3). For instance, out of the 25 strains, HN001 is able to use the largest number of carbon sources (42) while strain GG displayed higher growth efficiency on a reduced number of carbohydrates compared to other strains. Several carbohydrates could be utilized by only few strains (less than four), e.g., D-galactonic acid-γ-lactone, 2-deoxy-D-ribose, dihydroxyacetone, L-arabinose and maltose. Dulcitol, maltitol and gentiobiose appeared to be utilized by only a single strain.

Notably, 11 out of the 25 strains tested cannot use L-rhamnose for growth in either API50 or Biolog phenotyping, while their genomes still appear to encode the complete operon predicted to be involved in this phenotypic trait (Chapter 3, Chapter 6). This is remarkable since this carbohydrate can be considered as the discriminator for the name of the species. Taken together, it can be proposed that a so-far unidentified mutation of the rhamnose utilization operon is responsible for the inactivation of this phenotype in these strains (24).

Genome sequencing for 40 diverse *L. rhamnosus* strains complemented the phenotyping studies. Sequencing followed by gene annotation generated a large amount of information about the evolution of a microbial species that can be used to explain and predict discriminating physiological and other phenotypic traits that are strain specific rather than common to the members of the species. As presented in Chapter 6, *L. rhamnosus* strains are characterized by a large genetic diversity, hallmarked by carbohydrate transport and utilization, proteolytic activity, bacterial immunity systems, regulatory systems and a large array of surface molecules that are potentially involved in attachment and communication with the host. This inherent genetic diversity is in agreement with a versatile species that has the ability to survive in very different environments (28), which is a characteristic shared with the species *L. plantarum*, and which is also reflected by the relatively large genome size of both these species within the Lactobacillus genus (29).

Genome analysis with the automated bacteriocin mining software Bagel revealed the presence of six potential bacteriocin production systems of which only one appeared to be conserved in all strains. The other five systems appeared to be variably present in the *L. rhamnosus* strains (this thesis, Chapter 6) and the possible involvement of this particular genetic variation in differential inhibitory effects of individual *L. rhamnosus* strains against pathogens was evaluated experimentally. To this end, pathogen inhibition assays were performed using culture supernatant from *L. rhamnosus* strains in well-diffusion assays targeting five candidate pathogenic bacteria, i.e., *Listeria monocytogenes, Salmonella enteridis, Staphylococcus aureus, Escherichia coli and Pseudomonas aeruginosa*. The strongest pathogen-inhibitory effects were observed with supernatants obtained from the strains Lrh044, Lrh139, Lrh141, and with *S. enteriditis and L. monocytogenes* as target pathogens. However, the inhibitory effects did not appear to correlate with the presence of any of the predicted bacteriocin production systems, thereby failing to pinpoint a role of the bacteriocin genes in this phenotype. However, since the expression of the bacteriocin clusters in individual strains was not evaluated in this study, existing correlations may have been obscured by differential regulation of bacteriocin expression in the different strains encoding them, which may include previously reported variability of quorum-sensing mediated control of bacteriocin production by lactic acid bacteria [1, 2].

The remarkable synchronicity between strain phylogeny (studied by core gene conservation) and genetic diversity (presence – absence of genes) proposes that unlike other lactobacilli, evolution of this species is highly dependent on horizontal transfer events from other species and that inter-species genetic information exchange only contributes to a limited extent to its evolutionary differentiation. This characteristic may have a molecular explanation in the clustered regularly interspaced short palindromic repeats (CRISPR)/cas system, which are hypervariable loci widely distributed in bacteria and archaea, that provide acquired immunity against foreign genetic elements (34). In *L. rhamnosus*, the locus was found to be acting primarily against phages, transposases, and plasmids of the same or closely related strains. Nevertheless, only about half of the specificities for the CRISPR spacers could be identified, due to the limitations of the current knowledge on the mobile elements of lactobacilli. An overview of CRISPR spacers belonging to LAB species showed a similar profile to the one identified in *L. rhamnosus*, with about 75% of the sequences being mapped to phage sequences, and the remaining sequences mapped to either plasmids or chromosomal genes (35). Compared to our study of 40 strains of a single species that allowed the mapping of 112 of the overall 245 spacers, the previously

reported generic analysis of LAB spacer-mapping that included 11 LAB species, could only map a much smaller number of spacers relative to the number of studied species (35), leaving much room for further study. The CRISPR-mediated immunity could be of significant industrial interest since industrially relevant strains could be adapted to be resistant to phages known to exist in the industrial environment by CRISPR mediated immunity acquisition (36). This could provide an effective strategy for the development of robust next-generation starter cultures with increased industrial resilience.

The analysis of strain phylogeny based on core genes variation indicates that the group including strain HN001 was probably the last to emerge, preceded by the group of strains that contains strain GG. The recently discovered mucus binding pili (37) found in multiple strains of the GG genetic clade suggests that this clade might be a specialized intestinal group of strains, a finding that was also reinforced by the results of the carbohydrate utilization phenotype characterization (this thesis, Chapter 3).

Harnessing genomic diversity; the quest for phenotype effector molecules

Genotype phenotype matching was previously used successfully, providing functional annotations to health related proteins. The approach is also suitable for the identification of genes involved in complex phenotypes like strain-specific immunomodulation, as was illustrated by the identification of *L. plantarum* genes involved in modulation of responses in dendritic cells (38) and peripheral blood mononuclear cells (39). However, the efficacy of the procedure is very much enhanced by the presence of an evenly spread, but substantial phenotypic variation between strains and requires a relatively low level of experimental noise (e.g. high reproducibility of phenotypic outcomes).

Association analysis for genes and phenotypes, using low-pass genome sequences of a subset of the strains tested, revealed candidate operons for the utilization of L-Sorbose and α–Methyl–D–Glucoside, discriminating sugars that allow for separation of the metabolic groups (Chapter 3). Genetic engineering is an established approach to determine the functional properties of genes, and has been extensively used to provide proof of function of genes, for example through the characterization of expression or deletion or (over-)expression mutants (18, 40–43). In this thesis, the genes predicted to encode the utilization capacity for the discriminative sugars Sorbose and α–Methyl–D–Glucoside was confirmed using a complementation / expression strategy, employing heterologous plasmid-borne expression of the candidate operons under control of their own promotors, in strains that lack these operons (Chapter 3).

The relevance of L-Fucose metabolism in intestinal physiology is exemplified in Chapter 4 (44). Notably, fucose display on the surface of bacteria might stimulate bacterial recognition by immune cells since dendritic cells have specific receptors that recognize terminally linked fucose moieties (45). Therefore, bacterial strains with the capacity to produce surface glycans that have fucose incorporated are likely to be differentially recognized compared to non-fucosylated bacteria. Moreover, the body of evidence supporting a role for L-Fucose release and metabolism in intestinal persistence of bacteria is increasing (46). The core genome of *L. rhamnosus* is predicted to encode two fucosidases, which are enzymes involved in the release of fucose from environmental oligo- and polysaccharides. The carbohydrate utilization profiling indicated that L-Fucose was among the carbohydrates differentially utilized by the *L. rhamnosus* strains, which allowed us to identify a candidate operon responsible for this phenotype (Chapter 4). The role of this operon was investigated by the construction of deletion mutant for the fucosyl kinase (*fucK*), a gene that appeared to be unique for the species *L. rhamnosus* and is part of the operon and predicted to be essential for the L-Fucose catabolic pathway. However, although the phenotype of the mutant strain clearly supported the role of the fucose kinase function in the efficient utilization of L-Fucose as a carbon source for fermentation and growth, this function appeared to be redundant in *L. rhamnosus* GG since the mutant strain could in time reinstate L-fucose utilization albeit with a lower efficiency as compared to the parental strain.

Bacterial fitness studies in highly variable environments have shown that environmental flexibility and robustness can in part be achieved by genetic redundancy of critical functions, where the functional failure of a specific gene product through mutation or environmental challenge can be compensated by duplication of such function using either a very similar or quite distinct gene duplicate that can exert the same function (47). The evidence obtained using the *fucK* mutant strain provides support for the redundancy of the genes involved in fucose utilization in *L. rhamnosus* where the fucose kinase function can be compensated by the expression of an alternative (unidentified) gene and the additional L-Fucose utilization functions present in the operon have potential homologs in the genome.

Much effort has been given to the development of optimal immune models to distinguish strains of the species *L. rhamnosus* on basis of their immunomodulatory capacities. Several published models that were successfully employed with other species revealed no or limited variation between different strains of *L. rhamnosus* (Figure 1). Below a short summary of the models where strains showed uniformity and

therefore couldn't be used in the genotype – phenotype matching is given. The immune response of different strains of this species was addressed *in vitro*, in co-culture with immune cells, which were either stabilized cell lines (monocytes, macrophages) or primary cells (neutrophils, monocytes, B and T lymphocytes) isolated from blood.

In contrast, in Chapter 5, PBMC stimulation assays revealed considerable diversity in strain specific cytokine induction capacities. Gene trait matching can reveal candidate genetic markers involved in this difference, but verification of such associations by genetic engineering (expression or deletion derivatives of natural strains) is required to establish the role of such candidate immuno-effector molecules. Identification of such effector molecules can drive the selection of strains with enhanced functional capacities on basis of their molecular characteristics (e.g., effector molecule expression), and provides product quality control criteria that are based on molecular insight in the working mechanisms of the strains and the presence of the identified effector molecule in a product, rather than on the number of viable cells administered, which currently is the only functional specification provided to the consumers of probiotic products. It has been proposed that cell surface bound polymers and/or polysaccharides provide concealment of bacterial immune stimulatory ligands from the host immune system (43, 53–55), which may explain why human epithelial cells did not differentially respond to the different *L. rhamnosus* isolates. In addition, a generally conserved immunosuppressive capacity of *L. rhamnosus* could also be responsible for the immune inactivity of the strains of this species in some of these assays. *L. rhamnosus* cells produce on their surface, and partially shed to their environment, some anti-inflammatory molecules, including the major secreted proteins that derive from peptidoglycan hydrolases Msp2 (p40) and Msp1 (p75), which were shown to reduce cytokine production in epithelial cell lines (56) and mouse tissues *ex vivo* (57).

While all *L. rhamnosus* strains encode the anti-inflammatory compounds Msp1 and Msp2 (57) (this thesis and (24)) and induce similar cytokine production levels in epithelial and immune cell lines (this thesis and (58)) that were not previously stimulated (this thesis and (59)), there may still be other (un)known components that play a role in the differential immune recognition of *L. rhamnosus* strains. The strain-specific extracellular polysaccharide (EPS) production in *L. rhamnosus* could in part be responsible for this phenomenon. In *L. plantarum,* deletion of all four *cps* encoding gene clusters led to increased toll-like receptor (TLR)-mediated signalling of the mutant strain (43).

Table 1. *In vitro* models and their results in testing strain variation in *Lactobacillus rhamnosus* strains.

***In vitro* models**	**CK**	**Result**	**Ref.**
Co- culture CaCo2 epithelial cells with *L. rhamnosus* cells	IL-8	Strains and supernatants are not differentiated.	[3]
Co-culture HT29 epithelial cells with *L. rhamnosus* cells	TNF-α, IFN-γ, IL-10, IL-17, IL-12	Strains and supernatants are not differentiated.	[4]
Co-culture THP1 monocytes with *L. rhamnosus* cells	NFkB activity	Hyper variability.	[4]
Co-culture THP1 PMA-stimulated macrophages with *L. rhamnosus* cells	NFkB activity	Hyper variability.	[5]
Co-culture TNF-α activated PMA-stimulated THP-1 macrophages with *L. rhamnosus* cells	NFkB activity	Hyper variability.	[6]
Co-culture THP1 macrophages with *L. rhamnosus* cells	Phagocytosis assay using Phrodo.	Strains and supernatants are not differentiated.	[7]
Co-culture peripheral blood mononuclear cells (PBMCs)	IL-1, IL-8, IL-10, IL-12, TNF-α, IFN-γ.	Variation observed only for IL-1, IL-10 and TNF-α.	[8]

L. rhamnosus strains are predicted to encode two to four different loci that are predicted to be involved in exopolysaccharides production, which may in part explain their differential immune-modulatory capacity. In addition, only some strains contain genes encoding mucus- and fibrinogen-binding surface proteins (this thesis, Chapter 6), which may lead to differential binding to mucus or epithelial cells. This phenotypic variation of binding to mucus (37, 60, 61) and epithelial cells (6, 8) has previously

been reported and may influence immune or epithelial cell recognition as well as the persistence of strains in the intestinal tract.

Lessons from uniformity in *L. rhamnosus*

Evolution is driven by the 'universal' availability of genetic information combined with environmental fitness selection, leading to the increased abundance of better-adapted variants of a species. Thereby environmental selection can be considered to provide a filter for genetic information that supports fitness. Consequently, bacterial genomes can be interpreted as a record of environmental conditions, and genomic uniformity can thus be seen as the resultant of continuous and consistent environmental conditions that drive the selection of clonal populations characterized by gene-loss with optimal fitness for the most important changes/fluctuations of the niche they inhabit (62). It was shown that with a relieved selective pressure, *Lactobacillus* species tend to lose genes, giving rise to strain specific differences and environmental differentiation of lineages of strains within the species (63, 64). It is therefore important to understand under which conditions the environment exerts strong and variable, or weak and consistent selective pressure on bacteria, which may drive evolutionary diversity, and clonal populations and gene loss, respectively.

In our evaluation of phenotypic diversity among strains of the species *L. rhamnosus*, many phenotype differences between strains were statistically insignificant. These evaluations span a wide range of characteristics, including surface properties (hydrophobicity, auto-aggregation), immune recognition by epithelial cells (Figure 1) or macrophages (Table 1) and metabolic properties like utilization of a particular set of carbohydrates (Chapter 3).

The genes encoding the capacity to utilize D-Ribose, N-Acetyl-D-Glucosamine, D-Galactose, D-Tagatose, D-Trehalose, D-Mannose, α-D-Glucose, L-Lyxose, Salicin, D-Mannitol, L-Arabinose, 2-Deoxy-D-Ribose appear to be essential for the survival of *L. rhamnosus* strains in their diverse habitats and were universally conserved among members of the species (Chapter 6). While this is understandable for some of these carbohydrates that are frequently encountered in a variety of environmental niches, some others such as D-Mannitol, D-Tagatose, D-Trehalose, L-Lyxose, Salicin, are considered to be present only in small amounts and are predominantly associated with plant-associated habitats. The conservation of these phenotypic capacities and their encoding genes is quite interesting and might be a reflection of the regular recurrence of all *L. rhamnosus* lineages in the plant environment. Bachman et al. described elegant experimental evolution experiments that illustrated the genetic adaptation

(termed domestication) of plant-derived *Lactococcus lactis* to the dairy environment, showing that typical plant-environment related functions were suppressed during the adaptation, leading to adapted variants that converged to phenotypic characteristics that are typically found in isolates of the same species that have been cultured for an extended period in the dairy environment (66). Analogous experiments may elucidate the adaptations of *L. rhamnosus* isolates when they are cultured for extended periods in a specific habitat, to obtain a clearer understanding of the functions in this species that determine habitat-specific fitness.

The consistency of cell surface properties of different strains of *L. rhamnosus* in terms of aggregation and hydrophobicity (Figure 1) implies a high degree of conservation among enzymes and pathways involved in cell wall biosynthesis and turnover. This is agreement with the observation that all genes with an annotation that implicates their role in cell wall formation and modification are among the core genomic functions of the *L. rhamnosus* species (Chapter 6). Conversely, one of the functional categories that display a considerable degree of strain-specific genomic variation is the category of cell-surface proteins, which suggests that these functions play a minor role in the determination of cell surface characteristics that lead to aggregation, or modulated surface hydrophobicity.

L. rhamnosus strains also appear to have a universally conserved pyruvate dissipation capacity, which is in agreement with the consistent acid production profiles of the strains (Figure 1).

After overnight growth on media containing excess concentrations of glucose, all *L. rhamnosus* strains tested produced similar levels of acetic acid (42±3 mmol*L-1), D-lactic acid (5.5±1 mmol*L-1), L-lactic acid (214±22 mmol*L-1), and did not produce detectable levels of butyrate or propionate. This relatively inflexible fermentative capacity of *L. rhamnosus* could reflect the importance of this metabolic behaviour for the bacterium in any environment, being dairy, plant or host. In the intestine, carbohydrates are used for anaerobic fermentation and lead to production of fermentation end products, e.g., short chain fatty acids (SCFA). These organic acids directly impact the intestinal milieu by lowering the local pH and serving as an energy source for both the host and other bacteria (65).

Strain	*Surface properties*			*Metabolism*			*Pathogen inhibition*				
	Auto-aggregation	**Hydrophobicity**	**CaCo2 secretion of IL-8**	**Acetic acid (mmol/l)**	**D-lactate (mmol/l)**	**L-lactate (mmol/l)**	***S. enteritidis***	***P. aeruginosa***	***E. coli***	***E. faecalis***	***L. monocytogenes***
Lrh004	1.12	1.30	<DL								
Lrh009	0.96	1.15	<DL								
Lrh010	1.29	1.09	<DL								
Lrh026	0.79	0.67	<DL								
Lrh032	0.89	1.01	<DL								
Lrh035	1.29	1.26	<DL								
Lrh037	0.61	0.63	<DL								
Lrh040	1.09	0.85	<DL								
Lrh044	0.66	0.65	<DL								
Lrh047	1.35	1.34	<DL								
Lrh053	0.66	0.63	<DL								
Lrh071	1.07	0.93	<DL								
Lrh074	1.19	1.21	<DL								
Lrh075	1.04	1.26	<DL								
Lrh108	1.34	1.27	<DL								
Lrh110	1.32	0.97	<DL								
Lrh133	0.87	0.82	<DL								
Lrh134	1.16	1.27	<DL								
Lrh136	1.36	1.25	<DL								
Lrh138	0.54	0.70	<DL								
Lrh139	1.30	1.25	<DL								
Lrh140	0.89	0.58	<DL								
Lrh141	0.64	0.69	<DL								
Lrh142	1.28	1.22	<DL								

*DL - detection limit

Figure 1. Phenotypic traits in *L. rhamnosus* (heatmap). Data is processed by mean averaging, in order to make it comparable between different methods. While surface properties and metabolism traits show very little variation, strains have diverse pathogen inhibition effects that can be used in a gene trait matching analysis.

Future directions

Insight into the mechanisms by which bacteria can affect human health can contribute considerably to the improvement of probiotics as functional foods, which is needed to strengthen the scientific support of health claims associated with such products that are currently subjected to stringent evaluation by legislation (67). Selected strains of *L. rhamnosus* have been shown to confer health benefits in experimental animal models (57, 68, 69) as well as in human clinical trials (10, 32, 70–72), although some of these human trials have presented contrasting outcomes, for example in the case of prevention of antibiotic-associated diarrhea (73) and reduction of atopic dermatitis risk (74–78). Both the literature and the research described in this thesis support that effects of probiotics are highly dependent on the strain's properties.

Currently identified molecular mechanisms by which *L. rhamnosus* probiotic strains affect their hosts do not fully explain their reported probiotic health benefits. The exploration of *L. rhamnosus* genomes in this thesis revealed several tempting candidates for further research. Large surface proteins are primary candidates for immune modulation (79, 80) as they are most likely accessible at the cell surface and can therefore readily be recognized by immune cells. The same protein category is also potentially involved in recognition of surfaces and may play a role in cell-adhesion as well as biofilm formation (81). The protein tentatively assigned the name 'Potential epithelium binding' (PEB) is the largest gene of the species and specific for 17 *L. rhamnosus* strains. It contains 25 DUF1542 domains that are associated with adhesion and protection from antibiotics (82, 83). Another large surface exposed protein is LGG_02923 that contains four copies of a leucine rich repeat-domain, which is rare among prokaryotes but has been proven to be essential for the virulence process of some pathogens, including *Yersinia pestis* and *Listeria monocytogenes* (84–87). Glycan hydrolases are abundantly identified in the core genome, some of these functions may influence immune modulation by modifying surface glycans structures that shield bacteria from immune recognition, or may contribute to mucus degradation that potentially is of importance in the intestinal habitat (88). The comparative genomics study also identified several novel bacteriocins and exopolysaccharides clusters that could have industrial applications in the dairy industry. Microbial exopolysaccharides (EPSs) *in situ* produced by lactic acid bacteria (LAB) can contribute to the texture of fermented dairy products such as yoghurt or cheese (89–91). In addition, bio-preservation by fermentation with antimicrobial peptide producing strains may contribute to the prevention of growth of spoilage and pathogenic

microorganisms, thereby contributing to product shelf life and safety. Adhering to this scenario, the use of lactic acid bacteria as a source of *in situ* produced antimicrobial compounds has been growing in the last decades as a natural method of food preservation (92, 93). Moreover, some reports imply that bacterial antimicrobial peptides may also play a role in signalling to host cells, which is consistent with the identification of several genes within the antimicrobial peptide encoding plantaricin cluster in *L. plantarum* as one of the effector genes in immunomodulation (38, 94). We also propose that attention should be given to functional studies on CRISPR spacers acquisition in this species, for which half of the sequences could be mapped to phages, plasmids and other mobile genetic elements, whereas the origin of the other half of the sequences captured in these loci remains unknown and may provide novel insight in (the lack of) genome mobility between strains this species. Moreover, the CRISPR system offers the possibility for immunization to phages that could support the formulation of starter cultures with broad spectrum phage-resistance, which is of clear relevance in an industrial setting since phage predation can interfere with the industrial performance of starter cultures (36).

The knowledge derived from the carbohydrate fermentation studies described here may be a starting point in the development of synbiotic concepts based on *L. rhamnosus* strains with desired health promoting properties. The combined administration of a probiotic strain together with a carefully selected growth-substrate could provide a competitive advantage to the probiotic strain relative to the endogenous intestinal microbes or even strains of the same genus (95).

When it comes to selection of new strains for *in vivo* research, extensive comparative studies should be done beforehand, because metabolic properties and effects on the immune system can vary considerably within a species (97–100). Moreover, despite recent breakthroughs (37, 101–103), the precise mechanism of action of individual strains is mostly unknown, preventing the selection of more effective strains on basis of a functional parameter. Nevertheless, current studies (38, 39, 104), including those presented in this thesis suggest that specific *in vitro* assays can provide a foundation for the rationalized selection of candidate strains for further exploration. In the context of this thesis, the metabolic capacities like fucose utilization and EPS production could be characteristics that contribute significantly to the survival and persistence in the intestine and should be taken into account when evaluating strains with potential probiotic activity (53, 105).

Genetic and metabolic comparisons also identified strains with properties that are substantially different from both the type strain and those currently used as

probiotics, e.g., GG and HN001, which are interesting strains to evaluate in terms of their industrial characteristics and possible health benefits. Strain ATCC 8530, a beer isolate, is unique by encoding only two strain specific genes, the pediocin *pedB* and a thioredoxin *txnX*, which could be part of the lifestyle adaptation of this strain. The bacteriocin may help to prevent competition by other bacterial species, while thioredoxin may contribute to tolerance to oxidative stress encountered in the beer environment (106). Interestingly, oxidative stress response related functions were induced upon ethanol exposure of *L. plantarum* (107), corroborating that these stress associated functions could play a role in the adaptation of *L. rhamnosus* to alcoholic fermentation conditions, like the production of beer or wine, and underpinning the potential added value of the thioredoxin system in this strain. The impact these genes might have for the functioning of this strain in alcoholic fermentations (beer, wine or sake) or even dairy environments could be tested by engineering of these genes in this strain (mutation) or other members of the *L. rhamnosus* group (heterologous expression) or by verification of their function by complementation of mutations in model species such as *E. coli* or *Lactococcus lactis* (108).

The ability of bacterial strains to displace and inhibit pathogens is significant both for industrial application, as shown above, as well as therapeutic manipulation of the endogenous microbiota (109, 110). In some cases, mixing strains that produce different antimicrobial compounds may prove to be an even better approach to suppress intestinal pathogens. This is exemplified by the finding that a mixture of probiotic species and strains was more effective at inhibiting the enteric pathogens *Clostridium difficile*, *Escherichia coli* and *S. typhimurium* as compared to its composing constituents, when tested at approximately equal concentrations *in vitro* (31). Notably, *L. rhamnosus* strains Lrh110, Lrh044, Lrh138 and Lrh141 could be shown to exhibit an increased ability to inhibit the growth of the enteric pathogens *S. enteriditis*, *E. fecalis* and *L. monocytogenes* (Figure 1). The candidate genes underlying these activities were identified by gene trait matching *in silico*, and encompassed genes that were annotated as hypothetical proteins. Their functional elucidation would be best served by the construction of mutant strains and their functional testing in pathogen inhibition assays in vitro, and eventually *in vivo* in an infection model in animals.

While much evidence of probiotics mode of action was investigated *in vitro*, there is a large gap between the currently used *in vitro* models and the complexity of the whole organism. The consequence of this is that some molecules that appear very efficient in the laboratory have a minimal impact when evaluated in animal models or

clinical settings. Moreover, translation of findings in animal models to efficacy in humans remains quite challenging and frequently fails, which may be due to many factors, including disparity between the animal and human metabolism and immune system. Reliable experimental systems that are derived from humans and have predictive value are currently considered with the aim to at least in part replace the animal model systems,.

In the current work, we realized the need for more suitable diversity screening models when only a single immune model led to the determination of results of sufficient diversity among the strains tested to enable gene trait matching (Table 1). Similarly to the pathogen inhibition variation, gene trait matching on basis of the variation in cytokine production profiles induced in immune cells by individual strains led to the identification of several candidate genes that may encode the effector molecules involved, including genes encoding cell-surface proteins, proteases, glycan hydrolases and bacteriocins. However, the postulated role of these genes in modulation of immune cell phenotype remains to be further validated in the laboratory by genetic engineering approaches.

During the last few years, innovative *in vitro* testing methodologies have been developed that could potentially reduce the gap between the *in vitro* screening models and the *in situ* situation. One of these systems employed the *in vitro* apical stimulation with bacteria of whole human intestinal mucosal explants that can be kept in culture for a relatively short period of time (100). The predictive value of this and similar systems needs to be further established and also aspects of its screening throughput need further attention, but this approach may offer an attractive alternative to immune assays on basis of blood-derived cells or other more artificial and highly simplified *in vitro* models that are currently available Another high potential system in this context could be constructed on basis of intestinal stem cells that can be cultured *in vitro* to form multi-cellular and differentiated organoids (111–113) that may better reflect the complexity of the human epithelium in the intestinal mucosa.

The novel developments in *in vitro* screening models that may more reliably reflect the *in vivo* situation holds great promise for the more reliable identification of bacterial molecules of probiotic and/or beneficial commensal strains that are directly involved in strain-specific host-interaction phenotypes. Such more reliable measurements can serve as input for gene-trait matching approaches to identify the bacterial genes that encode the functions responsible for these strain-specific effects. Alternatively, approaches referred to as transcriptome trait matching may also allow the identification of functions that are conserved among different strains of a species

but play a role in the host interaction potential of individual strains on basis of their differential level of expression in different representatives of the species. This approach has recently been illustrated by the identification of *L. plantarum* genes that contribute to the survival of this species under intestinal conditions (114, 115) and may hold promise for the elucidation of effector molecules and mechanisms underlying other complex and strain-specific host modulation phenotypes.

References:

1. Ottman N, Smidt H, de Vos WM, Belzer C. 2012. *The function of our microbiota: who is out there and what do they do?* Front. Cell. Infect. Microbiol.

2. Schluter J, Foster KR. 2012. *The evolution of mutualism in gut microbiota via host epithelial selection.* PLoS Biol. 10.

3. Turroni F. et al. 2008. *Human gut microbiota and bifidobacteria: From composition to functionality*. Antonie van Leeuwenhoek, Int. J. Gen. Mol. Microbiol. 94:35–50.

4. Kant R. Et al. 2011. *Comparative genomics of Lactobacillus*. Microb. Biotechnol. 4:323–32.

5. Turroni F. et al. 2013. *Molecular dialogue between the human gut microbiota and the host: a Lactobacillus and Bifidobacterium perspective*. Cell. Mol. Life Sci.

6. Lebeer S. et al. 2012. *Functional analysis of Lactobacillus rhamnosus GG pili in relation to adhesion and immunomodulatory interactions with intestinal epithelial cells.* Appl. Environ. Microbiol. 2012 Jan;78(1):185-93.

7. Grompone G, Martorell P, Llopis S. 2012. *Anti-inflammatory Lactobacillus rhamnosus CNCM I-3690 strain protects against oxidative stress and increases lifespan in Caenorhabditis elegans*. PLoS One 2012;7(12):e52493.

8. Forestier C, De Champs C, Vatoux C, Joly B. 2001. *Probiotic activities of Lactobacillus casei rhamnosus: In vitro adherence to intestinal cells and antimicrobial properties*. Res. Microbiol. 152:167–173.

9. Vélez MP. et al. 2010. *Characterization of MabA, a modulator of Lactobacillus rhamnosus GG adhesion and biofilm formation.* FEMS Immunol. Med. Microbiol. 59:386–398.

10. Pessi T, Sütas Y, Hurme M, Isolauri E. 2000. *Interleukin-10 generation in atopic children following oral Lactobacillus* rhamnosus GG. Clinical and experimental allergy: journal of the British Society for Allergy and Clinical Immunology, 2000 Dec;30(12):1804-8.

11. Manzoni P. et al. 2011. *Routine Lactobacillus rhamnosus GG administration in VLBW infants: A retrospective, 6-year cohort study*. Early Hum. Dev. 87:S35–S38.

12. Horvath a., Dziechciarz P, Szajewska H. 2011. *Meta-analysis: Lactobacillus rhamnosus GG for abdominal pain-related functional gastrointestinal disorders in childhood*. Aliment. Pharmacol. Ther. 33:1302–1310.

13. De Keersmaecker SCJ. et al. 2006. *Strong antimicrobial activity of Lactobacillus rhamnosus GG against Salmonella typhimurium is due to accumulation of lactic acid*. FEMS Microbiol. Lett. 259:89–96.

14. Van Baarlen P. et al. 2011. *Human mucosal in vivo transcriptome responses to three lactobacilli indicate how probiotics may modulate human cellular pathways*. Proc. Natl. Acad. Sci. U. S. A. 108 Suppl :4562–4569.

15. Kleerebezem M. et al. 2010. *The extracellular biology of the lactobacilli*. FEMS Microbiol. Rev. 34:199–230.

16. Bron P a, van Baarlen P, Kleerebezem M. 2012. *Emerging molecular insights into the interaction between probiotics and the host intestinal mucosa.* Nat. Rev. Microbiol. 10:66–78.

17. Bochner BR, Gadzinski P, Panomitros E. 2001. *Phenotype microarrays for high-throughput phenotypic testing and assay of gene function.* Genome Res. 11:1246–1255.

18. Tohsato Y, Mori H. 2008. *Phenotype profiling of single gene deletion mutants of E. coli using Biolog technology.* Genome Inform. 21:42–52.

19. Johnson D. et al. 2008. *High-throughput phenotypic characterization of Pseudomonas aeruginosa membrane transport genes*. PLoS Genet. 4:e1000211.

20. Tamang B. et al. 2008. *Phenotypic and genotypic identification of lactic acid bacteria isolated from ethnic fermented bamboo tender shoots of North East India.* Int. J. Food Microbiol. 121:35–40.

21. Lamy B. et al. 2010. *Accuracy of 6 commercial systems for identifying clinical Aeromonas isolates*. Diagn. Microbiol. Infect. Dis. 67:9–14.

22. Di Cagno R. et al. 2007. *Genotypic and phenotypic diversity of Lactobacillus rossiae strains isolated from sourdough.* J. Appl. Microbiol. 103:821–835.

23. Larsen M.V. et al. 2012. *Multilocus sequence typing of total-genome-sequenced bacteria.* J. Clin. Microbiol. 50:1355–61.

24. Douillard FP. et al. 2013. *Comparative genomic and functional analysis of 100 Lactobacillus rhamnosus strains and their comparison with strain GG*. PLoS Genet. 9:e1003683.

25. Broadbent JR. et al. 2012. *Analysis of the Lactobacillus casei supragenome and its influence in species evolution and lifestyle adaptation*. BMC Genomics 13:533.

26. Klein CC. et al. 2012. *Exploration of the core metabolism of symbiotic bacteria*. BMC Genomics 13:438.

27. Salminen MK. et al. 2002. *Lactobacillus bacteremia during a rapid increase in probiotic use of Lactobacillus rhamnosus GG in Finland*. Clin. Infect. Dis. 35:1155–1160.

28. Ricciardi A. et al. 2012. *Genotypic diversity of stress response in Lactobacillus plantarum, Lactobacillus paraplantarum and Lactobacillus pentosus*. Int. J. Food Microbiol. 157:278–85.

29. Molenaar D. et al. 2005. *Exploring Lactobacillus plantarum genome diversity by using microarrays.* J. Bacteriol. 187:6119–6127.

30. Muñoz-Quezada S. et al. 2013. *Competitive inhibition of three novel bacteria isolated from faeces of breast milk-fed infants against selected enteropathogens.* Br. J. Nutr. 109 Suppl :S63–9.

31. Chapman CMC, Gibson GR, Rowland I. 2012. *In vitro evaluation of single- and multi-strain probiotics: Inter-species inhibition between probiotic strains, and inhibition of pathogens*. Anaerobe 18:405–413.

32. Sikorska H, Smoragiewicz W. 2013. *Role of probiotics in the prevention and treatment of meticillin-resistant Staphylococcus aureus infections*. Int. J. Antimicrob. Agents 42:475–81.

33. Riley MA, Wertz JE. 2002. *Bacteriocin diversity: Ecological and evolutionary perspectives*. Biochimie.

34. Makarova KS. et al. 2011. *Evolution and classification of the CRISPR-Cas systems.* Nat. Rev. Microbiol. 9:467–77.

35. Horvath P. et al. 2009. *Comparative analysis of CRISPR loci in lactic acid bacteria genomes*. Int. J. Food Microbiol. 131:62–70.

36. Barrangou R. et al. 2013. *Genomic impact of CRISPR immunization against bacteriophages*. Biochem. Soc. Trans. 41:1383–91.

37. Kankainen M. et al. 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.

38. Meijerink M, Hemert S Van, Taverne N. 2010. *Identification of genetic loci in Lactobacillus plantarum that modulate the immune response of dendritic cells using comparative genome hybridization.* PLoS One 2010 May 13;5(5):e10632.

39. Van Hemert S. et al. 2010. *Identification of Lactobacillus plantarum genes modulating the cytokine response of human peripheral blood mononuclear cells*. BMC Microbiol. 10:293.

40. Goh YJ, Klaenhammer TR. 2013. *A functional glycogen biosynthesis pathway in Lactobacillus acidophilus: expression and analysis of the glg operon.* Mol. Microbiol. 89:1187–200.

41. Rodríguez-Díaz J, Rubio-del-Campo A, Yebra MJ. 2012. *Lactobacillus casei ferments the N-Acetylglucosamine moiety of fucosyl-α-1,3-N-acetylglucosamine and excretes L-fucose*. Appl. Environ. Microbiol. 78:4613–9.

42. Lambert JM, Bongers RS, Kleerebezem M. 2007. *Cre-lox-based system for multiple gene deletions and selectable-marker removal in Lactobacillus plantarum.* Appl. Environ. Microbiol. 73:1126–35.

43. Remus DM. et al. 2012. *Impact of 4 Lactobacillus plantarum capsular polysaccharide clusters on surface glycan composition and host cell signalling.* Microb. Cell Fact. 11:149.

44. Bienenstock J. et al. 2013. *Fucosylated but not sialylated milk oligosaccharides diminish colon motor contractions.* PLoS One 8:e76236.

45. Van Liempt E. et al. 2006. *Specificity of DC-SIGN for mannose- and fucose-containing glycans*. FEBS Lett. 580:6123–6131.

46. Hooper L.V. et al. 1999. *A molecular sensor that allows a gut commensal to control its nutrient foundation in a competitive ecosystem*. Proc. Natl. Acad. Sci. U. S. A. 96:9833–9838.

47. Mendonça AG, Alves RJ, Pereira-Leal JB. 2011. *Loss of genetic redundancy in reductive genome evolution.* PLoS Comput. Biol. 7:e1001082.

48. Bahrami B, Macfarlane S, Macfarlane GT. 2011. *Induction of cytokine formation by human intestinal bacteria in gut epithelial cell lines*. J. Appl. Microbiol. 110:353–363.

49. Ménard S. et al. 2004. *Lactic acid bacteria secrete metabolites retaining anti-inflammatory properties after intestinal transport*. Gut 53:821–828.

50. Bai X. et al. 2010. *IL-32 is a host protective cytokine against Mycobacterium tuberculosis in differentiated THP-1 human macrophages*. J. Immunol. 184:3830–3840.

51. Das A. et al. 2013. *Proteomic changes induced by histone demethylase JMJD3 in TNF alpha-treated human monocytic (THP-1) cells*. Mol. Immunol. 56:113–122.

52. Miksa M. et al. 2009. *A novel method to determine the engulfment of apoptotic cells by macrophages using pHrodo succinimidyl ester*. J. Immunol. Methods 342:71–77.

53. Lebeer S. et al. 2009. *Identification of a Gene Cluster for the Biosynthesis of a Long, Galactose-Rich Exopolysaccharide in Lactobacillus rhamnosus GG and Functional Analysis of the Priming Glycosyltransferase*. Appl. Environ. Microbiol. 75:3554–3563.

54. Lebeer S. et al. 2011. *Exopolysaccharides of Lactobacillus rhamnosus GG form a protective shield against innate immune factors in the intestine*. Microb. Biotechnol. 4:368–374.

55. Bron P. et al. 2012. *Lactobacillus plantarum possesses the capability for wall teichoic acid backbone alditol switching.* Microb. Cell Fact. 11:123.

56. Claes IJJ. et al. 2012. *Genetic and biochemical characterization of the cell wall hydrolase activity of the major secreted protein of Lactobacillus rhamnosus GG*. PLoS One 7:e31588.

57. Bäuerl C, Pérez-Martínez G, Yan F, Polk DB, Monedero V. 2010. *Functional analysis of the p40 and p75 proteins from Lactobacillus casei BL23*. J. Mol. Microbiol. Biotechnol. 19:231–241.

58. Lehtoranta L, Pitkäranta A, Korpela R. 2012. *The effects of probiotic combination Lactobacillus rhamnosus GG and Lactobacillus rhamnosus Lc705 in cytokine and chemokine response in human macrophages*. Int. J. Probiotics Prebiotics 7:17–22.

59. Zhang L, Li N, Caicedo R, Neu J. 2005. *Alive and dead Lactobacillus rhamnosus GG decrease tumor necrosis factor-alpha-induced interleukin-8 production in Caco-2 cells*. J. Nutr. 135:1752–1756.

60. Sanchez B. et al. 2010. *Adhesive properties, extracellular protein production, and metabolism in the Lactobacillus rhamnosus GG strain when grown in the presence of mucin.* J. Microbiol. Biotechnol. 20:978–984.

61. Gagic D, Wen W, Collett M a, Rakonjac J. 2013. *Unique secreted-surface protein complex of Lactobacillus rhamnosus, identified by phage display*. Microbiologyopen 2:1–17.

62. Elitzur AC. 1997. *Constancy, uniformity and symmetry of living systems: the computational functions of morphological invariance.* Biosystems 43:41–53.

63. Levin BR, Cornejo OE. 2009. *The population and evolutionary dynamics of homologous gene recombination in bacteria*. PLoS Genet. 5.

64. 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.

65. 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. 54:2325–40.

66. Bachmann H. et al. 2012. *Microbial domestication signatures of Lactococcus lactis can be reproduced by experimental evolution*. Genome Res. 22:115–124.

67. De Vos WM, Bron P a, Kleerebezem M. 2004. *Post-genomics of lactic acid bacteria and other food-grade bacteria to discover gut functionality*. Curr. Opin. Biotechnol. 15:86–93.

68. Pant N. et al. 2007. *Effective prophylaxis against rotavirus diarrhea using a combination of Lactobacillus rhamnosus GG and antibodies*. BMC Microbiol. 7:86.

69. Martin F-PJ. et al. 2008. *Probiotic modulation of symbiotic gut microbial-host metabolic interactions in a humanized microbiome mouse model.* Mol. Syst. Biol. 4:157.

70. De Champs C. et al. 2003. *Persistence of colonization of intestinal mucosa by a probiotic strain, Lactobacillus casei subsp. rhamnosus Lcr35, after oral consumption*. J. Clin. Microbiol. 41:1270–1273.

71. Van der Kleij H. et al. 2008. *Protective effects of Lactobacillus rhamnosus and Bifidobacterium infantis in murine models for colitis do not involve the vagus nerve.* Am. J. Physiol. Regul. Integr. Comp. Physiol. 295:R1131–R1137.

72. Wickens K. et al. 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.

73. Hawrelak JA, Whitten DL, Myers SP. 2005. *Is Lactobacillus rhamnosus GG effective in preventing the onset of antibiotic-associated diarrhoea: a systematic review*. Digestion 72:51–56.

74. Nermes M. et al. 2011. *Interaction of orally administered Lactobacillus rhamnosus GG with skin and gut microbiota and humoral immunity in infants with atopic dermatitis*. Clin. Exp. Allergy 41:370–377.

75. Sawada J. et al. 2007. *Ingestion of heat-treated Lactobacillus rhamnosus GG prevents development of atopic dermatitis in NC/Nga mice*. Clin. Exp. Allergy 37:296–303.

76. Rose MA. et al. 2010. *Efficacy of probiotic Lactobacillus GG on allergic sensitization and asthma in infants at risk*. Clin. Exp. Allergy 40:1398–1405.

77. Brouwer ML. et al. 2006. *No effects of probiotics on atopic dermatitis in infancy: A randomized placebo-controlled trial*. Clin. Exp. Allergy 36:899–906.

78. Yeşilova Y, Çalka Ö, Akdeniz N, Berktaş M. 2012. *Effect of probiotics on the treatment of children with atopic dermatitis.* Ann. Dermatol. 2012 May;24(2):189-93.

79. Scott JR, Barnett TC. 2006. *Surface proteins of gram-positive bacteria and how they get there.* Annu. Rev. Microbiol. 60:397–423.

80. Kelly D, Conway S, Aminov R. 2005. *Commensal gut bacteria: mechanisms of immune modulation.* Trends Immunol. 2005 Jun;26(6):326-33.

81. Lasa I, Penadés JR. 2006. *Bap: A family of surface proteins involved in biofilm formation*. Res. Microbiol. 2006 Mar;157(2):99-107.

82. Clarke SR, Harris LG, Richards RG, Foster SJ. 2002. *Analysis of Ebh, a 1.1-megadalton cell wall-associated fibronectin-binding protein of Staphylococcus aureus*. Infect. Immun. 70:6680–7.

83. Wu SW, De Lencastre H. 1999. *Mrp--a new auxiliary gene essential for optimal expression of methicillin resistance in Staphylococcus aureus*. Microb. Drug Resist. 5:9–18.

84. Kobe B, Kajava A V. 2001. *The leucine-rich repeat as a protein recognition motif.* Curr. Opin. Struct. Biol. 11:725–732.

85. Kobe B, Deisenhofer J. 1994. *The leucine-rich repeat: a versatile binding motif*. Trends Biochem. Sci. 19:415–21.

86. Bierne H, Sabet C, Personnic N, Cossart P. 2007. *Internalins: a complex family of leucine-rich repeat-containing proteins in Listeria monocytogenes*. Microbes Infect. 2007 Aug;9(10):1156-66.

87. Evdokimov AG. et al. 2001. *Unusual molecular architecture of the Yersinia pestis cytotoxin YopM: a leucine-rich repeat protein with the shortest repeating unit*. J. Mol. Biol. 312:807–821.

88. Ivanov II, Littman DR. 2011. *Modulation of immune homeostasis by commensal bacteria*. Curr. Opin. Microbiol. 2011 Feb;14(1):106-14.

89. Jolly L, Stingele F. 2001. *Molecular organization and functionality of exopolysaccharide gene clusters in lactic acid bacteria.* Int. Dairy J. 11:733–745.

90. Badel S, Bernardi T, Michaud P. 2011. *New perspectives for Lactobacilli exopolysaccharides*. Biotechnol. Adv. 2011 Jan-Feb;29(1):54-66.

91. Duboc P, Mollet B. 2001. *Applications of exopolysaccharides in the dairy industry*, p. 759–768. *In* International Dairy Journal.

92. Tiwari BK. et al. 2009. *Application of natural antimicrobials for food preservation*. J. Agric. Food Chem. 57:5987–6000.

93. Gálvez A, Abriouel H, López RL, Omar N Ben. 2007. *Bacteriocin-based strategies for food biopreservation.* Int. J. Food Microbiol. 120:51–70.

94. Van Hemert S. et al. 2010. *Identification of Lactobacillus plantarum genes modulating the cytokine response of human peripheral blood mononuclear cells*. BMC Microbiol. 10:293.

95. Vitali B. et al. 2010. *Impact of a synbiotic food on the gut microbial ecology and metabolic profiles.* BMC Microbiol. 10:4.

96. Basholli-Salihu M. et al. 2013. *The use of cellobiose and fructooligosaccharide on growth and stability of Bifidobacterium infantis in fermented milk*. Food Nutr. Sci. 04:1301–1306.

97. Herbel SR, Vahjen W, Wieler LH, Guenther S. 2013. *Timely approaches to identify probiotic species of the genus Lactobacillus*. Gut Pathog. 5:27.

98. Coudeyras S, Marchandin H, Fajon C, Forestier C. 2008. *Taxonomic and strain-specific identification of the probiotic strain Lactobacillus rhamnosus 35 within the Lactobacillus casei group.* Appl. Environ. Microbiol. 74:2679–89.

99. Van Baarlen P, Wells JM, Kleerebezem M, Baarlen P Van. 2013. *Regulation of intestinal homeostasis and immunity with probiotic lactobacilli.* Trends Immunol. 34:208–215.

100. Tsilingiri K, Rescigno M. 2012. *Should probiotics be tested on ex vivo organ culture models?* Gut Microbes.

101. Abou Hachem M. et al. 2013. *Recent insight into oligosaccharide uptake and metabolism in probiotic bacteria.* Biocatal. Biotransformation 31:226–235.

102. Hevia A. et al. 2013. *An extracellular Serine/Threonine-rich protein from Lactobacillus plantarum NCIMB 8826 is a novel aggregation-promoting factor with affinity to mucin.* Appl. Environ. Microbiol. 79:6059–66.

103. Baarlen P van, Wells J, Kleerebezem M. 2013. *Regulation of intestinal homeostasis and immunity with probiotic lactobacilli*. Trends Immunol. 4:1286–1294.

104. Tsilingiri K. et al. 2012. *Probiotic and postbiotic activity in health and disease: comparison on a novel polarised ex-vivo organ culture model*. Gut 61:1007–15.

105. Pacheco AR. et al. 2012. *Fucose sensing regulates bacterial intestinal colonization*. Nature 491:113–117.

106. Vanderhaegen B. et al. 2003. *Evolution of Chemical and Sensory Properties during Aging of Top-Fermented Beer*. J. Agric. Food Chem. 51:6782–6790.

107. Van Bokhorst-van de Veen H. et al. 2011. *Short- and long-term adaptation to ethanol stress and its cross-protective consequences in Lactobacillus plantarum*. Appl. Environ. Microbiol. 77:5247–56.

108. Crook N, Alper H. 2012. Classical Strain Improvement, p. 1–33. In Engineering Complex Phenotypes in Industrial Strains.

109. Collado MC, Meriluoto J, Salminen S. 2007. *Role of commercial probiotic strains against human pathogen adhesion to intestinal mucus.* Lett. Appl. Microbiol. 45:454–460.

110. Vanderpool C, Yan F, Polk DB. 2008. *Mechanisms of probiotic action: Implications for therapeutic applications in inflammatory bowel diseases*. Inflamm. Bowel Dis. 14:1585–1596.

111. Schwank G, Andersson-Rolf A, Koo BK, Sasaki N, Clevers H. 2013. *Generation of BAC transgenic epithelial organoids.* PLoS One 8.

112. Sato T. et al. 2011. *Long-term expansion of epithelial organoids from human colon, adenoma, adenocarcinoma, and Barrett's epithelium*. Gastroenterology 141:1762–1772.

113. Buske P. et al. 2012. *On the biomechanics of stem cell niche formation in the gut--modelling growing organoids.* FEBS J. 279:3475–87.

114. Van Bokhorst-van de Veen H. et al. 2012. *Modulation of Lactobacillus plantarum gastrointestinal robustness by fermentation conditions enables identification of bacterial robustness markers*. PLoS One 7:e39053.

115. Bron PA. et al. 2012. *Transcriptomes reveal genetic signatures underlying physiological variations imposed by different fermentation conditions in Lactobacillus plantarum*. PLoS One 7:e38720.

Thesis summary

The use of microbes in the preparation and conservation of human food raw-materials has been a common practice for millennia. For example, the prehistoric Egyptians brewed beer and the Greeks had a high appreciation for wine. Likewise, ancient Japanese miso and shoyu breweries were widespread and visited daily. Likewise, fermented dairy products have made a major and long term contribution to the human diet and have been proposed to extend the lifespan of the Bulgarian population, a proposition that was harnessed by Ilya Ilyich Mechnikov through the isolation of *Lactobacillus delbrueckii subsp. bulgaricus*, the bacterial species used in the production of yoghurt, where it acts in concert with *Streptococcus thermophilus*.

Fermentation plays an important role in our foods, where it can be exploited in the form of fermented foods, but can also lead to food-waste by spoilage of food products. Microbial fermentation provides us with basic food products, like yoghurt, cheese, and bread. However, the mechanisms underlying fermentation was not understood until the groundbreaking discovery of the involvement of microorganisms by Louis Pasteur and Robert Koch in the latter part of the nineteenth century, which provided the foundation for the fermentation research that followed. In modern microbiology, the discovery that single bacterial strains can modulate various functions of the host's intestine has led to the concept of probiotics – live microorganisms that, when administered in adequate amounts, confer a health benefit on the consumer (World Health Organization, 2009). *Bifidobacterium* and *Lactobacillus* are the most well-known probiotic genera and are also among the commensal members of our intestinal microbiota. Various representative species of these genera, like *Bifidobacterium lactis, Lactobacillus plantarum,* and *Lactobacillus rhamnosus*, encompass several strains that are marketed as probiotics. Despite recent advances, we have yet to discover the mechanisms by which these microbes are interacting with our intestine and the diversity of host responses they can elicit.

In order to uncover some of these mechanisms, this thesis explores the diversity of *Lactobacillus rhamnosus*, a species from which strains are studied for their anti-inflammatory, allergy, and diarrhea preventing effects. The work combines observations on the behavior of the bacteria in a simplified laboratory setting (this thesis, **Chapters 3, 4** and **5**) with genomic information obtained by sequencing (**Chapter 6**) with the aim to pinpoint genes that could be relevant for bacterial survival and metabolic capacities.

In **Chapter 3** we analyzed the repertoire of carbohydrates that *L. rhamnosus* strains can use for growth. The 25 strains we investigated cluster into three main metabolic groups, of which two encompass generalists that were predominantly isolated from dairy or plant origin, whereas the third group mainly gathered isolates from the host intestine. In addition, we could discern carbohydrates that could potentially be used in synbiotic concepts (combination of probiotics and non-digestible carbohydrates or prebiotics) based on their selective stimulation of one of the metabolic groups, and/or individual strains. This information can be employed by the food industry in strain selection procedures for different food applications, including health promoting synbiotics, as well as selected fiber degradation systems for organic acid production.

Immune stimulation of the human host as well as other phenotype characteristics of genetically diverse strains were investigated *in vitro*, some of them displaying uniformity within the strain collection (**Chapter 7**). Information concerning bacterial phenotype-uniformity is relevant since it can pinpoint microbial functions that are of importance for the ecosystem-fitness of the species. Conversely, immune modulation by the *L. rhamnosus* strains displayed significant variability when bacteria were co-cultured with blood-derived immune cells, enabling the classification of the strains in four groups with distinct profiles of immune modulation capacity (**Chapter 5**). Immune-profiling is of interest in the light of clinical applications, where it is envisioned that pro-inflammatory strains may boost vaccine responsiveness, while anti-inflammatory strains can serve as

therapeutic tools for intestinal inflammatory diseases.

Comparative genomics of 40 strains of *Lactobacillus rhamnosus* revealed extensive genomic diversity and suggested a role of the CRISPR-Cas system as a genome evolution driver. The variable genes encompassed genes of potential technological relevance (bacteriocins, hydrolases, exopolysaccharides production), including those that are predicted to be involved in the interaction with the host (**Chapter 6**).

Phenotypic and genotypic profiling analyses congruently revealed that carbohydrate metabolism and transport is essential for this species' adaptation to the environment. Genotype–phenotype correlation analysis enabled us to fill some of the gaps in our understanding of gene functionality in this species. Particularly, the impact of carbohydrate utilization has a large impact in species adaptation, and some of the gene functions discovered could be confirmed by engineered expression or deletion of the identified genes. This is illustrated in **Chapter 3**, where the function of genes predicted to be responsible for the utilization of two metabolic-group discriminating carbohydrates, L-Sorbose and α-D-Methyl Glycoside, could be confirmed, and in **Chapter 4**, where the role of the fucosyl kinase encoding gene in the metabolization of L-fucose was established.

Samenvatting

Het gebruik van microorganismen voor de bereiding en conservering van voedsel-grondstoffen is al millenia lang een gewoonte van vele bevolkingen. Bijvoorbeeld, de prehistorische Egyptenaren brouwden bier en de Grieken hadden een grote waardering voor wijn. Ook traditionele Japanse miso en shoyu brouwerijen waren wijd verspreid en werden dagelijks bezocht. Ook gefermenteerde zuivelproducten leveren al millennia lang, een belangrijke bijdrage aan het humane voedingspalet. Aan het begin van de vorige eeuw werd gesuggereerd dat de levensduur van de Bulgaarse bevolking verlengd werd door de consumptive van gefermenteerde zuivelproducten, een suggestive die door Ilja Iljitsj Mechnikov werd uitgebaat door de isolatie van *Lactobacillus delbrueckii subsp. bulgaricus*, de bacterie die gebruikt wordt bij de productie van yoghurt, waarin deze bacterie samenwerkt met *Streptococcus thermophilus*.

Fermentatie speelt een belangrijke rol in onze voedselvoorziening, waar het kan worden uitgebuit bij de productie van gefermenteerde voedingsmiddelen, maar waar het ook kan bijdragen aan verlies van voedingsmiddelen doordat het bederf veroorzaakt. Microbiële fermentatie levert ons basale voedingsmiddelen, zoals yoghurt, kaas en brood. Maar hoe fermentatie eigenlijk werkt werd niet begrepen, tot de baanbrekende ontdekking van de rol van micro-organismen door Louis Pasteur en Robert Koch in het laatste deel van de negentiende eeuw, die de basis verschafte voor het fermentatie onderzoek dat volgde. De ontdekking dat specifieke bacteriestammen verschillende functies van het slijmvlies van de darm van de gastheer kunnen beïnvloeden heeft in de moderne microbiologie geleid tot het concept van probiotica - levende micro-organismen die, indien ze in voldoende hoeveelheden worden toegediend, een voordeel kunnen opleveren voor de gezondheid van de consument (World Health Organization, 2009). *Bifidobacterium* en *Lactobacillus* zijn de meest gebruikte probiotische genera en behoren ook tot de natuurlijke bacteriën van de darmflora. Verschillende representatieve soorten van deze genera, zoals *Bifidobacterium lactis*, *Lactobacillus plantarum* en *Lactobacillus rhamnosus*, omvatten verschillende stammen die worden verkocht als probiotica. Ondanks recente inzichten, moeten we de mechanismen waarmee deze microben een interactie aangaan met onze darm nog ontrafelen, en weten we ook niet hoe variabel de gastheer reacties zijn die ze kunnen opwekken.

Om sommige van deze mechanismen te kunnen ontcijferen, heeft dit proefschrift de diversiteit van *Lactobacillus rhamnosus* onderzocht, een bacteriesoort waarvan

specifieke stammen worden ingezet vanwege hun ontstekingsremmende effecten, hun allergie onderdrukkende effecten of hun capaciteit om diarree te voorkomen. Het werk combineert observaties van het fenotype van de bacteriën onder eenvoudige laboratorium condities (dit proefschrift, **Hoofdstukken 3, 4 en 5**) met genomische informatie die met behulp van sequentiebepaling is verkregen (**Hoofdstuk 6**) en beoogt genen te identificeren die relevant zijn voor het overleven van de bacterie en voor zijn metabolische capaciteiten.

In **Hoofdstuk 3** hebben we het koolhydraten repertoire waarop de *L. rhamnosus* stammen kunnen groeien onderzocht. De 25 bacteriestammen die onderzochten kunnen we indelen in drie metabole groepen, waarvan twee groepen zogenaamde 'generalisten' omvatten die voornamelijk werden geïsoleerd uit zuivel of uit plantaardig materiaal, terwijl de derde groep vooral bestaat uit stammen die uit de darm van de mens of andere zoogdieren werd geïsoleerd. Daarnaast identificeerde deze studie specifieke koolhydraten die gebruikt zouden kunnen worden in synbiotische concepten (combinatie van probiotica en onverteerbare koolhydraten of prebiotica), op basis van hun selectieve stimulatie van de groei van één van de metabole groepen en/of individuele stammen. Deze informatie kan door de voedingsindustrie worden gebruikt in de selectieprocedure van stammen die geschikt zijn voor verschillende voedingstoepassingen, waaronder gezondheids-bevorderende synbiotica, of voor biotechnologische vezelafbraak systemen voor de productie van organische zuren.

Bacteriële stimulatie van het immuunsysteem van de mens werd onderzocht voor een verzameling van genetisch diverse bacterie stammen, waarvoor ook andere fenotypische kenmerken werden bepaald. Sommige eigenschappen bleken uniform aanwezig in alle bacterie stammen die getest werden (**Hoofdstuk 7**), hetgeen informatief is omdat het microbiële functies kan ontdekken die van belang zijn voor de fitheid van de bacteriesoort binnen een ecosysteem. Daarentegen bleken de immuunmodulatie karakteristieken van de *L. rhamnosus* stammen grote variabiliteit te vertonen, wanneer bacteriën samen gekweekt werden met uit het bloed verkregen immuuncellen. Met deze informatie konden de bacterie stammen in vier groepen worden verdeeld, met verschillende immuun systeem stimulerende profielen (**Hoofdstuk 5**). Immune systeem interacties zijn van belang in het licht van klinische toepassingen, waarin het wordt voorgesteld dat pro-inflammatoire stammen de reactie op vaccinatie kunnen verbeteren, terwijl anti-inflammatoire stammen kunnen dienen als therapeutische instrumenten gericht tegen ontstekingsziekten van de darm.

Genomische vergelijking van 40 stammen van *Lactobacillus rhamnosus* toonde een uitgebreide genomische diversiteit aan en suggereerde een sturende rol voor het

CRISPR-Cas-systeem in genoom evolutie. De gene die varieerden in de bacterie stammen omvatten ook genen van mogelijke technologische relevantie (productie van bacteriocines, hydrolase-enzymen en polysachariden), waaronder genen die voorspeld worden een rol te spelen in de interactie met de gastheer (**Hoofdstuk 6**).

Fenotypische en genotypische bepalingen onthulden beide dat variatie in koolhydraat transport en metabolisme een belangrijke vorm van aanpassing aan de omgevingsomstandigheden is in deze bacterie soort. Genotype - fenotype correlatie-analyse maakte het mogelijk om onze kennis met betrekking tot de gen-functionaliteit in deze soort uit te breiden. Met name het gebruik van specifieke koolhydraten speelt een grote rol in het aanpassingsvermogen van de bacterie soort en een aantal van de betrokken gen-functies die werden voorspeld konden worden bevestigd door expressie of deletie van de betreffende genen. Dit wordt geïllustreerd in **Hoofdstuk 3**, waarin de voorspelde functie van genen betrokken bij het gebruik van twee koolhydraten die de metabole-groepen kunnen onderscheiden, L-sorbose en α-D-Methyl Glycoside, bevestigd kon worden. Bovendien bevestigt het werk in **Hoofdstuk 4** de rol van het gen dat voorspeld werd te coderen voor de fucosyl-kinase functie, die een sleutelrol speelt in het metabolisme van L-fucose.

Acknowledgements

They tell me this is the most often read part of the thesis. If so, and if you read this, Dear Reader, please turn your attention also to one (or several) of the chapters, the one whose title catches most your interest, and read it as well, and perhaps send me an email with comments, either good or bad. I will feel like it was worth writing it!

My stay in The Netherlands was an amazing part of my life, full of discoveries, bikes (and bike accidents), joys and sorrows, football through the fields and noisy parties, Romanian, English, Dutch, Spanish, French and others I don't (yet) understand, sometimes all spoken at once, and I have all of you to thank for it!

Firstly, I would like to express my gratitude towards my promotor Michiel for the continuous support of my Ph.D study and related research, for his patience, enthusiasm, and immense knowledge. He constantly and convincingly inspired me in regard to research, keeping me motivated even throughout the most difficult moments. Without his supervision and constant help this dissertation would not have been possible.

Besides my promotor, I would like to thank my thesis co-promotors: Jan and Jolanda, and my managers Richele and Raish, for their insightful observations and encouragement, for keeping my progress on schedule but also for the hard questions which incented me to view my research from various perspectives.

I am particularly grateful for the assistance given by my reading committee members: Prof. Dr. T. Abee, Prof. Dr. R.J. Siezen, Dr. W. van Schaik, Prof. dr. D. van Sinderen. Thank you for taking the time to do the thesis review, for your insightful comments and for making my thesis defense day a special one!

And let me not forget my paranymphs, Loo Wee and Marta, I appreciate enormously your moral support and being my left and right arms when preparing the booklet, defense and party.

My sincere thanks also go to my team in Danone Research (nowadays Danone Nutricia Research): Roger, Kees, Amra, Rob, Harm, Gerben, Tijmen, Jeroen, Claudia. Without their precious support it would not have been possible to conduct this research.

To the Cross Talk EU network (www.**cross-talk.eu**/) that funded part of my research and where I had the opportunity to subject my research to peers scrutiny, many thanks! Your influence shaped my ideas about science and life in general and about host-microbe interactions in particular.

To my fellow labmates from the Microbiology Department, too many to add here, thank you for the stimulating discussions after the MolEco and department presentations and for all the fun we have had in the last six years. Special thanks for my PhD trip co-organizers: you were awesome, brave and efficient; it was a big success and a step in choosing my next endeavor!

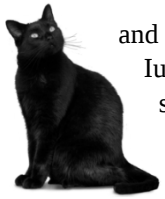
To my workmates from Merck: Danny (also contributor to this thesis by correcting the many mistakes in my Dutch summary), Stefanie, Rik, Loes, Bert, Rezie, and all the MDPS team, I was working with you when most of the (frustrating, time-demanding) writing took place. You guys showed me that you can find a family at your work place and that as a team, you can achieve unsurmountable goals!

To my students: Marie-Jose, Gosia and Loo Wee, you've done great work and made me very proud! Without you, I could not have explored so many facets of this diverse bacterial species.

To my wonderful friends in Wageningen, Nijmegen and Utrecht: Anna, Marta, Miriam and Wiebe, Irene (x3), Paula, Juanan and Naiara, Rocio and Hans, Javi and Teresita, Natalia, Monica, Jaume, Esther, Fanny, Cecilia, Manon, Panam, Amra, Inga, Sérgio, Leo, Veronika and Karel, Ebru, Danny, Loes, Rik, Stéphanie, Maxime, Urmila, Uttara, Sven and Sara, Janneke, Noora and Vicente, Tom, Gerben, Thomas, Monika, Emilio, Cecilia and Martin, Diana, Ale and Pia, Carlos, Femke, Jose, Rupert, Erwin, Ana, Sarah, Aroa, Mark, Ivan and Livia, Liviu and Oana, Eva, Delia and Mugurel, Laurentia, Suki, Arunima, Katerina, Antje … the list is endless… I find myself lucky to have met you and shared so many experiences with you! I'll always remember the beers after long workdays, dinners, game nights, sports events, surprise parties, paintball matches as well as advice, rides to the airport, cat-sittings, and overall an ocean of help and friendship. To my friends scattered around the world, thank you for your thoughts, well-wishes, phone calls, e-mails, texts, visits, and for being there whenever I needed a friend.

To my family away from home, Adi and Alex, and the Romanian group in Wageningen, Smaranda, Cozmina and Rob, Adrian and Luminita, Maria and Viro, Laura and Cătălin, Ionică, thank you from all my heart! Together, we kept the Romanian traditions alive, made new traditions, protested, voted, we've helped each other and others in times of need and been there at all those important moments in life: weddings, baptisms, birthdays, graduations. Thank you for being there, today and always!

To my dear Armando, there are not enough words of gratitude for your love and encouragement. You're my love, my best friend, my artist and scientist, my balance. In the words of Carl Sagan, "in the vastness of space and immensity of time, it is my joy to spend a planet and an epoch with you".



Last but not least, I would like to thank my family: my parents Paula and Florin Ceapa, my nana Livia Fratiman, my sister and brother-in-law, Iulia and Alex Popa and my cousin Arina Fratiman, for supporting me spiritually throughout writing this thesis and my life in general. You give me life and purpose!

About the author

Photo taken by Armando Hernandez

Corina Diana Ceapa was born in Barlad, Romania, on September 10th, 1983. She obtained her Bachelor Degree in **Biochemistry** from the University of Bucharest in 2006, after finishing a Bachelor thesis at the Center for Molecular Biology, under the supervision of Dr. Cristina Munteanu, Dr. Radu Huculeci, and Prof. Conf. Dr. Anca Dinischiotu. The thesis investigated stress proteins expression changes of human cells following manganese (II) and mycotoxin treatment. The author continued this work as a Research Assistant in the same laboratory while earning a Master of Science Degree in **Biochemistry and Molecular Biology** from the University of Bucharest in 2008. She combined her Masters thesis with an Erasmus Fellowship to work at the Structural and Functional Glycobiology Unit in Lille, France under the supervision of Prof. Yann Guerardel, where she investigated the diversity and composition of the glycans involved in the fertilization of the sea pineapple, thus generating the thesis *"Glycosylation profile for the tunicate Halocinthya roretzi. A structural glycobiology study"*.

Thereafter, the author was recruited by the Cross Talk Project, a Marie Curie Initial Training Network (2008-2012), focusing on health and promoting positive interactions between the **microbiota and the human host**, on the project entitled *"Impact of genomic diversity of a probiotic on its immunomodulation effects"*. The project was established as a collaboration between Danone Nutricia Research and Wageningen University, under the supervision of Prof. Dr. Michiel Kleerebezem, Prof. Dr. Jan Knol, Dr. Richele Wind and Dr. Jolanda Lambert and resulted in this thesis.

From June to December 2014, Corina worked as a scientist in the **Analytics** Department in Merck, Oss, The Netherlands.

Currently, the author is employed by the Laboratory of Microbiology, University of Chicago, USA, where she is studying the host receptors for the main *Bacillus anthracis* adhesin, under the supervision of Prof. Dominique Missiakas.

List of publications

From this thesis:

1. *Influence of fermented milk products, prebiotics and probiotics on microbiota composition and health*, **Ceapa C.**, Wopereis H., Rezaïki L, Kleerebezem M, Knol J, Oozeer R., Best Practice & Research Clinical Gastroenterology, Volume 27, Issue 1, February 2013, Pages 139–155.
2. *Correlation of Lactobacillus rhamnosus genotypes and carbohydrate utilization signatures determined by phenotype profiling*, **Corina Ceapa**, Jolanda Lambert, Kees van Limpt, Michiel Wels, Tamara Smokvina, Jan Knol, and Michiel Kleerebezem, Appl. Environ. Microbiol., August 2015, 81:16, 5458-5470.
3. *L-fucose metabolization in Lactobacillus rhamnosus: identification and analysis of the operon*, **Corina Ceapa**, Roger S. Bongers, Loo Wee Chia, Jolanda M. Lambert, Jan Knol, Michiel Kleerebezem, submitted manuscript.
4. *Genetically diverse strains of Lactobacillus rhamnosus display fermentation dependent strain-specific immune modulation effects*, **Corina Ceapa**, Jeroen van Bergenhenegouwen, Jolanda Lambert, Michiel Kleerebezem, Jan Knol, submitted manuscript.
5. *Variable regions of Lactobacillus rhamnosus genomes suggest strong strains segregation by horizontal gene transfer events*, **Corina Ceapa**, Mark Davids, Jarmo Ritari, Jolanda Lambert, Michiel Wels, François P. Douillard, Tamara Smokvina, Willem M. de Vos, Jan Knol, Michiel Kleerebezem, submitted manuscript.
6. Provisional patent application: The use of *L. rhamnosus* strains for pathogen inhibition, **Corina Ceapa**, Kees van Limpt, Jan Knol.

Others:

1. *The effect of deoxynivalenol on hepatic cell line HepG2*, G. Dragomir, R. Huculeci, C. Munteanu, **C. Ceapa**, L. Bodea, A. Dinischiotu, Biotechnology in Animal Husbandry 23 (5-6-1), 583-588.
2. *Adapted response of the antioxidant defense system to oxidative stress induced by deoxynivalenol in Hek-293 cells*, Dinu D, Bodea GO, **Ceapa CD**, Toxicon, Volume 57, Issues 7–8, June 2011, Pages 1023–1032.

Overview of completed training activities

Discipline specific activities

Cross Talk - Kick Off Meeting, Paris, France (2009)

NIZO Bioinformatics Seminar, NIZO, Ede, The Netherlands (2009)

Genetics & physiology of food associated microorganisms, VLAG, Wageningen, The Netherlands (2010)

Systems Biology: Statistical analysis of ~omics data, VLAG, Wageningen, The Netherlands (2010)

1st MetaHIT Conference on Human Metagenomics, Shenzhen, China (2010)

13th Gut Day Symposium, Wageningen, The Netherlands (2011)

Lessons from host-pathogens interactions, Cross Talk, Milan, Italy (2011)

Network meeting, Cross Talk, Wageningen, The Netherlands (2011)

Comparative Genomics: from evolution to function, NBIC, Utrecht, The Netherlands (2011)

Functional Metagenomics of the Intestinal Tract and Food-Related Microbes, Helsinki, Finland (2011)

3d Beneficial Microbes Conference, TNO, Noordwijkerhout, the Netherlands (2012)

EU - US Bioinformatics Training, Jacobs University, Bremen, Germany (2012)

Host-microbes cross-talk: from animal models to human patients, Cross Talk, Oslo, Norway (2012)

Laboratory Animal Science, Utrecht University, Utrecht, The Netherlands (2012)

Cross Talk Final Symposium, Paris, France (2013)

General activities

VLAG Ph.D. Week, Graduate School VLAG, Wageningen, The Netherlands (2008)

Presentation skills, Impact and Danone Research, Wageningen, The Netherlands (2011)

General skills development (intellectual property, project management, network building, science communication, presentation skills, careers in industry and academia), Cross Talk, Aberdeen, England (2009)

General skills development (team building, bio-pharma products, scientific posters and resilience among PhD students), Cross Talk, Debrecen, Hungary (2010)

Scientific publishing, WGS (2013)

Dutch Language Course, 140 hours, European Language Centre, Wageningen, The Netherlands (2009-2010)

VMT Biosafety, Fontys, Nijmegen, The Netherlands (2009)

Leadership training, Danone Research (2012)

Techniques for Writing and Presenting a Scientific Paper, WIMEK & SENSE (2012)

Last Stretch of the PhD Programme, WGS (2014)

Optionals

Preparing PhD proposal, VLAG (2009)

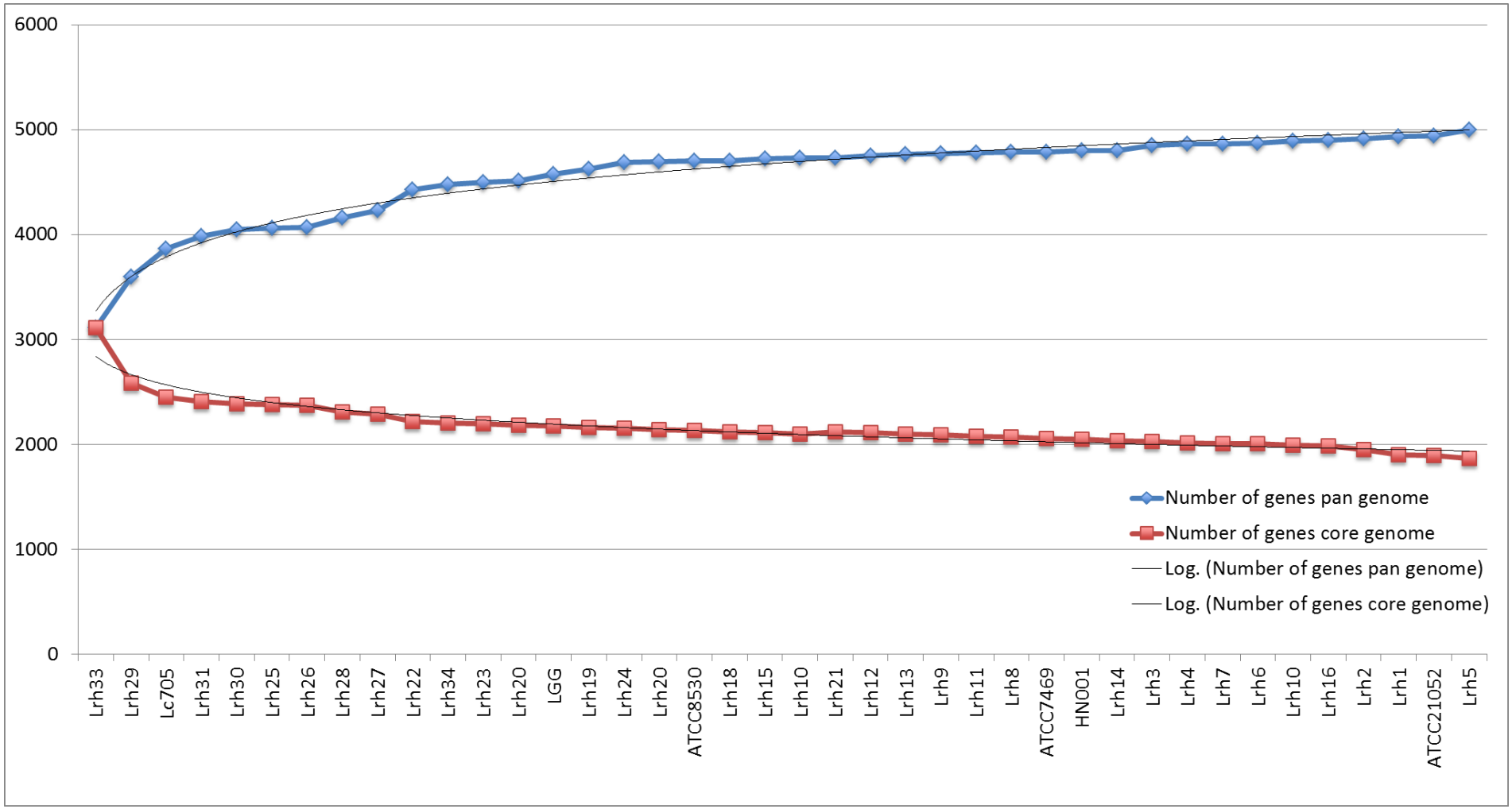
PhD/PostDoc meetings, Laboratory of Microbiology, WUR, Wageningen, The Netherlands (2009/2014)

Cross Talk Newsletter – writer, editor (2009/2012)

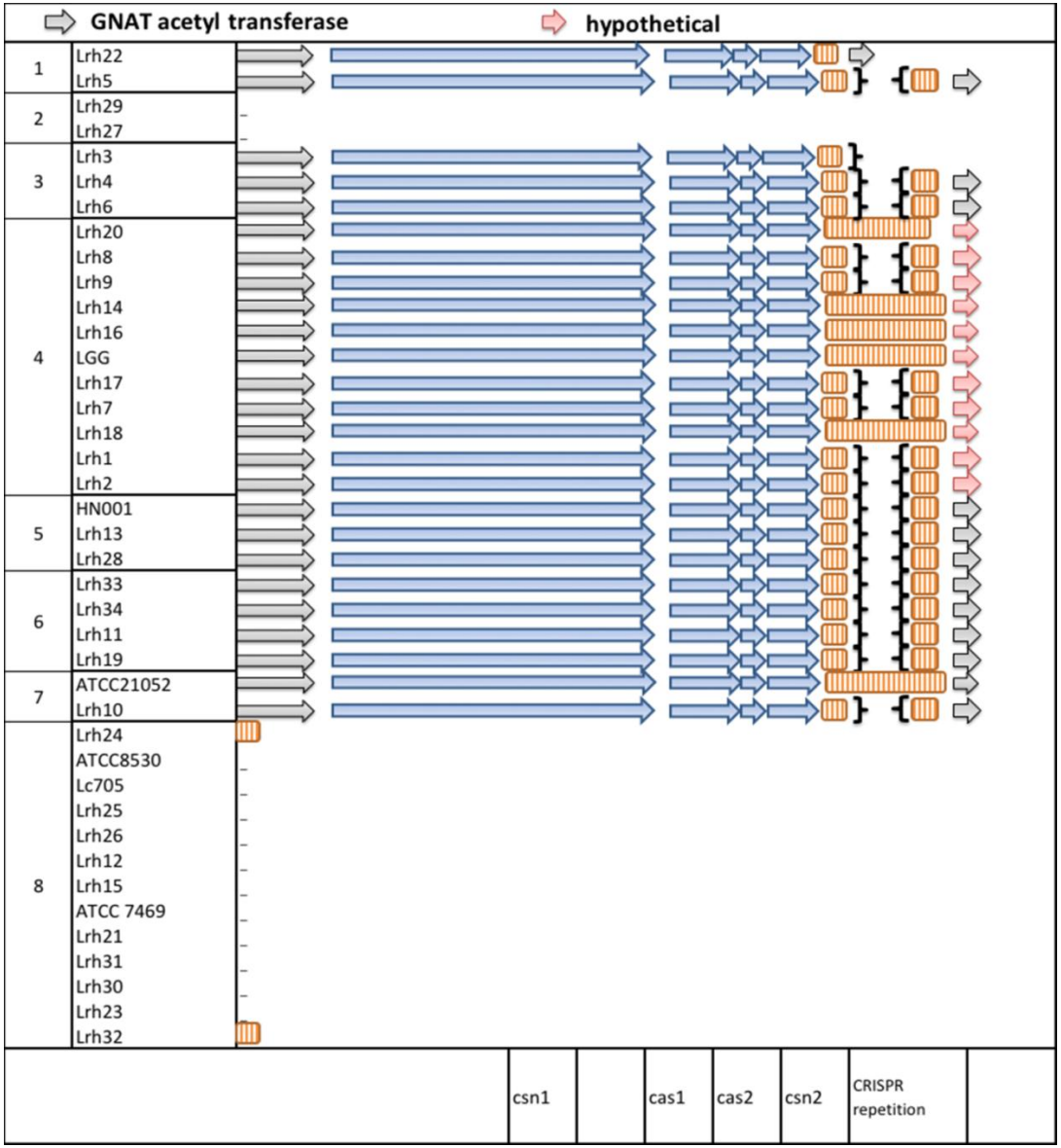
Organizer and participant for the PhD trip of MIB and SSB: USA and Canada (2013)

Research presented in this thesis was supported by a Marie Curie International Training Mobility Network (Cross-Talk, grant agreement No: 21553-2) and by Danone Nutricia Research, Utrecht, The Netherlands.

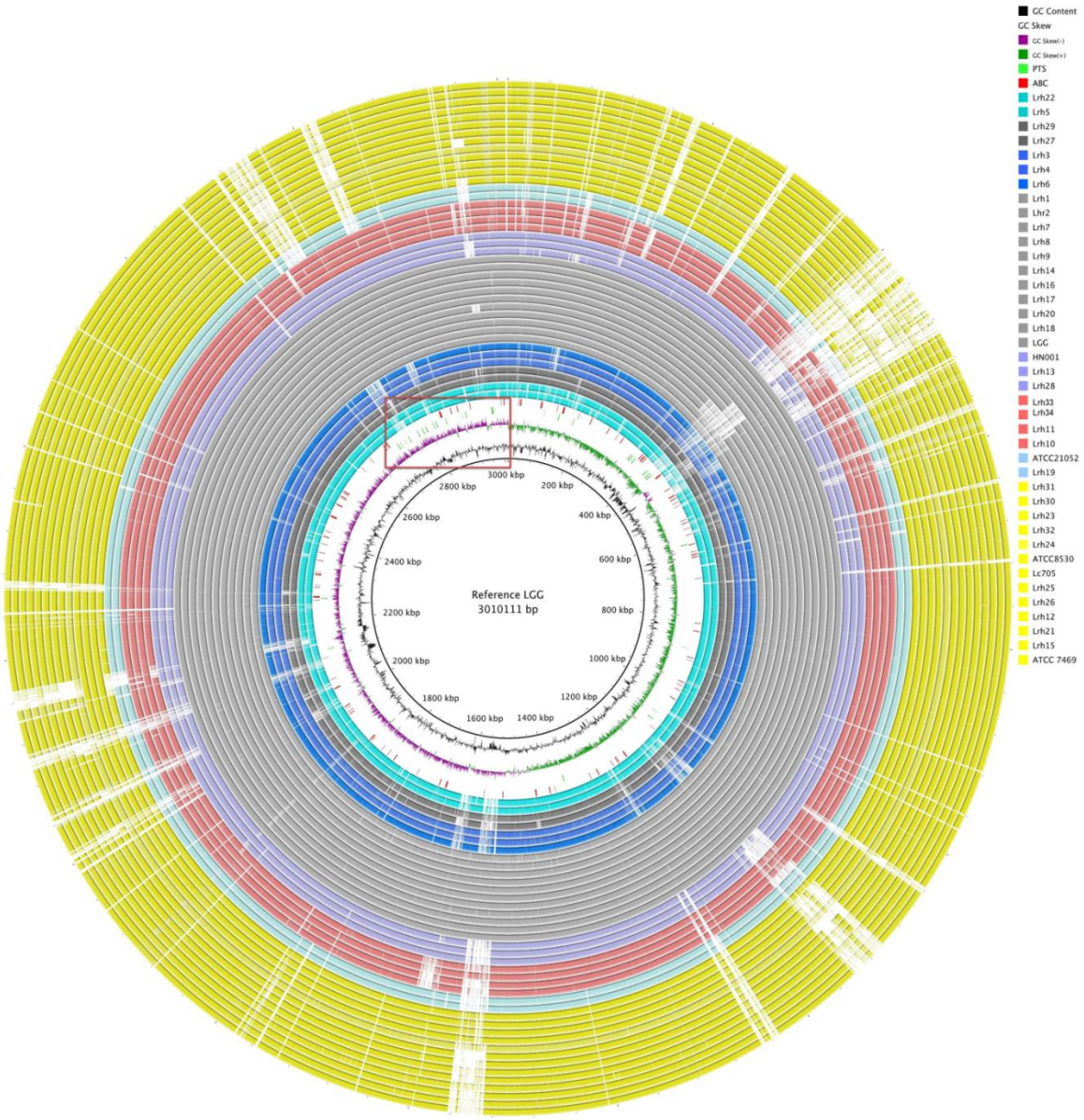
Supplementary Figure 1. Core/ pangenome OG saturation.



Supplementary Figure 2. Cas genes/ CRISPR architecture in *L. rhamnosus* strains.



Supplementary Figure 3. Genomes alignment to strain GG. Highlighted region contains an enrichment of carbohydrate transport PTS genes.

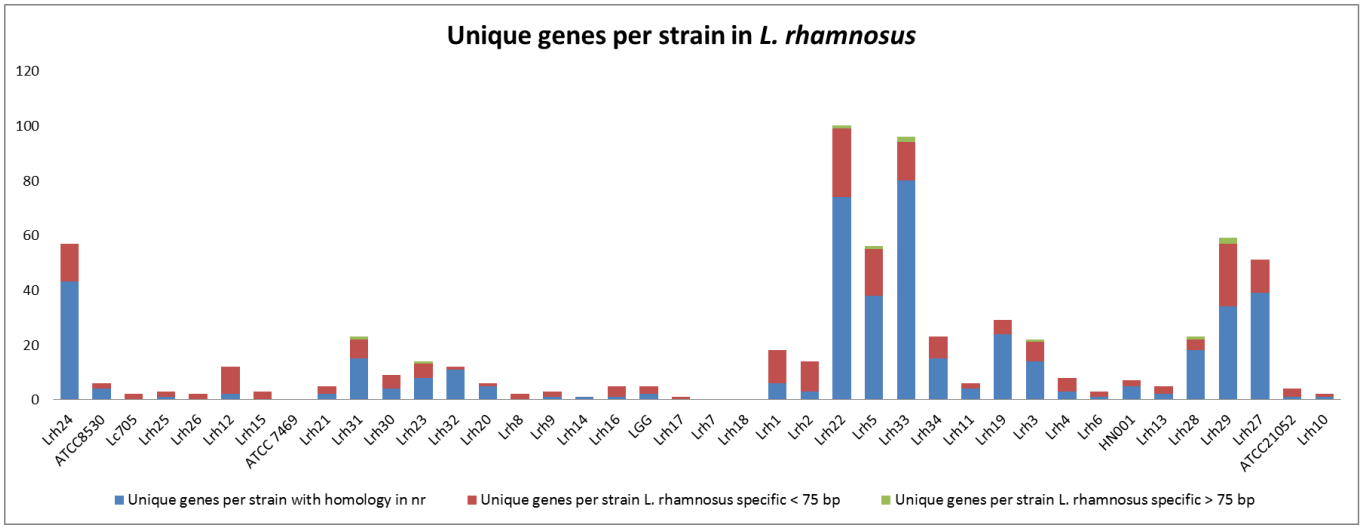


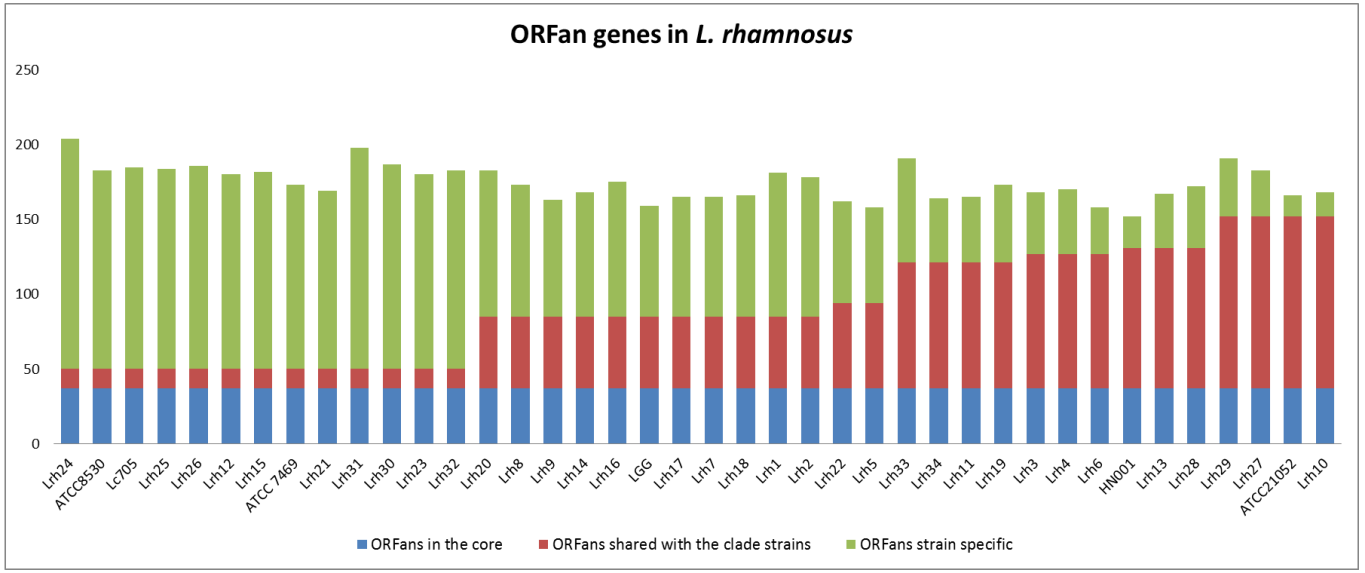
Supplementary Table.1. A general overview of the origin and genome statistics of the 40 Lactobacillus *rhamnosus* genomes used in the study: size, GC content, number of genes, niche of isolation, sequencing parameters and previous literature. The gene locuses names used for GG genome are based on the original locus tag published by Kankainen et al. and not the newly assigned tags by NCBI. The original names are still there within the Genbank files and are called old locus tag.

Strain	Length (Mb	G+C (%)	ORFs	Isolated from	Contigs	Coverage	LPXTG gene	Reference	NCBI accesion number	Bioproject ID	Biosample ID
ATCC21052	2.87	46.7	2909	healthy intestine	1	20	19	**Reference genome for the Human Microbiome Project**			
ATCC7469	2.93	46.7	2934	clinical, type strain	81	17	21	**this work**			
ATCC8530	2.96	46.8	2933	beer	1	20	21	[35]			
Lrh33	3.12	46.4	3163	fermented dairy product	261	13	21	**this work**	JTIX00000000	PRJNA253894	SAMN03196620
Lrh34	3	46.5	3006	fermented dairy product	136	13	22	**this work**	JTIW00000000	PRJNA253894	SAMN03196621
HN001	2.91	46.6	2973	fermented dairy product	96	36	17				
Lc705	3.03	46.6	2881	fermented dairy product	1	25	23	[21]			
LGG	3.01	46.7	2926	healthy intestine	1	25	22	**[21, 36]**			
Lrh1	2.79	46.6	2779	human feces	535	9	19	**this work**	JTIV00000000	PRJNA253894	SAMN03196625
Lrh10	2.88	46.7	2898	fermented dairy product	91	15	19	**this work**	JTIU00000000	PRJNA253894	SAMN03196618
Lrh11	2.89	46.6	2923	blood	112	10	20	**this work**	JTIT00000000	PRJNA253894	SAMN03196637
Lrh12	2.89	46.6	2884	human feces	384	10	21	**this work**	JTIS00000000	PRJNA253894	SAMN03196628
Lrh13	2.9	46.7	2891	blood	209	9	22	**this work**	JTIR00000000	PRJNA253894	SAMN03196647
Lrh14	2.9	46.6	2883	blood	260	9	17	**this work**	JTIQ00000000	PRJNA253894	SAMN03196644
Lrh15	2.92	46.6	2923	blood	147	12	21	**this work**	JTIP00000000	PRJNA253894	SAMN03196633
Lrh16	2.92	46.6	2901	clinical	145	15	23	**this work**	JTIO00000000	PRJNA253894	SAMN03196622
Lrh17	2.92	46.6	2895	human feces	127	16	23	**this work**	JTIN00000000	PRJNA253894	SAMN03196627
Lrh18	2.93	46.6	2918	blood	222	9	22	**this work**	JTIM00000000	PRJNA253894	SAMN03196638
Lrh19	2.93	46.6	2933	healthy intestine	99	12	25	**this work**	JTIL00000000	PRJNA253894	SAMN03196650
Lrh2	2.8	46.6	2793	human feces	401	9	18	**this work**	JTIK00000000	PRJNA253894	SAMN03196626
Lrh20	2.93	46.6	2929	blood	102	16	21	**this work**	JTIJ00000000	PRJNA253894	SAMN03196645
Lrh21	2.95	46.6	2945	human feces	85	22	21	**this work**	JTII00000000	PRJNA253894	SAMN03196623
Lrh22	2.96	46.6	2988	blood	195	9	19	**this work**	JTIH00000000	PRJNA253894	SAMN03196648
Lrh23	2.96	46.6	2971	blood	61	15	21	**this work**	JTIG00000000	PRJNA253894	SAMN03196649
Lrh24	2.98	46.6	3008	animal (goat) feces	345	11	23	**this work**	JTIF00000000	PRJNA253894	SAMN03196631
Lrh25	2.98	46.6	2994	fermented dairy product	132	15	23	**this work**	JTIE00000000	PRJNA253894	SAMN03196619
Lrh26	2.99	46.6	3004	human feces	112	16	23	**this work**	JTID00000000	PRJNA253894	SAMN03196629
Lrh27	3	46.5	3061	human feces	240	14	20	**this work**	JTIC00000000	PRJNA253894	SAMN03196624
Lrh28	3	46.6	2982	blood	112	19	21	**this work**	JTIB00000000	PRJNA253894	SAMN03196643
Lrh29	3.01	46.5	3082	blood	437	11	20	**this work**	JTIA00000000	PRJNA253894	SAMN03196632
Lrh3	2.82	46.7	2856	fermented dairy product	362	10	17	**this work**	JTHZ00000000	PRJNA253894	SAMN03196636
Lrh30	3.01	46.6	3010	blood	96	12	21	**this work**	JTHY00000000	PRJNA253894	SAMN03196639
Lrh31	3.02	46.5	3061	vagina	124	10	25	**this work**	JTHX00000000	PRJNA253894	SAMN03196651
Lrh32	2.95	46.6	2980	blood	64	18	21	**this work**	JTHW00000000	PRJNA253894	SAMN03196642
Lrh4	2.83	46.8	2852	blood	287	11	17	**this work**	JTHV00000000	PRJNA253894	SAMN03196634
Lrh5	2.84	46.7	2845	human feces	161	12	18	**this work**	JTHU00000000	PRJNA253894	SAMN03196630
Lrh6	2.85	46.8	2876	blood	173	14	17	**this work**	JTHT00000000	PRJNA253894	SAMN03196635
Lrh7	2.86	46.7	2851	blood	86	17	17	**this work**	JTHS00000000	PRJNA253894	SAMN03196646
Lrh8	2.86	46.7	2875	blood	86	17	18	**this work**	JTHR00000000	PRJNA253894	SAMN03196640
Lrh9	2.88	46.7	2864	blood	112	13	18	**this work**	JTHQ00000000	PRJNA253894	SAMN03196641

Supplementary Table.2. Presence absence map for all orthologous groups (OGs) present in the L. rhamnosus strains. Fields representing number of genomes the protein is present in size, closest hit in the NCBI database, ATCC 53103 gene number from the public databases and whether the genes contains an LPXTG motif are added to the table. Genomes order is based on the genetic clades. (see separate pdf file)

Supplementary Table.3. Unique and ORF-an genes of the *L. rhamnosus* strains.





Supplementary Table 4. Clade specific OGs.

Clade 1	Clade 2	Clade 3	Clade 4	Clade 5	Clade 6	Clade 7	Clade 8
orth4194; SortaseA1,_surface_protein_transpeptidase	orth4328; replication_protein_B	orth2719; hypothetical_protein	orth3689; transcription_regulator	orth1054; hypothetical_protein	orth509; hypothetical_protein	orth4465; Nicotinate_phosphoribosyltransferase_(EC 2.4.2.11)	orth2099; hypothetical
orth4106; D-lactate_dehydrogenase_(EC_1.1.1.28)/EC_number=1.1.1.28	orth2472; putative_type_II_restriction_enzyme_methylasesubunit	orth2740; hypothetical_protein	orth3177; hypothetical_protein	orth1109; hypothetical_protein	orth577; FIG036446:_hypothetical_protein	orth4702; hypothetical_protein	orth2371; Stage_V_sporulation_protein_whose_disruptionleads_to_the_production_of_immature_spores_(SpoVK)
orth4093; Malate_permease	orth2522; hypothetical_protein	orth2767; hypothetical_protein	orth3174; hypothetical_protein	orth4008; hypothetical_protein	orth595; hypothetical_protein	orth1304; hypothetical_protein	orth2372; FIG00742155:_hypothetical_protein
orth358; hypothetical_protein	orth2523; hypothetical_protein	orth2787; hypothetical_protein	orth619; hypothetical_protein	orth4014; Type_I_restriction-modification_system,restriction_subunit_R_(EC_3.1.21.3)	orth1544; Nucleotidyltransferase_(EC_2.7.7.-)	orth4629; hypothetical_protein	orth2637; FIG00754890:_hypothetical_protein
orth4276; nucleoid_DNA-binding_protein	orth4007; FIG00753307:_hypothetical_protein	orth2823; hypothetical_protein	orth916; putative_bacteriophage_related_protein	orth4016; hypothetical_protein	orth1581; Putative_uncharacterized_protein	orth4665; hypothetical_protein	orth3243; hypothetical_protein
orth4076; Uncharacterized_protein,_YUKC_B.subtilis homolog	orth4015; putative_cytoplasmic_protein	orth2846; Undecaprenyl-phosphategalactosephosphotransferase(_EC:2.7.8.6_)	orth885; hypothetical_protein	orth4025; hypothetical_protein	orth1648; hypothetical_protein	orth4667; hypothetical_protein	orth3866; conserved_domain_protein

orth4236; hypothetical_protein	orth4021; putative_type_II_restriction_enzyme_methylasesubunit	orth2941; Acetyltransferase,_GNAT_family	orth38; ATP-dependent_DNA_helicase_UvrD/PcrA	orth4054; hypothetical_protein	orth1686; hypothetical_protein	orth3984; hypothetical_protein	orth780; Beta-glucoside_bgl_operon_antiterminator,_BglGfamily
orth4011; hypothetical_protein	orth4037; conserved_hypothetical_protein,_probablyfragment	orth2733; FIG00745703:_hypothetical_protein	orth3670; hypothetical_protein	orth4069; Type_I_restriction-modification_system,DNA-methyltransferase_subunit_M_(EC_2.1.1.72)	orth1699; hypothetical_protein	orth3978; hypothetical_protein	orth3225; PTS_system,_beta-glucoside-specific_IIBcomponent_(EC_2.7.1.69)_/_PTS_system,
orth4329; ESAT-6/Esx_family_secreted_protein_EsxA/YukE	orth4051; hypothetical_protein	orth4055; polysaccharide_polymerase	orth1177; hypothetical_protein	orth4079; Type_I_restriction-modification_system,specificity_subunit_S_(EC_3.1.21.3)	orth1731; hypothetical_protein	orth1246; unknown	orth2170; GO:0008706
orth4112; Aggregation_promoting_factor	orth4070; CpsX_protein	orth4078; Relaxase/Mobilisation_nuclease_domain	orth660; hypothetical_protein	orth4105; hypothetical_protein	orth1732; hypothetical_protein	orth3974; phage_anti-repressor_protein	orth781; Phosphoenolpyruvate-dependent_sugarphosphotransferase_system,_EIIA_2
orth4128; Zeta_toxin	orth4100; putative_acetyl_transferase	orth4082; Glycosyltransferase	orth3524; hypothetical_protein	orth4171; hypothetical	orth1762; hypothetical_protein	orth1245; hypothetical_protein	orth3223; PTS_system,_maltose_and_glucose-specific_IICcomponent_(EC_2.7.1.69)_/_PTS_system,_maltose_and

orth4337; hypothetical_protein	orth4117; hypothetical_protein	orth4089; Lipopolysaccharide_1,6-galactosyltransferase(EC_2.4.1.-)	orth3539; hypothetical_protein	orth4174; hypothetical_protein	orth1802; hypothetical_protein	orth3966; hypothetical_protein	orth3832; Cellobiose_phosphotransferase_system_YdjC-likeprotein
orth4282; hypothetical_protein	orth4127; ParA_family_protein	orth4097; Glycosyltransferase	orth606; hypothetical_protein	orth4240; hypothetical	orth1757; hypothetical_protein	orth3964; Modification_methylase_EcoRV_(EC_2.1.1.72)(Adenine-specific_methyltransferase_EcoRV)_(M.EcoRV)	orth3067; Transcriptional_regulator,_RpiR_family
orth4165; putative_plasmid_partition_protein	orth4140; DnaD_domain_protein	orth4179; hypothetical_protein	orth62; hypothetical_protein	orth4247; hypothetical_protein	orth543; hypothetical_protein	orth3963; Chain_A,_Ecorv_Restriction_EndonucleaeC-Terminal_Deletion_Mutant GATATCCA2+	orth2362; hypothetical_protein
orth4376; hypothetical_protein	orth4146; hypothetical_protein	orth4183; truncated_RepA	orth1194; hypothetical_protein	orth4254; hypothetical_protein	orth554; hypothetical_protein	orth3962; hypothetical_protein	orth3870; hypothetical_protein
orth4389; hypothetical_protein	orth4160; hypothetical_protein	orth4197; hypothetical_protein	orth1907; hypothetical_protein	orth4274; hypothetical_protein	orth566; Galactofuranose_transferase	orth3961; Modification_methylase	orth3048; hypothetical_protein
orth4296; hypothetical_protein	orth4172; hypothetical_protein	orth4293; mobilization_protein	orth3129; hypothetical_protein	orth4326; Type_I_restriction-modification_system,DNA-methyltransferase_subunit_M_(EC_2.1.1.72)	orth567; hypothetical_protein	orth3946; hypothetical_protein	orth3741; hypothetical_protein

orth4283; hypothetical_protein	orth4191; hypothetical_protein	orth4299; N-acetylmuramic_acid_6-phosphate_etherase_(EC4.2.-.-)	orth1169; hypothetical_protein	orth4327; HsdR	orth571; hypothetical_protein	orth1250; hypothetical_protein	orth2685; hypothetical_protein
orth4061; hypothetical_Cytosolic_Protein	orth4210; hypothetical_protein	orth4303; Transcriptional_regulator,_xre_family	orth10; hypothetical_protein	orth4330; hypothetical	orth599; hypothetical_protein	orth1274; hypothetical_protein	orth2003; hypothetical_protein
orth4049; FtsK/SpoIIIE_family_protein,_putative_secretionsystem_component_EssC/YukA	orth4213; FIG00745257:_hypothetical_protein	orth4314; killer_suppression_protein_HigA,_putative	orth130; hypothetical_protein	orth4341; mobile_element_protein	orth1559; hypothetical_protein	orth1297; hypothetical_protein	orth1350; hypothetical_protein
orth4516; hypothetical_protein	orth4214; hypothetical_protein	orth4375; hypothetical_protein	orth4384; hypothetical_protein	orth4349; Type_I_restriction-modification_system,DNA-methyltransferase_subunit_M_(EC_2.1.1.72)	orth1723; hypothetical_protein	orth3932; hypothetical_protein	orth725; hypothetical_protein
orth4605; hypothetical_protein	orth4217; putative_scaffolding_protein	orth4424; hypothetical_protein	orth636; hypothetical_protein	orth4351; hypothetical_protein	orth1735; hypothetical_protein	orth3920; hypothetical_protein	orth2356; hypothetical_protein
orth4511; YafQ_toxin_protein	orth4224; hypothetical_protein	orth4441; hypothetical_protein	orth798; hypothetical_protein	orth4388; phage_Mu_protein_F_like_protein	orth1788; Glycosyltransferase	orth3917; hypothetical_protein	orth3276; Type_I_restriction-modification_system,specificity_subunit_S_(EC_3.1.21.3)

orth4517; hypothetical_protein	orth4225; FIG00754149:_hypothetical_protein	orth4454; Glycyl-tRNA_synthetase_alpha_chain_(EC6.1.1.14)	orth911; hypothetical_protein	orth4405; Replication	orth1790; Possible_glycosyltransferase	orth3915; hypothetical_protein	orth2186; hypothetical
orth4338; hypothetical_protein	orth4235; hypothetical_protein	orth4455; hypothetical_protein	orth1973; hypothetical_protein	orth4435; unnamed_protein_product	orth1795; hypothetical_protein	orth1255; hypothetical_protein	orth2676; hypothetical_protein
orth4494; hypothetical_protein	orth4241; hypothetical_protein	orth4460; hypothetical_protein	orth3580; hypothetical_protein	orth4461; hypothetical_protein	orth565; Putative_mannosyltransferase_involved_inpolysaccharide_biosynthesis	orth1270; hypothetical_protein	orth2665; FIG00754910:_hypothetical_protein
orth4366; hypothetical_protein	orth4243; major_capsid_protein_gpP	orth4470; hypothetical_protein	orth3588; hypothetical_protein	orth4467; hypothetical_protein	orth1787; putative_polysaccharide_polymerase_protein		orth1340; hypothetical_protein
orth4090; hypothetical_protein	orth4256; hypothetical_protein	orth4481; copy-number_control_protein	orth3590; hypothetical_protein	orth4483; hypothetical_protein	orth1789; Lipopolysaccharide_1,6-galactosyltransferase(EC_2.4.1.-)		orth3009; hypothetical_protein
orth4043; hypothetical_protein	orth4257; hypothetical_protein	orth4498; hypothetical_protein	orth3593; hypothetical_protein	orth4484; replication	orth529; phage_terminase_large_subunit		orth4013; phage_tail_length_tape-measure_protein
orth4453; hypothetical_protein	orth4263; hypothetical_protein	orth4506; hypothetical_protein	orth3709; hypothetical_protein	orth4485; hypothetical_protein	orth4017; Surface_antigen_p40		orth3263; hypothetical_protein

orth4144; hypothetical_protein	orth4268; orf15	orth4540; hypothetical_protein	orth3712; hypothetical_protein	orth4539; Type_I_restriction-modification_system,DNA-methyltransferase_subunit_M_(EC_2.1.1.72)	orth4041; hypothetical_protein		orth2677; putative_antirepressor_protein
orth4630; hypothetical_protein	orth4272; hypothetical_protein	orth4544; hypothetical_protein	orth4151; hypothetical_protein	orth4608; hypothetical_protein	orth4347; hypothetical_protein		orth1033; hypothetical_protein
orth4346; hypothetical_protein	orth4273; minor_capsid_protein	orth4546; hypothetical_protein	orth4221; hypothetical_protein	orth4609; hypothetical_protein	orth4369; hypothetical_protein		orth2659; hypothetical_protein
orth4149; hypothetical_protein	orth4275; hypothetical_protein	orth4565; hypothetical_protein	orth4271; Chaperone_protein_DnaJ	orth4656; hypothetical	orth4397; hypothetical_protein		orth2660; transcriptional_regulator,_Cro/CI_family
orth4139; hypothetical_protein	orth4277; ORF42	orth4604; hypothetical_protein	orth4333; transport_protein	orth4662; hypothetical_protein	orth4406; hypothetical_protein		orth3197; hypothetical_protein
orth4447; hypothetical_protein	orth4288; hypothetical_protein	orth4620; hypothetical_protein	orth4345; Alpha-galactosidase_(EC_3.2.1.22)/EC_number=3.2.1.22	orth4681; hypothetical	orth4545; hypothetical_protein		orth3210; hypothetical_protein
orth4674; hypothetical_protein	orth4294; hypothetical_protein	orth4672; hypothetical_protein	orth4373; FIG00750939:_hypothetical_protein	orth4692; hypothetical_protein	orth1446; hypothetical_protein		orth3241; FIG045374:_Type_II_restriction_enzyme,methylase_subunit_YeeA
orth4132; hypothetical_protein	orth4297; hypothetical_protein	orth4679; hypothetical_protein	orth4383; hypothetical_protein	orth1070; hypothetical_protein	orth4006; FIG00745879:_hypothetical_protein		orth3271; hypothetical_protein

orth4502; hypothetical_protein	orth4311; mobilization_protein	orth4710; hypothetical_protein	orth4398; FIG00743351:_hypothetical_protein	orth1072; hypothetical_protein	orth4012; glycosyl_hydrolase_53 (_EC:3.2.1.89_)		orth3272; hypothetical_protein
orth4400; hypothetical_protein	orth4321; hypothetical_protein	orth4711; hypothetical_protein	orth4399; hypothetical_protein	orth1094; hypothetical_protein	orth4018; FIG00629370:_hypothetical_protein		orth3347; FIG006126:_DNA_helicase,restriction/modification_system_component_YeeB
orth4238; hypothetical_protein	orth4340; hypothetical_protein	orth2695; hypothetical_protein	orth4420; hypothetical_protein	orth173; hypothetical_protein	orth4019; virulence-associated_E		orth3348; YeeC-like_protein
orth4177; hypothetical_protein	orth4350; BS_yobM_related_protein	orth2702; hypothetical_protein	orth4425; hypothetical_protein	orth1820; hypothetical	orth4020; relaxase_Mob_DEI		orth3742; FIG131328:_Predicted_ATP-dependent_endonucleaseof_the_OLD_family
orth4663; hypothetical_protein	orth4352; hypothetical_protein	orth2736; hypothetical_protein	orth4429; hypothetical_protein	orth1862; hypothetical	orth4024; hypothetical_protein		orth2681; site-specific_recombinase,_phage_integrasefamily
orth4129; Hypothetical_similar_to_thiamin_biosynthesislipoprotein_ApbE	orth4353; hypothetical_protein	orth2778; Family_13_glycosyl_hydrolase,_row_724	orth4440; hypothetical_protein	orth1866; hypothetical	orth4029; Restriction_enzyme_BgcI_alpha_subunit_(EC3.1.21.-)_[Includes:_Adenine-specific_methyltransferase		orth3837; hypothetical_protein

orth4232; FIG00750704:_hypothetical_protein	orth4356; hypothetical_protein	orth2781; hypothetical_protein	orth4442; hypothetical_protein	orth1889; GO:0004322	orth4030; hypothetical_protein		orth3052; hypothetical_protein
orth4190; Triosephosphate_isomerase_(EC_5.3.1.1)/EC_number=5.3.1.1	orth4362; hypothetical_protein	orth2796; hypothetical_protein	orth4443; hypothetical_protein	orth218; hypothetical_protein	orth4032; putative_transposase		orth2183; hypothetical
orth4142; Fructose-bisphosphate_aldolase_class_II_(EC4.1.2.13)	orth4365; hypothetical_protein	orth2815; hypothetical_protein	orth4444; hypothetical_protein	orth249; hypothetical_protein	orth4034; Arsenical_pump-driving_ATPase_(EC_3.6.3.16)/EC_number=3.6.3.16		orth3040; hypothetical_protein
orth4266; hypothetical_protein	orth4372; Alpha-galactosidase_(EC_3.2.1.22)/EC_number=3.2.1.22	orth2818; hypothetical_protein	orth4449; ATP-dependent_nuclease,_subunit_B		orth4036; hypothetical_protein		orth2187; FIG00754149:
orth4046; Transcriptional_antiterminator_of_lichenanoperon,_BglG_family	orth4378; hypothetical_protein	orth2821; hypothetical_protein	orth4451; hypothetical_protein		orth4039; Membrane_protein_involved_in_the_export_ofO-antigen,_teichoic_acid_lipoteichoic_acids		orth3807; hypothetical_protein
orth4113; 4-hydroxythreonine-4-phosphate_dehydrogenase(EC_1.1.1.262)	orth4381; Type_I_restriction-modification_system,specificity_subunit_S_(EC_3.1.21.3)	orth2833; hypothetical_protein	orth4459; hypothetical_protein		orth4045; bacillolysin(_EC:3.4.24.27_)		orth2188; hypothetical

orth4066; hypothetical_protein	orth4385; hypothetical_protein	orth2863; HNH_homing_endonuclease	orth4462; hypothetical_protein		orth4047; hypothetical_protein		orth3042; hypothetical_protein
orth4108; 2-keto-3-deoxygluconate_permease_(KDGpermease)	orth4386; hypothetical_protein	orth2875; Putative_uncharacterized_protein	orth4469; hypothetical_protein		orth4052; membrane_protein,_putative		orth328; Manganese-dependent_inorganic_pyrophosphatase(EC_3.6.1.1)
orth4033; hypothetical_protein	orth4390; hypothetical_protein	orth2884; hypothetical_protein	orth4486; hypothetical_protein		orth4053; FIG00749762:_hypothetical_protein		orth330; CrcB_protein
orth4336; hypothetical_protein	orth4396; hypothetical_protein	orth2905; collagen_triple_helix_repeat_domain_protein	orth4501; hypothetical_protein		orth4056; Signal_transduction_histidine_kinase		orth763; hypothetical_protein
orth4302; hypothetical_protein	orth4401; Ig_domain_protein,_group_2_domain_protein	orth2906; Putative_exporter_protein	orth4503; hypothetical_protein		orth4060; DNA-cytosine_methyltransferase_(EC_2.1.1.37)/EC_number=2.1.1.37		orth788; hypothetical_protein
orth4157; hypothetical_protein	orth4404; hypothetical_protein	orth2918; hypothetical_protein	orth4513; hypothetical_protein		orth4062; Arsenic_efflux_pump_protein		orth948; FIG00747635:_hypothetical_protein
orth4428; hypothetical_protein	orth4409; hypothetical_protein	orth2933; Possible_glucoamylase_(diverged),_15_family	orth4515; hypothetical_protein		orth4067; FIG00744946:_hypothetical_protein		orth949; hypothetical_protein
orth4267; hypothetical_protein	orth4415; hypothetical_protein	orth2977; hypothetical_protein	orth4522; hypothetical_protein		orth4072; Beta-lactamase_class_C_and_other_penicillinbinding_proteins		orth950; hypothetical_protein

orth4668; hypothetical_protein	orth4419; hypothetical_protein	orth2978; Death-on-curing_protein	orth4527; hypothetical_protein		orth4073; Integrase,_superantigen-encoding_pathogenicityislands_SaPI		orth951; transcriptional_regulator,_Cro/CI_family
orth4426; hypothetical_protein	orth4427; hypothetical_protein	orth2982; hypothetical_protein	orth4549; hypothetical_protein		orth4074; transposase_(partial)		orth952; hypothetical_protein
orth4618; hypothetical_protein	orth4430; hypothetical_protein	orth2990; Sugar_transporter	orth4552; Single-stranded-DNA-specific_exonuclease_RecJ(EC_3.1.-.-)		orth4081; FIG00750967:_hypothetical_protein		orth1021; hypothetical_protein
orth4643; hypothetical_protein	orth4446; PrgO-like_protein_for_plasmid_replication	orth2991; Transcriptional_regulator	orth4555; hypothetical_protein		orth4086; Glycosyltransferase		orth2350; hypothetical_protein
orth4521; hypothetical_protein	orth4464; hypothetical_protein	orth2992; hypothetical_protein	orth4569; hypothetical_protein		orth4091; FIG00712184:_hypothetical_protein		orth3261; hypothetical_protein
orth4542; hypothetical_protein	orth4480; IMP_cyclohydrolase_(EC_3.5.4.10)_/Phosphoribosylaminoimidazolecarboxamide_formyltransferase		orth4573; hypothetical_protein		orth4092; hypothetical_protein		orth3341; hypothetical_protein

orth4615; hypothetical_protein	orth4492; hypothetical_protein		orth4575; hypothetical_protein		orth4095; dTDP-glucose_4,6-dehydratase_(EC_4.2.1.46)/EC_number=4.2.1.46		orth2094; hypothetical
orth4148; hypothetical_protein	orth4496; hypothetical_protein		orth4590; Transcriptional_regulator		orth4107; Ornithine_cyclodeaminase_(EC_4.3.1.12)/EC_number=4.3.1.12		orth2097; hypothetical
orth4135; hypothetical_protein	orth4509; hypothetical_protein		orth4592; hypothetical_protein		orth4116; LtrC		orth2126; hypothetical
orth4085; hypothetical_protein	orth4519; hypothetical_protein		orth4603; hypothetical_protein		orth4119; Beta-1,3-glucosyltransferase		orth2180; hypothetical
orth4038; FIG00513666:_hypothetical_protein	orth4520; hypothetical_protein		orth4607; hypothetical_protein		orth4121; Glycosyltransferase		orth1338; hypothetical_protein
orth4479; hypothetical_protein	orth4524; hypothetical_protein		orth4616; hypothetical_protein		orth4138; hypothetical_protein		orth1346; hypothetical_protein
orth4071; Tn916,_transcriptional_regulator,_putative	orth4532; hypothetical_protein		orth4617; hypothetical_protein		orth4141; FIG00745596:_hypothetical_protein		orth283; hypothetical_protein
orth4305; hypothetical_protein	orth4538; hypothetical_protein		orth4623; hypothetical_protein		orth4147; hypothetical_protein		orth284; hypothetical_protein
orth4239; Hypothetical_membrane_protein	orth4551; hypothetical_protein		orth4624; hypothetical_protein		orth4150; Tn5252,_Orf23		orth313; hypothetical_protein
orth4368; hypothetical_protein	orth4561; hypothetical_protein		orth4635; hypothetical_protein		orth4153; Helicase_loader_DnaI		orth336; hypothetical_protein

orth4651; hypothetical_protein	orth4563; hypothetical_protein		orth4639; hypothetical_protein		orth4155; Glycosyl_transferase,_family_2		orth706; hypothetical_protein
orth4587; hypothetical_protein	orth4566; FIG00751638:_hypothetical_protein		orth4657; hypothetical_protein		orth4166; Hypothetical_membrane_spanning_protein		orth731; hypothetical_protein
orth4614; hypothetical_protein	orth4567; hypothetical_protein		orth4666; hypothetical_protein		orth4168; hypothetical_protein		orth779; hypothetical_protein
orth4126; hypothetical_protein	orth4571; hypothetical_protein		orth4671; hypothetical_protein		orth4173; FIG00751682:_hypothetical_protein		orth787; hypothetical_protein
orth4463; hypothetical_protein	orth4589; hypothetical_protein		orth4682; hypothetical_protein		orth4180; hypothetical_cytosolic_protein		orth2368; integral_membrane_protein_(putative)
orth4445; hypothetical_protein	orth4594; hypothetical_protein		orth4684; hypothetical_protein		orth4186; 2-C-methyl-D-erythritol_4-phosphatecytidylyltransferase_(EC_2.7.7.60)		orth2401; hypothetical_protein
orth4695; hypothetical_protein	orth4595; hypothetical_protein		orth4688; hypothetical_protein		orth4188; FIG00742740:_hypothetical_protein		orth3059; hypothetical_protein
orth4478; hypothetical_protein	orth4601; hypothetical_protein		orth4701; hypothetical_protein		orth4189; response_regulator		orth3076; hypothetical_protein
orth4159; Integrase	orth4606; hypothetical_protein		orth4703; hypothetical_protein		orth4196; Membrane-associated_phospholipid_phosphatase		orth3230; hypothetical_protein
orth4402; Integrase	orth4625; hypothetical_protein		orth4707; hypothetical_protein		orth4200; hypothetical_protein		orth3720; hypothetical_protein

orth4198; hypothetical_protein	orth4627; hypothetical_protein		orth4437; PTS_system,_cellobiose-specific_IIA_component(EC_2.7.1.69)		orth4204; membrane_protein		orth3723; hypothetical_protein
orth4264; hypothetical_protein	orth4633; hypothetical_protein		orth2; hypothetical_protein		orth4207; DedA_protein		orth3781; hypothetical_protein
orth4583; hypothetical_protein	orth4634; hypothetical_protein		orth3094; Putative_superfamily_I_DNA_helicases		orth4209; major_tail_protein,_phi13_family		orth3824; hypothetical_protein
orth4697; hypothetical_protein	orth4637; hypothetical_protein		orth58; hypothetical_protein		orth4211; hypothetical_protein		orth3825; hypothetical_protein
orth4152; transposase,_IS4	orth4640; hypothetical_protein		orth455; hypothetical_protein		orth4215; putative_resolvase		orth4472; hypothetical
orth4370; hypothetical_protein	orth4646; hypothetical_protein		orth348; Integrase,_superantigen-encoding_pathogenicityislands_SaPI		orth4216; hypothetical_protein		orth4644; hypothetical
orth4413; hypothetical_protein	orth4648; hypothetical_protein		orth2036; FIG00753262:_hypothetical_protein		orth4223; hypothetical_protein		orth4562; hypothetical_protein
orth4325; hypothetical_protein	orth4652; hypothetical_protein		orth3665; hypothetical_protein		orth4227; FIG00748045:_hypothetical_protein		orth4300; PapB
orth4596; hypothetical_protein	orth4659; hypothetical_protein		orth100; hypothetical_protein		orth4231; hypothetical_protein		orth4316; Thioredoxin
orth4628; hypothetical_protein	orth4664; Putative_uncharacterized_protein		orth129; phage_tail_fibers		orth4237; hypothetical_protein		orth4342; hypothetical_protein,_nickel_resistancedeterminant

orth4114; Sucrose_operon_repressor_ScrR,_LacI_family	orth4669; hypothetical_protein		orth836; phage_tail_fibers		orth4244; hypothetical_protein		orth4412; hypothetical_protein
orth4584; hypothetical_protein	orth4670; hypothetical_protein		orth26; hypothetical_protein		orth4245; Teichoic_acid_glycosylation_protein		orth4632; hypothetical_protein
orth4022; hypothetical_protein	orth4676; hypothetical_protein		orth3638; hypothetical_protein		orth4258; transcriptional_regulator,_XRE_family		orth4124; truncated_RepA
orth4023; hypothetical_protein	orth4680; hypothetical_protein		orth3940; ABC_transporter,_ATP-binding_protein		orth4261; Putative_uncharacterized_protein		orth4169; hypothetical_protein
orth4058; hypothetical_protein	orth4687; hypothetical_protein		orth39; FIG00746435:_hypothetical_protein		orth4269; hypothetical_protein		orth4579; hypothetical_protein
orth4080; hypothetical_protein	orth4689; hypothetical_protein		orth3558; Lyzozyme_M1_(1,4-beta-N-acetylmuramidase)		orth4279; Hypothetical_protein		orth4164; putative_plasmid_partition_protein
orth4102; predicted_ATP-dependent_endonuclease,_OLDfamily	orth4690; hypothetical_protein		orth2490; hypothetical_protein		orth4286; Putative_uncharacterized_protein		orth4203; hypothetical_protein
orth4104; acyltransferase/acetyltransferase	orth4691; hypothetical_protein		orth3099; hypothetical_protein		orth4298; hypothetical_protein		orth4230; Alpha/beta_superfamily_hydrolase

orth4115; Orf55	orth4005; phage_tail_length_tape-measure_protein		orth139; hypothetical_protein		orth4304; Arsenical_resistance_operon_trans-actingrepressor_ArsD		orth4250; hypothetical_protein
orth4125; FIG00518155:_hypothetical_protein	orth4042; phage_minor_capsid_protein		orth502; Undecaprenyl-phosphategalactosephosphotransferase_(EC_2.7.8.6)		orth4306; Hypothetical_protein		orth4322; hypothetical_protein
orth4137; hypothetical_protein	orth4084; phage_minor_capsid_protein		orth901; hypothetical_protein		orth4307; Putative_uncharacterized_protein		orth4332; DNA-damage-inducible_protein_J
orth4156; minor_tail_protein	orth4088; Prophage_Lp2_protein_6		orth43; predicted_ORF		orth4308; Non-specific_DNA-binding_protein_Dps_/ Iron-binding_ferritin-like_antioxidant_protein_/		orth4411; hypothetical_protein
orth4158; hypothetical_protein	orth4110; phage_major_capsid_protein		orth119; hypothetical_protein		orth4313; Arsenical_resistance_operon_repressor		orth315; DNA-binding_response_regulator,_OmpR_family
orth4167; hypothetical_protein_within_pathogenicityisland	orth4154; phage_replication_protein_#_ACLAME_208		orth800; Type_III_restriction-modification_systemmethylation_subunit_(EC_2.1.1.72)		orth4319; hypothetical_protein		orth4027; FIG045374:_Type_II_restriction_enzyme,methylase_subunit_YeeA

orth4178; hypothetical_protein	orth4318; Hypothetical_prophage_lsa1_protein		orth930; Type_III_restriction-modification_system_StyLTIenzyme_res_(EC_3.1.21.5)		orth4320; hypothetical_protein		orth4040; Transposase_subunit
orth4185; hypothetical_protein	orth4331; phage_transcriptional_regulator,_Cro/CI_family		orth2867; unknown		orth4324; Transposase,_IS30_family		orth4048; Membrane_protein_involved_in_the_export_ofO-antigen,_teichoic_acid_lipoteichoic_acids
orth4192; Hypothetical_SAV2027_homolog_insuperantigen-encoding_pathogenicity_islands_SaPI	orth718; hypothetical_protein		orth50; ThiJ/PfpI_family		orth4334; FIG00751019:_hypothetical_protein		orth4057; hypothetical_protein
orth4205; conserved_hypothetical_protein	orth402; hypothetical_protein		orth1156; FIG00746973:_hypothetical_protein		orth4335; Predicted_cell-wall-anchored_protein_SasC(LPXTG_motif)		orth4064; FIG01282627:_hypothetical_protein
orth4208; predicted_ATP-dependent_endonuclease,_OLDfamily	orth2607; ABC_transporter_ATP-binding_protein		orth29; FIG00745049:_hypothetical_protein		orth4339; hypothetical_protein		orth4077; truncated_transposase_for_IS1272

orth4222; hypothetical_protein	orth2311; hypothetical_protein		orth1962; FIG00749233:_hypothetical_protein		orth4348; Modification_methylase_EcoRI_(EC_2.1.1.72)/EC_number=2.1.1.72		orth4083; Glycosyl_transferase,_group_1
orth4251; hypothetical_protein	orth3444; hypothetical_protein		orth3420; hypothetical_protein		orth4355; Putative_uncharacterized_protein		orth4087; hypothetical_protein
orth4252; hypothetical_protein	orth3910; Transcriptional_regulator,_ArsR_family		orth3562; hypothetical_protein		orth4371; bovine_pathogenicity_island_protein_Orf19		orth4094; transcriptional_regulator,_Cro/CI_family
orth4255; conserved_hypothetical_protein	orth1240; Transcriptional_regulator,_ArsR_family		orth3417; hypothetical_protein		orth4374; hypothetical_protein		orth4096; hypothetical_protein
orth4284; putative_head_to_tail_joining	orth3415; putative_MDR_permease;_possible_multidrugefflux_pump		orth3147; putative_amino_acid_ABC_transporter,ATP-binding_protein		orth4379; hypothetical_protein		orth4098; hypothetical_protein
orth4290; hypothetical_protein	orth1261; hypothetical_protein		orth3146; putative_amino_acid_ABC_transporter,periplasmic_amino_acid-binding_protein		orth4382; FIG00754370:_hypothetical_protein		orth4101; hypothetical_protein
orth4291; ORF042	orth3418; hypothetical_protein		orth28; putative_amino_acid_ABC_transporter,_permeaseprotein		orth4391; hypothetical_protein		orth4109; putative_glycosyltransferase

orth4315; hypothetical_protein	orth1514; Putative_uncharacterized_protein		orth3597; hypothetical_protein		orth4392; hypothetical_protein		orth4111; Beta-1,3-glucosyltransferase
orth4344; hypothetical_protein	orth2453; putative_methyltransferase		orth1974; mobile_element_protein		orth4394; hypothetical_protein		orth4118; protein_of_unknown_function_DUF23
orth4354; hypothetical_protein	orth1436; phage_terminase,_small_subunit		orth1968; hypothetical_protein		orth4395; hypothetical_protein		orth4120; putative_glycosyltransferase
orth4361; hypothetical_protein	orth2524; DNA_adenine_methyltransferase,phage-associated		orth633; hypothetical_protein		orth4403; FIG00751021:_hypothetical_protein		orth4122; Alpha-L-Rha_alpha-1,3-L-rhamnosyltransferase(EC_2.4.1.-)
orth4364; Plasmid_maintenance_system_killer_protein	orth2059; 2-dehydropantoate_2-reductase_(EC_1.1.1.169)/EC_number=1.1.1.169		orth103; ABC_transporter,_ATP-binding_protein		orth4407; COG0477:_Permeases_of_the_major_facilitatorsuperfamily		orth4123; Lysophospholipase_(EC_3.1.1.5);_Monoglyceridelipase_(EC_3.1.1.23);_putative
orth4377; hypothetical_protein	orth2078; hypothetical_protein		orth856; hypothetical_protein		orth4422; hypothetical_protein		orth4133; Glycosyl_transferase,_group_2_family_protein;dTDP-rhamnosyl_transferase_RfbF_(EC_2.-.-.-)
orth4387; hypothetical_protein	orth2482; ORF052		orth1135; Galactofuranose_transferase		orth4431; hypothetical_protein		orth4134; hypothetical_protein

orth4439; hypothetical_protein	orth2553; hypothetical_protein		orth70; Alpha-L-Rha_alpha-1,3-L-rhamnosyltransferase(EC_2.4.1.-)		orth4432; hypothetical_protein		orth4136; ATPase_involved_in_chromosome_partitioning
orth4458; hypothetical_protein	orth3380; ORF039		orth3130; Glycosyltransferase_(EC_2.4.1.-)		orth4436; Transposase		orth4145; transposase
orth4474; hypothetical_protein	orth3389; Integrase		orth4317; Membrane_protein_involved_in_the_export_of O-antigen,_teichoic_acid_lipoteichoic_acids		orth4448; hypothetical_protein		orth4162; FIG045374:_Type_II_restriction_enzyme,methylase_subunit_YeeA
orth4491; hypothetical_protein	orth2238; hypothetical_protein		orth857; hypothetical_protein		orth4468; hypothetical_protein		orth4163; glycosyltransferase_family_2
orth4493; hypothetical_protein	orth2545; hypothetical_protein		orth1963; Inososse_isomerase_(EC_5.3.99.-)		orth4473; hypothetical_protein		orth4170; IstB_domain_protein_ATP-binding_protein
orth4526; hypothetical_protein	orth3378; hypothetical_protein		orth3688; 2-deoxy-D-gluconate_3-dehydrogenase_(EC1.1.1.125)		orth4476; Putative_uncharacterized_protein		orth4181; Oxygen-insensitive_NADPH_nitroreductase_(EC1.-.-.-)
orth4533; hypothetical_protein	orth2564; hypothetical_protein		orth3634; Shikimate/quinate_5-dehydrogenase_I_beta_(EC1.1.1.282)		orth4477; hypothetical_protein		orth4187; putative_transposase

orth4536; hypothetical_protein	orth2450; hypothetical_protein		orth1964; transport_protein		orth4482; hypothetical_protein		orth4195; Lipid_carrier_:UDP-N-acetylgalactosaminyltransferase_(EC_2.4.1.-)
orth4543; hypothetical_protein	orth2463; hypothetical_protein		orth1965; Shikimate/quinate_5-dehydrogenase_I_beta_(EC1.1.1.282)		orth4488; hypothetical_protein		orth4199; putative_mobilization_protein
orth4553; hypothetical_protein	orth2464; hypothetical_protein		orth630; Regulator_of_polyketide_synthase_expression		orth4497; hypothetical_protein		orth4201; hypothetical_protein
orth4554; hypothetical_protein	orth2468; hypothetical_protein		orth825; Short-chain_dehydrogenase/reductase_SDR		orth4499; hypothetical_protein		orth4212; hypothetical_protein
orth4568; hypothetical_protein	orth2483; hypothetical_protein		orth826; Acetoacetate_decarboxylase		orth4504; hypothetical_protein		orth4229; Transposase
orth4599; hypothetical_protein	orth2491; hypothetical_protein		orth3602; hypothetical_protein		orth4505; hypothetical_protein		orth4234; hypothetical_protein
orth4647; hypothetical_protein	orth2492; TolA_protein		orth41; hypothetical_protein		orth4507; hypothetical_protein		orth4248; hypothetical_protein
orth4649; hypothetical_protein	orth2493; Structural_feature(s)_predicted_byPsort:Transmembrane:_3056018_-_3056034;_Transmembrane:		orth3152; hypothetical_protein		orth4518; FIG00745584:_hypothetical_protein		orth4260; FIG00632156:_hypothetical_protein

orth4650; hypothetical_protein	orth2494; hypothetical_protein		orth635; hypothetical_protein		orth4537; hypothetical_protein		orth4285; FIG00743773:_hypothetical_protein
orth4653; hypothetical_protein	orth2496; hypothetical_protein		orth3679; hypothetical_protein		orth4548; hypothetical_protein		orth4292; hypothetical_protein
orth4685; hypothetical_protein	orth2505; hypothetical_protein		orth3131; capsular_polysaccharide_biosynthesis_protein		orth4557; hypothetical_protein		orth4295; hypothetical_protein
orth4059; PTS_system,_galactitol-specific_IIC_component(EC_2.7.1.69)	orth2510; Phosphate_butyryltransferase_(EC_2.3.1.19)/EC_number=2.3.1.19		orth610; putative_transcriptional_regulatorphage-related		orth4559; HNH_homing_endonuclease		orth4301; ATP-dependent_protease_(EC_3.4.21.-)
orth4242; PTS_system,_galactitol-specific_IIA_component(EC_2.7.1.69)	orth2511; Dihydrolipoamide_dehydrogenase_ofbranched-chain_alpha-keto_acid_dehydrogenase_(EC				orth4564; hypothetical_protein		orth4309; hypothetical_protein
orth4249; phage_terminase,_small_subunit	orth2537; hypothetical_protein				orth4572; hypothetical_protein		orth4310; ORF-11
orth4028; PTS_system,_sucrose-specific_IIB_component_(EC2.7.1.69)_/_PTS_system,_sucrose-specific_IIC_component	orth2541; Sorbitol_dehydrogenase_(EC_1.1.1.14)/EC_number=1.1.1.14				orth4578; hypothetical_protein		orth4323; hypothetical_protein

orth4202; ABC_transporter_ATP-binding_protein	orth2543; Glycerol_uptake_operon_antiterminatorregulatory_protein				orth4580; hypothetical_protein		orth4357; hypothetical_protein
orth4009; phage_tail_length_tape-measure_protein	orth2546; hypothetical_protein				orth4581; hypothetical_protein		orth4358; hypothetical_protein
orth4044; phage_portal	orth2563; hypothetical_protein				orth4597; hypothetical_protein		orth4360; hypothetical_protein
orth4063; phage_terminase,_large_subunit	orth2579; hypothetical_protein				orth4598; FIG00749217:_hypothetical_protein		orth4363; hypothetical_protein
orth4130; phage_major_capsid_protein	orth2582; hypothetical_protein				orth4600; hypothetical_protein		orth4367; hypothetical_protein
orth4226; phage_capsid_and_scaffold	orth2586; hypothetical_protein				orth4610; hypothetical_protein		orth4393; hypothetical_protein
orth734; DNA-cytosine_methyltransferase_(EC_2.1.1.37)/EC_number=2.1.1.37	orth2589; hypothetical_protein				orth4613; hypothetical_protein		orth4410; hypothetical_protein
orth387; hypothetical_protein	orth2594; membrane_protein				orth4619; hypothetical_protein		orth4414; hypothetical_protein
orth1264; hypothetical_protein	orth2599; putative_dehydrogenase				orth4631; hypothetical_protein		orth4417; hypothetical_protein
orth422; hypothetical_protein	orth2600; transcription_regulator				orth4638; hypothetical_protein		orth4418; hypothetical_protein

orth492; Positive_transcriptional_regulator,_MutRfamily	orth2603; hypothetical_protein				orth4654; FIG00745443:_hypothetical_protein		orth4423; hypothetical_protein
orth402; hypothetical_protein	orth3371; putative_minor_capsid_protein				orth4655; hypothetical_protein		orth4433; hypothetical_protein
orth2607; ABC_transporter_ATP-binding_protein	orth3372; Terminase_large_subunit				orth4661; hypothetical_protein		orth4438; hypothetical_protein
orth491; hypothetical_protein	orth3373; hypothetical_protein				orth4673; hypothetical_protein		orth4450; hypothetical_protein
orth2266; SOS-response_repressor_and_protease_LexA_(EC3.4.21.88)	orth3374; hypothetical_protein				orth4683; hypothetical_protein		orth4452; hypothetical_protein
orth1612; FIG00754149:_hypothetical_protein	orth3376; hypothetical_protein				orth4694; hypothetical_protein		orth4456; hypothetical_protein
orth446; Transcriptional_regulator,_xre_family	orth3377; hypothetical_protein				orth4699; hypothetical_protein		orth4457; hypothetical_protein
orth1449; FIG00748875:_hypothetical_protein	orth3379; Methyltransferase				orth4259; N-acetylglucosamine/galactosamine_PTS,_EIIA(EC:2.7.1.69_)		orth4475; hypothetical_protein
orth349; glycoside_hydrolase,_family_28	orth3381; hypothetical_protein				orth4206; YbbL_ABC_transporter_ATP-binding_protein		orth4487; hypothetical_protein

orth2245; FIG00742107:_hypothetical_protein	orth3382; hypothetical_protein				orth4161; mobile_element_protein		orth4489; hypothetical_protein
orth2225; L-arabinose_isomerase_(EC_5.3.1.4)/EC_number=5.3.1.4	orth3383; hypothetical_protein				orth4193; Lipopolysaccharide_cholinephosphotransferaseLicD3_(EC_2.7.8.-)		orth4490; hypothetical_protein
orth374; hypothetical_protein	orth3384; hypothetical_protein				orth4010; phage_tail_length_tape-measure_protein		orth4495; hypothetical_protein
orth375; hypothetical_protein	orth3386; hypothetical_protein				orth4026; Putative_phage-related_1,4-beta-N-acetylmuramidase_(cell_wall_hydrolase)(_EC:3.2.1.17_)		orth4500; hypothetical_protein
orth376; positive_transcriptional_regulator_MutR_family	orth3387; hypothetical_protein				orth4035; phage_terminase_large_subunit		orth4510; hypothetical_protein
orth380; Ribulokinase_(EC_2.7.1.16)/EC_number=2.7.1.16	orth3393; Butyrate_kinase_(EC_2.7.2.7)/EC_number=2.7.2.7				orth4065; Portal_protein,_phage_associated		orth4512; hypothetical_protein

orth2226; conserved_hypothetical_protein	orth3394; Branched-chain_alpha-keto_acid_dehydrogenase,E1_component,_alpha_subunit_(EC_1.2.4.4)				orth4175; phage_antirepressor_protein		orth4514; hypothetical_protein
orth2227; hypothetical_protein	orth3395; Branched-chain_alpha-keto_acid_dehydrogenase,E1_component,_beta_subunit_(EC_1.2.4.4)				orth4280; phage_protein		orth4523; hypothetical_protein
orth2251; FIG00744171:_hypothetical_protein	orth3396; Dihydrolipoamide_acyltransferase_component_of branched-chain_alpha-keto_acid_dehydrogenase_complex_(EC				orth4281; prophage_LambdaBa04,_Gp54		orth4525; hypothetical_protein
orth2228; putative_unsaturated_glucuronyl_hydrolase	orth3405; hypothetical_protein				orth4289; phage_protein		orth4528; hypothetical_protein
orth405; Beta-xylosidase_(EC_3.2.1.37)/EC_number=3.2.1.37	orth3409; hypothetical_protein				orth4312; phage_protein		orth4529; hypothetical_protein
orth426; hypothetical_protein	orth3421; hypothetical_protein				orth4343; prophage_pi2_protein_49		orth4530; hypothetical_protein

orth2258; hypothetical_protein	orth3424; hypothetical_protein				orth4184; Prophage_clp_protease-like_protein		orth4531; hypothetical_protein
orth447; Maltose_operon_transcriptional_repressor_MalR,LacI_family	orth3430; Serine/threonine_protein_kinase(_EC:2.7.1.37_)				orth1618; hypothetical_protein		orth4534; hypothetical_protein
orth448; Oligo-1,6-glucosidase_(EC_3.2.1.10)/EC_number=3.2.1.10	orth3450; hypothetical_protein				orth578; hypothetical_protein		orth4535; Phenylalanyl-tRNA_synthetase_beta_chain_(EC6.1.1.20)
orth2314; Glucan_1,6-alpha-glucosidase_(EC_3.2.1.70)/EC_number=3.2.1.70	orth3451; hypothetical_protein				orth1440; hypothetical_protein		orth4541; hypothetical_protein
orth2316; hypothetical_protein	orth3463; FIG00752411:_hypothetical_protein				orth1767; Predicted_membrane_protein		orth4547; hypothetical_protein
orth2265; Putative_uncharacterized_protein	orth3464; hypothetical_protein				orth2241; hypothetical_protein		orth4550; hypothetical_protein
orth456; Transcriptional_regulator,_xre_family	orth3480; Predicted_L-fucose-specific_phosphotransferasesystem,_EIIA_component				orth2302; Beta-glucosidase_(EC_3.2.1.21);6-phospho-beta-glucosidase_(EC_3.2.1.86)		orth4556; hypothetical_protein

orth466; Transcriptional_repressor_of_arabinosideutilization_operon,_GntR_family	orth2598; PTS_system,_IIb_component				orth2303; PTS_system,_cellobiose-specific_IIA_component(EC_2.7.1.69)		orth4558; hypothetical_protein
orth2247; hypothetical_protein	orth2597; PTS_system,_IIc_component				orth1670; PTS_system,_cellobiose-specific_IIC_component(EC_2.7.1.69)		orth4570; hypothetical_protein
orth2255; tributyrin_esterase	orth2495; Metallo-beta-lactamase_superfamily_domainprotein_in_prophage				orth1569; PTS_system,_cellobiose-specific_IIB_component(EC_2.7.1.69)		orth4574; hypothetical_protein
orth470; hypothetical_protein	orth2497; ATPase_involved_in_DNA_repair,_phageassociated				orth542; PRD/PTS_system_IIA_2_domain_protein		orth4576; hypothetical_protein
orth475; hypothetical_protein	orth2549; phage_minor_capsid_protein_#Fam0016				orth3922; hypothetical_protein		orth4577; hypothetical_protein
orth2294; hypothetical_protein	orth3385; ATPase_involved_in_DNA_repair,_phageassociated				orth523; hypothetical_protein		orth4582; hypothetical_protein
orth481; FIG00745047:_hypothetical_protein	orth3388; phage_protein				orth586; D-arabino-3-hexulose_6-phosphate_formaldehydelyase		orth4585; hypothetical_protein

orth482; ABC-type_uncharacterized_transport_system,permease_component	orth3391; prophage_pi3_protein_01				orth587; transcriptional_regulator,_DeoR-family		orth4586; hypothetical_protein
orth2293; ABC-type_oligopeptide_transport_system,_ATPasecomponent	orth3436; phage_anti-repressor_protein				orth602; hypothetical_protein		orth4591; hypothetical_protein
orth494; Transcriptional_regulator,_RpiR_family	orth3502; prophage_pi3_protein_01				orth1416; 6-phospho-beta-glucosidase_(EC_3.2.1.86)/EC_number=3.2.1.86		orth4593; hypothetical_protein
orth496; putative_transcriptional_regulator,_XRE_familyprotein					orth1417; Transposase		orth4602; hypothetical_protein
orth2222; putative_inosine-uridine_preferring_nucleosidehydrolase					orth1418; hypothetical_protein		orth4611; hypothetical_protein
orth2229; PTS_system,_IIa_component					orth1519; ISSth1,_transposase_(orf1),_IS3_family		orth4612; hypothetical_protein

orth441; PTS_system,_mannose-specific_IID_component_(EC2.7.1.69)					orth1547; Nucleotidyltransferase_(EC_2.7.7.-)		orth4621; hypothetical_protein
orth2257; PTS_system,_mannose-specific_IIC_component_(EC2.7.1.69)					orth1601; hypothetical_protein		orth4622; hypothetical_protein
orth2315; Maltose/maltodextrin_ABC_transporter,_substratebinding_periplasmic_protein_MalE					orth1603; hypothetical_protein		orth4626; hypothetical_protein
orth450; Maltose/maltodextrin_ABC_transporter,_permeaseprotein_MalF					orth1604; FIG036446:_hypothetical_protein		orth4636; hypothetical_protein
orth451; Maltose/maltodextrin_ABC_transporter,_permeaseprotein_MalG					orth1671; transposase		orth4641; hypothetical_protein

orth454; PTS_system_sorbose_subfamily_IIB_component					orth1675; Functional_role_page_for_Anaerobic_nitric_oxidereductase_transcription_regulator_NorR		orth4642; Mevalonate_kinase_(EC_2.7.1.36)/EC_number=2.7.1.36
orth2256; PTS_system,_IIB_component					orth1677; Putative_uncharacterized_protein		orth4645; hypothetical_protein
orth468; PTS_system,_mannose-specific_IIC_component_(EC2.7.1.69)					orth1678; COG0637:_Predictedphosphatase/phosphohexomutase		orth4658; hypothetical_protein
orth469; PTS_system,_mannose/fructose_family_IIDcomponent					orth1679; Mannitol-1-phosphate_5-dehydrogenase_(EC1.1.1.17)		orth4660; hypothetical_protein
orth495; PTS_system,_maltose_and_glucose-specific_IICcomponent_(EC_2.7.1.69)_/_PTS_system,_maltose_and					orth1681; hypothetical_protein		orth4675; hypothetical_protein
					orth1697; hypothetical_protein		orth4677; hypothetical_protein
					orth1781; 6-phospho-3-hexuloisomerase		orth4678; hypothetical_protein

					orth1811; hypothetical_protein		orth4686; hypothetical_protein
					orth1676; PTS_system,_fructose-_and_mannose-inducibleputative_EII_component		orth4693; hypothetical_protein
					orth588; PTS_system,_fructose-_and_mannose-inducible_IIDcomponent_(EC_2.7.1.69)		orth4696; hypothetical_protein
					orth589; PTS_system,_fructose-_and_mannose-inducible_IICcomponent_(EC_2.7.1.69)		orth4698; hypothetical_protein
					orth1558; PTS_system,_fructose-_and_mannose-inducible_IIBcomponent_(EC_2.7.1.69)		orth4700; hypothetical_protein

					orth1786; PTS_system,_fructose-_and_mannose-inducible_IIAcomponent_(EC_2.7.1.69)		orth4704; hypothetical_protein
					orth585; PTS_system,_mannitol-specific_IIB_component_(EC2.7.1.69)_/_PTS_system,_mannitol-specific_IIC_component		orth4705; hypothetical_protein
					orth590; PTS_system,_cellobiose-specific_IIB_component(EC_2.7.1.69)		orth4706; hypothetical_protein
					orth1472; PTS_system,_cellobiose-specific_IIA_component(EC_2.7.1.69)		orth4708; hypothetical_protein
					orth1474; PTS_system,_cellobiose-specific_IIC_component(EC_2.7.1.69)		orth4709; hypothetical_protein
					orth1619; mobile_element_protein		orth4068; mobile_element_protein

							orth4220; PTS_system,_N-acetylgalactosamine-specific_IICcomponent_(EC_2.7.1.69)
							orth4262; PTS_system,_N-acetylgalactosamine-_andgalactosamine-specific_IIA_component_(EC_2.7.1.69)
							orth4466; ABC_transporter,_ATP-binding_protein
							orth4050; polysaccharide_biosynthesis_protein
							orth4031; phage_terminase_large_subunit
							orth4075; phage_portal_protein
							orth4131; cII-like_protein,_phage_associated
							orth4228; phage_tail_fiber_protein

							orth4246; phage_terminase_small_subunit
							orth4253; prophage_pi1_protein_32
							orth4182; Prophage_clp_protease-like_protein
							orth2139; FIG00742964:
							orth1011; ABC_transporter,_ATP-binding_protein
							orth3787; D-3-phosphoglycerate_dehydrogenase_(EC1.1.1.95)
							orth722; Bifunctional_protein:_zinc-containing_alcoholdehydrogenase;_quinone_oxidoreductase_(_NADPH:quinone
							orth434; Transcriptional_regulator,_xre_family

							orth3286; FIG00750480:_hypothetical_protein
							orth270; ribose_operon_repressor,_putative
							orth2938; phosphoenolpyruvate-dependent_sugarphosphotransferase_system_EIIB,_probable_sorbose
							orth3069; Phosphotransferase_system,mannose/fructose/N-acetylgalactosamine-specific_component
							orth3734; phosphoenolpyruvate-dependent_sugarphosphotransferase_system_EIID,_probable_fructose
							orth3815; Maltodextrin_glucosidase_(EC_3.2.1.20)/EC_number=3.2.1.20

							orth3297; PTS_system,_mannose-specific_IIA_component_(EC2.7.1.69)
							orth2955; Oligo-1,6-glucosidase_(EC_3.2.1.10)/EC_number=3.2.1.10
							orth1254; FIG00743082:_hypothetical_protein
							orth1218; hypothetical_protein
							orth3247; FIG00753206:_hypothetical_protein
							orth271; Transcriptional_regulator,_TetR_family
							orth770; FIG00746376:_hypothetical_protein
							orth2633; FIG00749467:_hypothetical_protein

							orth3058; ABC-type_multidrug_transport_system(daunorubicin_resistance),_ATPase_component
							orth1009; ABC_transporter,_permease_protein
							orth3074; ABC_transporter,_ATP-binding_protein
							orth3344; FIG00746758:_hypothetical_protein
							orth2337; hypothetical_protein
							orth2148; hypothetical
							orth1023; hypothetical_protein
							orth3929; hypothetical_protein
							orth2624; hypothetical_protein
							orth2250; unknown_protein
							orth3868; Rep_protein

							orth2285; hypothetical_protein
							orth3220; hypothetical_protein
							orth971; hypothetical_protein
							orth1808; hypothetical_protein
							orth778; hypothetical_protein
							orth762; Glycosyl_transferase,_family_2
							orth995; Glycosyltransferase
							orth3326; modification_methylase_(Cytosine-specificmethyltransferase
							orth747; hypothetical_protein
							orth3319; Superfamily_II_DNA_and_RNA_helicase
							orth3810; hypothetical_protein
							orth2107; hypothetical

							orth2159; hypothetical
							orth3850; oligogalacturonide_transporter
							orth2160; hypothetical
							orth3763; hypothetical_protein
							orth2679; abortive_infection_protein_AbiGI

Supplementary Table 5. Predicted peptidase OGs of the *L. rhamnosus* strains.

KEGG	Gene name	Protease type	Annotation	OG/ GG gene	Presence
K03101	lspA	Aspartic Peptidases	signal peptidase II	orth347	core
K00764	purF	Cysteine Peptidases	amidophosphoribosyltransferase	orth1865	core
K00820	glmS	Cysteine Peptidases	glucosamine--fructose-6-phosphate aminotransferase	orth2970	core
K01304	pcp	Cysteine Peptidases	pyroglutamyl-peptidase	orth3100	core
K01372	pepC	Cysteine Peptidases	bleomycin hydrolase	orth569	core
K01951	guaA	Cysteine Peptidases	GMP synthase (glutamine-hydrolysing)	orth1928	core
K01953	asnB	Cysteine Peptidases	asparagine synthase (glutamine-hydrolysing)	orth2202	core
K05520	pfpI	Cysteine Peptidases	protease I	orth1891	core
K07010		Cysteine Peptidases	putative glutamine amidotransferase	orth2534, orth4421	core
K07284	srtA	Cysteine Peptidases	sortase A	orth1655, orth3244	core
K08659	pepDA, pepDB	Cysteine Peptidases	dipeptidase	orth1130, orth2096, orth2237	core
K01256	pepM	Metallo Peptidases	aminopeptidase N	orth2997	core
K01258	pepT	Metallo Peptidases	tripeptide aminopeptidase	orth122	core
K01262	pepP	Metallo Peptidases	Xaa-Pro aminopeptidase	orth3107	core
K01265	map	Metallo Peptidases	methionyl aminopeptidase	orth3605	core
K01271	pepQ	Metallo Peptidases	Xaa-Pro dipeptidase	orth2552	core
K01409	gcp	Metallo Peptidases	O-sialoglycoprotein endopeptidase	orth2555	core
K01436		Metallo Peptidases	amidohydrolase	orth1222	core

K03798	ftsH	Metallo Peptidases	cell division protease	orth255	core
K05823		Metallo Peptidases	N-acetyldiaminopimelate deacetylase	orth1901	core
K07386	pepO	Metallo Peptidases	putative endopeptidase	orth1148, orth2704	core
K08602	pepF, pepB	Metallo Peptidases	oligoendopeptidase F	orth2583	core
K08643	zmpB	Metallo Peptidases	zinc metalloprotease	orth470	variable
K11749	rseP	Metallo Peptidases	regulator of sigma E protease	orth2271	core
K12941	abgB	Metallo Peptidases	aminobenzoyl-glutamate utilization protein B	orth614	variable
K16203	dppA	Metallo Peptidases	D-amino peptidase	orth1368	variable
K01259	pip	Serine endopeptidases	proline iminopeptidase	orth858, orth3239	core
K01281	pepXP	Serine endopeptidases	X-Pro dipeptidyl-peptidase	orth3514	core
K01338	lon	Serine endopeptidases	ATP-dependent Lon protease	orth2471	variable
K01356	lexA	Serine endopeptidases	repressor LexA	orth978	core
K01358	clpP	Serine endopeptidases	ATP-dependent Clp protease, protease subunit	orth117	core
K01361		Serine endopeptidases	lactocepin	orth1104	core
K03100	lepB	Serine endopeptidases	signal peptidase I	orth2680	core
K03797	ctpA	Serine endopeptidases	carboxyl-terminal processing protease	orth292	core
K07258	dacA, dacB, dacC	Serine endopeptidases	D-alanyl-D-alanine carboxypeptidase (penicillin-binding protein	orth2197	core
K01419	hslV, clpQ	Threonine Peptidases	ATP-dependent HslUV protease	orth1201	core

Supplementary Table.6. Specific genes for *L. rhamnosus* strains – listing of OG numbers and RAST annotation.

OG	Size	Strain	Annotation	Discussed in the text
4174	247	Lrh13	hypothetical_protein	
4301	112	Lrh23	ATP-dependent_protease_(EC_3.4.21.-)	
4309	109	Lrh23	hypothetical_protein	
4124	308	Lrh23	truncated_RepA	
4199	221	Lrh23	putative_mobilization_protein	
4106	334	Lrh22	D-lactate_dehydrogenase_(EC_1.1.1.28)/EC_number=1.1.1.28	
4093	357	Lrh22	Malate_permease	*
358	746	Lrh22	hypothetical_protein	
4276	129	Lrh22	nucleoid_DNA-binding_protein	
4090	366	Lrh22	hypothetical_protein	
4043	513	Lrh22	hypothetical_protein	
4144	285	Lrh22	hypothetical_protein	
4149	278	Lrh22	hypothetical_protein	

4139	291	Lrh22	hypothetical_protein	
4194	228	Lrh22	SortaseA1,_surface_protein_transpeptidase	*
4132	302	Lrh22	hypothetical_protein	
4238	168	Lrh22	hypothetical_protein	
4177	244	Lrh22	hypothetical_protein	
4129	305	Lrh22	Hypothetical_similar_to_thiamin_biosynthesislipoprotein_ApbE	
4232	173	Lrh22	FIG00750704:_hypothetical_protein	* sugar cluster
4190	230	Lrh22	Triosephosphate_isomerase_(EC_5.3.1.1)/EC_number=5.3.1.1	* sugar cluster
4142	287	Lrh22	Fructose-bisphosphate_aldolase_class_II_(EC4.1.2.13)	* sugar cluster
4059	445	Lrh22	PTS_system,_galactitol-specific_IIC_component(EC_2.7.1.69)	* sugar cluster
4242	159	Lrh22	PTS_system,_galactitol-specific_IIA_component(EC_2.7.1.69)	* sugar cluster
4266	137	Lrh22	hypothetical_protein	* sugar cluster
4046	494	Lrh22	Transcriptional_antiterminator_of_lichenanoperon,_BglG_family	* sugar cluster

4113	326	Lrh22	4-hydroxythreonine-4-phosphate_dehydrogenase(EC_1.1.1.262)	* sugar cluster
4066	409	Lrh22	hypothetical_protein	* sugar cluster
4108	331	Lrh22	2-keto-3-deoxygluconate_permease_(KDGpermease)	* sugar cluster
4033	577	Lrh22	hypothetical_protein	* sugar cluster
4302	112	Lrh22	hypothetical_protein	
4157	270	Lrh22	hypothetical_protein	
4267	136	Lrh22	hypothetical_protein	
4148	280	Lrh22	hypothetical_protein	
4135	302	Lrh22	hypothetical_protein	
4085	374	Lrh22	hypothetical_protein	
4038	537	Lrh22	FIG00513666:_hypothetical_protein	
4071	403	Lrh22	Tn916,_transcriptional_regulator,_putative	
4305	110	Lrh22	hypothetical_protein	
4239	166	Lrh22	Hypothetical_membrane_protein	
4126	308	Lrh22	hypothetical_protein	

4159	269	Lrh22	Integrase	
4198	221	Lrh22	hypothetical_protein	
4264	139	Lrh22	hypothetical_protein	
4152	274	Lrh22	transposase,_IS4	
4249	151	Lrh22	phage_terminase,_small_subunit	
4076	392	Lrh22	Uncharacterized_protein,_YUKC_B.subtilishomolog	
4236	171	Lrh22	hypothetical_protein	
4011	958	Lrh22	hypothetical_protein	
4112	327	Lrh22	Aggregation_promoting_factor	
4128	307	Lrh22	Zeta_toxin	*toxin
4282	126	Lrh22	hypothetical_protein	
4165	263	Lrh22	putative_plasmid_partition_protein	*plasmid
4296	114	Lrh22	hypothetical_protein	
4283	124	Lrh22	hypothetical_protein	
4061	433	Lrh22	hypothetical_Cytosolic_Protein	

4049	477	Lrh22	FtsK/SpoIIIE_family_protein,_putative_secretionsystem_component_EssC/YukA	
4028	650	Lrh22	PTS_system,_sucrose-specific_IIB_component_(EC2.7.1.69)_/_PTS_system,_sucrose-specific_IIC_component	
4114	326	Lrh22	Sucrose_operon_repressor_ScrR,_LacI_family	
4320	103	Lrh11	hypothetical_protein	
4161	265	Lrh11	mobile_element_protein	
4201	220	Lrh30	hypothetical_protein	
4057	450	Lrh30	hypothetical_protein	
4068	408	Lrh32	mobile_element_protein	
4230	175	Lrh32	Alpha/beta_superfamily_hydrolase	
4250	150	Lrh32	hypothetical_protein	
4203	218	Lrh32	hypothetical_protein	
4322	102	Lrh32	hypothetical_protein	
4164	263	Lrh32	putative_plasmid_partition_protein	*plasmid
4169	252	Lrh32	hypothetical_protein	

4228	177	Lrh32	phage_tail_fiber_protein	
4069	406	Lrh28	Type_I_restriction-modification_system,DNA-methyltransferase_subunit_M_(EC_2.1.1.72)	
4079	387	Lrh28	Type_I_restriction-modification_system,specificity_subunit_S_(EC_3.1.21.3)	
4014	884	Lrh28	Type_I_restriction-modification_system,restriction_subunit_R_(EC_3.1.21.3)	
4016	842	Lrh28	hypothetical_protein	
4008	1082	Lrh28	hypothetical_protein	
4054	457	Lrh28	hypothetical_protein	
4025	669	Lrh28	hypothetical_protein	
4247	154	Lrh28	hypothetical_protein	
4105	335	Lrh28	hypothetical_protein	
4274	129	Lrh28	hypothetical_protein	
4254	148	Lrh28	hypothetical_protein	
4300	112	ATCC8530	PapB	*bacteriocin
4316	106	ATCC8530	Thioredoxin	*stress resistance

4259	142	Lrh33	N-acetylglucosamine/galactosamine_PTS,_EIIA(EC:2.7.1.69_)	
4168	253	Lrh33	hypothetical_protein	
4298	113	Lrh33	hypothetical_protein	
4024	678	Lrh33	hypothetical_protein	
4269	135	Lrh33	hypothetical_protein	
4306	110	Lrh33	Hypothetical_protein	
4189	231	Lrh33	response_regulator	
4056	452	Lrh33	Signal_transduction_histidine_kinase	
4207	216	Lrh33	DedA_protein	
4196	223	Lrh33	Membrane-associated_phospholipid_phosphatase	
4121	315	Lrh33	Glycosyltransferase	
4053	463	Lrh33	FIG00749762:_hypothetical_protein	
4020	735	Lrh33	relaxase_Mob_DEI	
4116	324	Lrh33	LtrC	
4081	385	Lrh33	FIG00750967:_hypothetical_protein	

4231	174	Lrh33	hypothetical_protein	
4017	825	Lrh33	Surface_antigen_p40	
4018	818	Lrh33	FIG00629370:_hypothetical_protein	
4150	277	Lrh33	Tn5252,_Orf23	
4030	633	Lrh33	hypothetical_protein	
4244	158	Lrh33	hypothetical_protein	
4045	499	Lrh33	bacillolysin(_EC:3.4.24.27_)	*toxin
4308	109	Lrh33	Non-specific_DNA-binding_protein_Dps_/Iron-binding_ferritin-like_antioxidant_protein_/	
4227	178	Lrh33	FIG00748045:_hypothetical_protein	
4245	157	Lrh33	Teichoic_acid_glycosylation_protein	
4032	582	Lrh33	putative_transposase	
4074	399	Lrh33	transposase_(partial)	
4215	198	Lrh33	putative_resolvase	
4313	107	Lrh33	Arsenical_resistance_operon_repressor	*arsenic resistance

4034	576	Lrh33	Arsenical_pump-driving_ATPase_(EC_3.6.3.16)/EC_number=3.6.3.16	*arsenic resistance
4062	431	Lrh33	Arsenic_efflux_pump_protein	*arsenic resistance
4304	111	Lrh33	Arsenical_resistance_operon_trans-actingrepressor_ArsD	*arsenic resistance
4324	101	Lrh33	Transposase,_IS30_family	
4206	216	Lrh33	YbbL_ABC_transporter_ATP-binding_protein	
4258	143	Lrh33	transcriptional_regulator,_XRE_family	
4216	197	Lrh33	hypothetical_protein	
4138	292	Lrh33	hypothetical_protein	
4261	141	Lrh33	Putative_uncharacterized_protein	
4006	1503	Lrh33	FIG00745879:_hypothetical_protein	
4188	231	Lrh33	FIG00742740:_hypothetical_protein	
4010	990	Lrh33	phage_tail_length_tape-measure_protein	*phage
4289	121	Lrh33	phage_protein	*phage
4209	215	Lrh33	major_tail_protein,_phi13_family	*phage
4280	127	Lrh33	phage_protein	*phage

4307	109	Lrh33	Putative_uncharacterized_protein	*phage
4312	108	Lrh33	phage_protein	*phage
4067	409	Lrh33	FIG00744946:_hypothetical_protein	*phage
4184	235	Lrh33	Prophage_clp_protease-like_protein	*phage
4065	411	Lrh33	Portal_protein,_phage_associated	*phage
4035	576	Lrh33	phage_terminase_large_subunit	*phage
4286	124	Lrh33	Putative_uncharacterized_protein	*phage
4281	126	Lrh33	prophage_LambdaBa04,_Gp54	*phage
4279	127	Lrh33	Hypothetical_protein	
4153	274	Lrh33	Helicase_loader_DnaI	
4173	249	Lrh33	FIG00751682:_hypothetical_protein	
4092	363	Lrh33	hypothetical_protein	
4060	435	Lrh33	DNA-cytosine_methyltransferase_(EC_2.1.1.37)/EC_number=2.1.1.37	
4180	240	Lrh33	hypothetical_cytosolic_protein	
4047	492	Lrh33	hypothetical_protein	

4029	639	Lrh33	Restriction_enzyme_BgcI_alpha_subunit_(EC3.1.21.-)_[Includes:_Adenine-specific_methyltransferase	
4091	363	Lrh33	FIG00712184:_hypothetical_protein	
4107	331	Lrh33	Ornithine_cyclodeaminase_(EC_4.3.1.12)/EC_number=4.3.1.12	
4141	289	Lrh33	FIG00745596:_hypothetical_protein	
4200	221	Lrh34	hypothetical_protein	
4237	171	Lrh34	hypothetical_protein	
4019	740	Lrh34	virulence-associated_E	
4073	401	Lrh34	Integrase,_superantigen-encoding_pathogenicityislands_SaPI	
4036	556	Lrh34	hypothetical_protein	
4319	103	Lrh34	hypothetical_protein	
4175	247	Lrh34	phage_antirepressor_protein	
4147	281	Lrh34	hypothetical_protein	
4204	218	Lrh19	membrane_protein	

4072	401	Lrh19	Beta-lactamase_class_C_and_other_penicillinbinding_proteins	
4041	519	Lrh19	hypothetical_protein	
4223	186	Lrh19	hypothetical_protein	
4166	256	Lrh19	Hypothetical_membrane_spanning_protein	*EPS cluster
4012	950	Lrh19	glycosyl_hydrolase_53(_EC:3.2.1.89_)	*EPS cluster
4026	651	Lrh19	Putative_phage-related_1,4-beta-N-acetylmuramidase_(cell_wall_hydrolase)(_EC:3.2.1.17_)	*EPS cluster
4052	463	Lrh19	membrane_protein,_putative	*EPS cluster
4086	372	Lrh19	Glycosyltransferase	*EPS cluster
4039	526	Lrh19	Membrane_protein_involved_in_the_export_ofO-antigen,_teichoic_acid_lipoteichoic_acids	*EPS cluster
4119	322	Lrh19	Beta-1,3-glucosyltransferase	*EPS cluster
4095	356	Lrh19	dTDP-glucose_4,6-dehydratase_(EC_4.2.1.46)/EC_number=4.2.1.46	*EPS cluster
4186	234	Lrh19	2-C-methyl-D-erythritol_4-phosphatecytidylyltransferase_(EC_2.7.7.60)	*EPS cluster

4193	228	Lrh19	Lipopolysaccharide_cholinephosphotransferaseLicD3_(EC_2.7.8.-)	*EPS cluster
4155	272	Lrh19	Glycosyl_transferase,_family_2	*EPS cluster
4211	205	Lrh19	hypothetical_protein	*EPS cluster
4240	163	HN001	hypothetical	
4171	251	HN001	hypothetical	
4212	202	Lrh31	hypothetical_protein	
4292	120	Lrh31	hypothetical_protein	
4295	115	Lrh31	hypothetical_protein	
4101	342	Lrh31	hypothetical_protein	
4163	263	Lrh31	glycosyltransferase_family_2	
4096	355	Lrh31	hypothetical_protein	
4133	302	Lrh31	Glycosyl_transferase,_group_2_family_protein	
4122	313	Lrh31	Alpha-L-Rha_alpha-1,3-L-rhamnosyltransferase(EC_2.4.1.-)	

4048	481	Lrh31	Membrane_protein_involved_in_the_export_ofO-antigen,_teichoic_acid_lipoteichoic_acids	
4118	322	Lrh31	protein_of_unknown_function_DUF23	
4248	153	Lrh31	hypothetical_protein	
4251	150	Lrh5	hypothetical_protein	
4167	254	Lrh5	hypothetical_protein_within_pathogenicityisland	
4178	242	Lrh5	hypothetical_protein	
4315	106	Lrh5	hypothetical_protein	
4252	149	Lrh5	hypothetical_protein	
4023	696	Lrh5	hypothetical_protein	
4202	218	Lrh5	ABC_transporter_ATP-binding_protein	
4222	187	Lrh5	hypothetical_protein	
4058	449	Lrh5	hypothetical_protein	
4158	269	Lrh5	hypothetical_protein	
4102	339	Lrh5	predicted_ATP-dependent_endonuclease,_OLDfamily	

4208	216	Lrh5	predicted_ATP-dependent_endonuclease,_OLDfamily	
4104	337	Lrh5	acyltransferase/acetyltransferase	
4137	298	Lrh5	hypothetical_protein	
4125	308	Lrh5	FIG00518155:_hypothetical_protein	
4063	430	Lrh5	phage_terminase,_large_subunit	*phage
4044	506	Lrh5	phage_portal	*phage
4185	235	Lrh5	hypothetical_protein	*phage
4226	180	Lrh5	phage_capsid_and_scaffold	*phage
4130	305	Lrh5	phage_major_capsid_protein	*phage
4291	120	Lrh5	ORF042	*phage
4284	124	Lrh5	putative_head_to_tail_joining	*phage
4290	121	Lrh5	hypothetical_protein	*phage
4205	217	Lrh5	conserved_hypothetical_protein	*phage
4255	146	Lrh5	conserved_hypothetical_protein	*phage
4009	998	Lrh5	phage_tail_length_tape-measure_protein	*phage

4156	272	Lrh5	minor_tail_protein	*phage
4022	702	Lrh5	hypothetical_protein	
4115	324	Lrh5	Orf55	
4080	386	Lrh5	hypothetical_protein	
4192	228	Lrh5	Hypothetical_SAV2027_homolog_insuperantigen-encoding_pathogenicity_islands_SaPI	
4131	305	Lrh24	cII-like_protein,_phage_associated	
4162	265	Lrh24	FIG045374:_Type_II_restriction_enzyme,methylase_subunit_YeeA	
4027	650	Lrh24	FIG045374:_Type_II_restriction_enzyme,methylase_subunit_YeeA	
4310	108	Lrh24	ORF-11	
4234	172	Lrh24	hypothetical_protein	
4094	357	Lrh24	transcriptional_regulator,_Cro/CI_family	
4187	232	Lrh24	putative_transposase	
4145	282	Lrh24	transposase	
4285	124	Lrh24	FIG00743773:_hypothetical_protein	

4064	413	Lrh24	FIG01282627:_hypothetical_protein	
4182	238	Lrh24	Prophage_clp_protease-like_protein	
4075	395	Lrh24	phage_portal_protein	
4031	630	Lrh24	phage_terminase_large_subunit	
4246	157	Lrh24	phage_terminase_small_subunit	
4253	149	Lrh24	prophage_pi1_protein_32	
4170	252	Lrh24	IstB_domain_protein_ATP-binding_protein	
4040	521	Lrh24	Transposase_subunit	
4181	238	Lrh24	Oxygen-insensitive_NADPH_nitroreductase_(EC1.-.-.-)	*EPS cluster
4077	390	Lrh24	truncated_transposase_for_IS1272	*EPS cluster
4220	189	Lrh24	PTS_system,_N-acetylgalactosamine-specific_IICcomponent_(EC_2.7.1.69)	*EPS cluster
4262	140	Lrh24	PTS_system,_N-acetylgalactosamine-_andgalactosamine-specific_IIA_component_(EC_2.7.1.69)	*EPS cluster
4120	320	Lrh24	putative_glycosyltransferase	*EPS cluster
4050	470	Lrh24	polysaccharide_biosynthesis_protein	*EPS cluster

4109	330	Lrh24	putative_glycosyltransferase	*EPS cluster
4087	371	Lrh24	hypothetical_protein	*EPS cluster
4111	327	Lrh24	Beta-1,3-glucosyltransferase	*EPS cluster
4098	351	Lrh24	hypothetical_protein	*EPS cluster
4083	383	Lrh24	Glycosyl_transferase,_group_1	*EPS cluster
4195	227	Lrh24	Lipid_carrier_:UDP-N-acetylgalactosaminyltransferase_(EC_2.4.1.-)	*EPS cluster
4136	299	Lrh24	ATPase_involved_in_chromosome_partitioning	
4229	176	Lrh24	Transposase	
4123	313	Lrh24	Lysophospholipase_(EC_3.1.1.5)	
4260	141	Lrh24	FIG00632156:_hypothetical_protein	
4037	548	Lrh29	conserved_hypothetical_protein,_probablyfragment	
4015	844	Lrh29	putative_cytoplasmic_protein	
4021	723	Lrh29	putative_type_II_restriction_enzyme_methylasesubunit	

2472	1212	Lrh29	putative_type_II_restriction_enzyme_methylasesubunit	
4007	1241	Lrh29	FIG00753307:_hypothetical_protein	
4214	199	Lrh29	hypothetical_protein	
4241	159	Lrh29	hypothetical_protein	
4146	282	Lrh29	hypothetical_protein	
4213	201	Lrh29	FIG00745257:_hypothetical_protein	
4172	250	Lrh29	hypothetical_protein	
4160	268	Lrh29	hypothetical_protein	
4070	404	Lrh29	CpsX_protein cold shock	*stress resistance
4235	172	Lrh29	hypothetical_protein	
4100	345	Lrh29	putative_acetyl_transferase	
4256	144	Lrh29	hypothetical_protein	
4117	324	Lrh29	hypothetical_protein	
4268	136	Lrh29	orf15	
4225	181	Lrh29	FIG00754149:_hypothetical_protein	

2523	215	Lrh29	hypothetical_protein	
4127	307	Lrh29	ParA_family_protein	
4197	222	Lrh4	hypothetical_protein	
2733	150	Lrh4	FIG00745703:_hypothetical_protein	
4299	112	Lrh6	N-acetylmuramic_acid_6-phosphate_etherase_(EC4.2.-.-)	
4078	389	Lrh3	Relaxase/Mobilisation_nuclease_domain	
4293	117	Lrh3	mobilization_protein	
4303	111	Lrh3	Transcriptional_regulator,_xre_family	
4314	107	Lrh3	killer_suppression_protein_HigA,_putative	
4055	457	Lrh3	polysaccharide_polymerase	
4097	352	Lrh3	Glycosyltransferase	
4082	383	Lrh3	Glycosyltransferase	
4089	367	Lrh3	Lipopolysaccharide_1,6-galactosyltransferase(EC_2.4.1.-)	
4179	242	Lrh3	hypothetical_protein	
4183	238	Lrh3	truncated_RepA	*plasmid

4323	102	Lrh21	hypothetical_protein	
4191	229	Lrh27	hypothetical_protein	
4154	273	Lrh27	phage_replication_protein_#_ACLAME_208	
4051	468	Lrh27	hypothetical_protein	
4318	103	Lrh27	Hypothetical_prophage_lsa1_protein	
4224	184	Lrh27	hypothetical_protein	
4321	102	Lrh27	hypothetical_protein	
4140	290	Lrh27	DnaD_domain_protein	
4311	108	Lrh27	mobilization_protein	
4005	1530	Lrh27	phage_tail_length_tape-measure_protein	
4210	206	Lrh27	hypothetical_protein	
4257	143	Lrh27	hypothetical_protein	
4243	158	Lrh27	major_capsid_protein_gpP	
4273	130	Lrh27	minor_capsid_protein	
4288	122	Lrh27	hypothetical_protein	

4294	116	Lrh27	hypothetical_protein	
4263	139	Lrh27	hypothetical_protein	
4110	330	Lrh27	phage_major_capsid_protein	
4217	197	Lrh27	putative_scaffolding_protein	
4275	129	Lrh27	hypothetical_protein	
4297	113	Lrh27	hypothetical_protein	
4084	381	Lrh27	phage_minor_capsid_protein	
4042	515	Lrh27	phage_minor_capsid_protein	
4277	128	Lrh27	ORF42	
4272	132	Lrh27	hypothetical_protein	
4088	370	Lrh27	Prophage_Lp2_protein_6	
4221	188	Lrh1	hypothetical_protein	
4151	276	Lrh1	hypothetical_protein	
4271	133	Lrh2	Chaperone_protein_DnaJ	
4134	302	Lrh12	hypothetical_protein	

Supplementary Table.7. All identified transporter OGs.

OG	Presence	PTS	Complete PTS	ABC	Import/ Export	RAST annotation	Core	LGG gene
3485	29	1				PTS_system,_fructose-specific_IIB_component_(EC2.7.1.69)_/_PTS_system,_fructose-specific_IIC_component	no	LGG_00090
1285	29	1				PTS_system,_fructose-specific_IIA_component_(EC2.7.1.69)	no	LGG_00092
141	40	2				PTS_system,_sucrose-specific_IIB_component_(EC2.7.1.69)_/_PTS_system,_sucrose-specific_IIC_component	core	LGG_00125
3525	40	3	19			PTS_system,_cellobiose-specific_IIB_component(EC_2.7.1.69)	core	LGG_00159
3491	40	3	19			PTS_system,_cellobiose-specific_IIA_component(EC_2.7.1.69)	core	LGG_00160
3157	37	3	19			PTS_system,_cellobiose-specific_IIC_component(EC_2.7.1.69)	core	LGG_00162
1419	40	4				PTS_system,_mannose-specific_IIB_component_(EC2.7.1.69)	core	LGG_00337
3529	40	4				PTS_system,_mannose-specific_IID_component_(EC2.7.1.69)	no	LGG_00339
1932	20	5	11			PTS_system,_galactitol-specific_IIC_component(EC_2.7.1.69)	no	LGG_00343

2720	20	5	11			PTS_system,_galactitol-specific_IIA_component(EC_2.7.1.69)	no	LGG_00345
1244	20	5	11			PTS_system,_galactitol-specific_IIB_component(EC_2.7.1.69)	no	LGG_00346
507	20	5	11			FIG00743883:_hypothetical_protein_galactitol pts related	no	LGG_00347
3165	40	6	1			PTS_system,_cellobiose-specific_IIC_component(EC_2.7.1.69)	core	LGG_00353
1980	40	6	1			PTS_system,_beta-glucoside-specific,_IIBcomponent	no	LGG_00354
161	40	6	1			PTS_system,_cellobiose-specific_IIA_component(EC_2.7.1.69)	no	LGG_00355
1378	29	7				PTS_system,_fructose-specific_IIB_component_(EC2.7.1.69)	no	LGG_00396
2573	29	7				PTS_system,_fructose-specific_IIA_component_(EC2.7.1.69)_/_PTS_system,_fructose-specific_IIB_component	no	LGG_00397
1630	10	8				Predicted_galactitol_operon_regulator(Transcriptional_antiterminator),_BglG_family_/_PTS	no	LGG_00400
3212	10	8				FIG00629163:_hypothetical_protein_galactitol PTS	no	LGG_00401
3339	10	9				Transport_protein_SgaT,_putative_galactitol PTS	core	LGG_00404
1379	10	10	20			PTS_system,_fructose-specific_IIA_component_(EC2.7.1.69)	no	LGG_00409
1629	10	10	20			PTS_system,_fructose-specific_IIBC_component(EC_2.7.1.69)	no	LGG_00411

2487	8	11				PTS_system,_mannose-specific_IIB_component_(EC2.7.1.69)	no	LGG_00415
716	8	11				PTS_system,_sorbose-specific_IID_component_(EC2.7.1.69)	no	LGG_00417
2948	40	12	2			PTS_system,_trehalose-specific_IIB_component(EC_2.7.1.69)_/_PTS_system,_trehalose-specific_IIC	core	LGG_00603
86	40	12	2			PTS_system,_lactose-specific_IIB_component_(EC2.7.1.69)_/_PTS_system,_lactose-specific_IIC_component	core	LGG_00649
1026	40	12	2			PTS_system,_lactose-specific_IIA_component_(EC2.7.1.69)	core	LGG_00652
2289	33	13				PTS_system,_galactose-specific_IIC_component(EC_2.7.1.69)	no	LGG_00659
2290	40	14				PTS_system,_glucitol/sorbitol-specific_IIAcomponent_(EC_2.7.1.69)	core	LGG_00872
1453	39	15				PTS_system,_cellobiose-specific_IIA_component(EC_2.7.1.69)	core	LGG_01045
1461	39	16				PTS_system,_beta-glucoside-specific_IIBcomponent_(EC_2.7.1.69)_/_PTS_system,	core	LGG_01065
2390	40	17				PTS_system,_fructose-specific_IIA_component_(EC2.7.1.69)_/_PTS_system,_fructose-specific_IIB_component	core	LGG_01359
743	40	18				PTS_system,_cellobiose-specific_IIC_component(EC_2.7.1.69)	core	LGG_02078

744	40	19				PTS_system,_sucrose-specific_IIB_component_(EC2.7.1.69)_/_PTS_system,_sucrose-specific_IIC_component	core	LGG_02104
2653	40	20				PTS_system,_beta-glucoside-specific_IIBcomponent_(EC_2.7.1.69)_/_PTS_system,	core	LGG_02192
3226	39	21				PTS_system,_galactose-specific_IIA_component(EC_2.7.1.69)	core	LGG_02574
239	39	22				PTS_system,_galactose-inducible_IIB_component(EC_2.7.1.69)_/_PTS_system,_galactose-inducible_IIC	core	LGG_02647
477	25	23				PTS_system,_mannose-specific_IIC_component_(EC2.7.1.69)	no	LGG_02654
2297	25	23				PTS_system,_mannose-specific_IID_component_(EC2.7.1.69)	no	LGG_02655
2421	22	24	9			PTS_system,_galactitol-specific_IIB_component(EC_2.7.1.69)	no	LGG_02666
3782	22	24	9			PTS_system,_galactitol-specific_IIA_component(EC_2.7.1.69)	no	LGG_02667
3014	22	24	9			Predicted_galactitol_operon_regulator(Transcriptional_antiterminator),_BglG_family_/_PTS	no	LGG_02669
3225	20	25				PTS_system,_cellobiose-specific_IIC_component(EC_2.7.1.69)	no	LGG_02678
3223	37	26				PTS_system,_IIbc_component	core	LGG_02702
2008	40	28	3			PTS_system,_glucitol/sorbitol-specific_IIAcomponent_(EC_2.7.1.69)	core	LGG_02717

1652	40	28	3			PTS_system,_glucitol/sorbitol-specific_IIBcomponent_and_second_of_two_IIC_components_(EC_2.7.1.69)	core	LGG_02718
205	39	28	3			PTS_system,_glucitol/sorbitol-specific_IICcomponent_(EC_2.7.1.69)	core	LGG_02719
3956	39	29	4			Ascorbate-specific_PTS_system,_EIIB_component(EC_2.7.1.69)	core	LGG_02730
3367	39	29	4			Ascorbate-specific_PTS_system,_EIIC_component	core	LGG_02731
3547	39	29	4			Ascorbate-specific_PTS_system,_EIIA_component(EC_2.7.1.-)	core	LGG_02732
1280	21	30	12			PTS_system,_IId_component	no	LGG_02745
223	21	30	12			PTS_system,_mannose-specific_IIC_component(EC:2.7.1.69_)	no	LGG_02746
3663	21	30	12			PTS_system,_mannose-specific_IIB_component_(EC2.7.1.69)_/_PTS_system,_mannose-specific_IIA_component	no	LGG_02747
277	17	30				PTS_system,_mannose-specific_IIA_component_(EC2.7.1.69)	no	LGG_02748
6	22	31	8			Putative_carbohydrate_PTS_system,_IIC_component(EC_2.7.1.69)	no	LGG_02753
1958	22	31	8			Putative_carbohydrate_PTS_system,_IIB_component(EC_2.7.1.69)	no	LGG_02754
3162	22	31	8			Putative_carbohydrate_PTS_system,_IIA_component(EC_2.7.1.69)	no	LGG_02755

1091	40	32	5			PTS_system,_mannose-specific_IIB_component_(EC2.7.1.69)_/_PTS_system,_mannose-specific_IIA_component	core	LGG_02777
2812	40	32	5			Phosphotransferase_system,mannose/fructose/N-acetylgalactosamine-specific_component	core	LGG_02778
2013	40	32	5			PTS_system,_IIC_component	core	LGG_02779
14	39	33	6			PTS_system,_mannose-specific_IID_component_(EC2.7.1.69)	core	LGG_02836
1166	39	33	6			PTS_system,_mannose-specific_IIC_component_(EC2.7.1.69)	core	LGG_02837
894	39	33	6			PTS_system,_mannose-specific_IIB_component_(EC2.7.1.69)_/_PTS_system,_mannose-specific_IIA_component	core	LGG_02838
3297	40	34				PTS_system,_mannitol-specific_IIA_component_(EC2.7.1.69)	core	LGG_02910
545	40	34				Mannitol_operon_activator,_BglG_family	core	LGG_02911
476	38	34				PTS_system,_mannitol-specific_IIB_component_(EC2.7.1.69)_/_PTS_system,_mannitol-specific_IIC_component	core	LGG_02912
1428	20	35				galactitol_PTS,_EIIB(_EC:2.7.1.69_)	no	0

711	21	35				PTS_system,_galactitol-specific_IIC_component(EC_2.7.1.69)	no	0
1167	23	36	7			PTS_system,_N-acetylgalactosamine-specific_IICcomponent_(EC_2.7.1.69)	no	0
641	23	36	7			PTS_system,hyaluronate-oligosaccharide-specific_IIB_component_(EC	no	0
2613	24	36	7			PTS_system,hyaluronate-oligosaccharide-specific_IIA_component_(EC	no	0
2378	22	36				PTS_system,_N-acetylgalactosamine-specific_IIDcomponent_(EC_2.7.1.69)	no	0
2379	20	37				PTS_system,_IIC_component	no	0
4259	20	37				PTS_system,_mannose-specific_IIB_component_(EC2.7.1.69)_/_PTS_system,_mannose-specific_IIA_component	no	0
590	24	38				PTS_system,_lactose-specific_IIB_component_(EC2.7.1.69)_/_PTS_system,_lactose-specific_IIC_component	no	0
1472	17	39				PTS_system,_fructose-specific_IIA_component_(EC2.7.1.69)	no	0
1474	17	39				PTS_system,_fructose-specific_IIA_component_(EC2.7.1.69)_/_PTS_system,_fructose-specific_IIB_component	no	0

585	24	40				Transcriptional_antiterminator_and_PTS_systemcomponent_IIA	no	0
438	21	41	21			PTS_system,_fructose-specific_IIA_component_(EC2.7.1.69)_/_PTS_system,_fructose-specific_IIB_component	no	0
588	24	41	21			PTS_system,_galactitol-specific_IIC_component(EC_2.7.1.69)	no	0
3000	22	42	10			PTS_system,_sorbose-specific_IIC_component_(EC2.7.1.69)	no	0
1401	22	42	10			PTS_system,_sorbose-specific_IID_component_(EC2.7.1.69)	no	0
3466	22	42	10			PTS_system,_mannose-specific_IIA_component_(EC2.7.1.69)_/_PTS_system,_mannose-specific_IIB_component	no	0
589	18	43				PTS_system,_galactitol-specific_IIC_component(EC_2.7.1.69)	no	0
500	19	44	13			pentitol_PTS_system_enzyme_II_B_component-likeprotein_lmo1972	no	0
3543	19	44	13			PTS_system,_mannitol-specific_IIA_component_(EC2.7.1.69)	no	0
244	19	44	13			PTS_system,_maltose_and_glucose-specific_IICcomponent_(EC_2.7.1.69)_/_PTS_system,_maltose_and	no	0
1558	12	45				PTS_system,_beta-glucoside-	no	0

						specific_IIBcomponent_(EC_2.7.1.69)_/_PTS_system,		
1676	12	46				PTS_system,_maltose_and_glucose-specific_IICcomponent_(EC_2.7.1.69)_/_PTS_system,_maltose_and	no	0
1786	19	47				PTS_system,_fructose-specific_IIA_component_(EC2.7.1.69)	no	0
542	19	47				PTS_system,_fructose-specific_IIB_component_(EC2.7.1.69)_/_PTS_system,_fructose-specific_IIC_component	no	0
1569	15	48				PTS_system,_mannose-specific_IIA_component_(EC2.7.1.69)	no	0
1670	21	49				PTS_system,_galactose-specific_IIA_component(EC_2.7.1.69)	no	0
2303	21	49				PTS_system,_galactose-specific_IIB_component(EC_2.7.1.69)	no	0
2373	6	50	14			PTS_system,_galactosamine-specific_IICcomponent_(EC_2.7.1.69)	no	0
2374	7	50	14			PTS_system,_mannose-specific_IIB_component_(EC2.7.1.69)_/_PTS_system,_mannose-specific_IIA_component	no	0
102	7	50	14			PTS_system,_galactosamine-specific_IIDcomponent_(EC_2.7.1.69)	no	0
396	2	50				N-acetylglucosamine/galactosamine_PTS,_EIIA(EC:2.7.1.69_)	no	0

750	5	51	16			PTS_system,_cellobiose-specific_IIB_component(EC_2.7.1.69)	no	0
3252	5	51	16			PTS_system,_cellobiose-specific_IIA_component(EC_2.7.1.69)	no	0
3933	5	51	16			PTS_system,_cellobiose-specific_IIC_component(EC_2.7.1.69)	no	0
1310	5	52				PTS_system,_mannitol-specific_IIB_component_(EC2.7.1.69)_/_PTS_system,_mannitol-specific_IIC_component	no	0
3939	9	53				PTS_system,_beta-glucoside-specific_IIBcomponent_(EC_2.7.1.69)_/_PTS_system,	no	0
175	4	54	17			PTS_system,_fructose-_and_mannose-inducible_IIDcomponent_(EC_2.7.1.69)	no	0
1663	4	54	17			PTS_system,_fructose-_and_mannose-inducible_IICcomponent_(EC_2.7.1.69)	no	0
3535	4	54	17			PTS_system,_fructose-_and_mannose-inducible_IIBcomponent_(EC_2.7.1.69)	no	0
597	4	54	17			PTS_system,_fructose-_and_mannose-inducibleputative_EII_component	no	0
399	4	54	17			PTS_system,_fructose-_and_mannose-inducible_IIAcomponent_(EC_2.7.1.69)	no	0
1631	6	55	15			PRD/PTS_system_IIA_2_domain_protein	no	0
243	6	55	15			PTS_system,_cellobiose-specific_IIB_component(EC_2.7.1.69)	no	0
2739	6	55	15			PTS_system,_cellobiose-specific_IIC_component(EC_2.7.1.69)	no	0

3456	6	55	15			PTS_system,_cellobiose-specific_IIA_component(EC_2.7.1.69)	no	0
3458	10	56				PTS_system,_galactitol-specific_IIB_component(EC_2.7.1.69)	no	0
2857	9	57				PTS_system,_galactitol-specific_IIA_component(EC_2.7.1.69)	no	0
2858	9	57				PTS_system,_galactitol-specific_IIC_component(EC_2.7.1.69)	no	0
4059	5	58				PTS_system,_fructose-specific_IIA_component_(EC2.7.1.69)	no	0
441	4	59				PTS_system,_Lactose_specific_IIB_subunitsubfamily	no	0
4242	4	59				PTS_system,_mannitol-specific_IIA_component_(EC2.7.1.69)	no	0
2229	1	60				PTS_system,_galactitol-specific_IIC_component(EC_2.7.1.69)	no	0
454	1	60				PTS_system,_galactitol-specific_IIA_component(EC_2.7.1.69)	no	0
2257	2	61				PTS_system,_mannose-specific_IID_component_(EC2.7.1.69)	no	0
468	2	61				PTS_system,_IIa_component	no	0
469	2	62				PTS_system_sorbose_subfamily_IIB_component	no	0
2256	2	62				PTS_system,_mannose-specific_IIC_component_(EC2.7.1.69)	no	0
1150	2	63	18			PTS_system,_mannose-specific_IIC_component_(EC2.7.1.69)	no	0
3700	2	63	18			PTS_system,_mannose/fructose_family_IIDcomponent	no	0

312	3	63	18			PTS_system,_IIB_component	no	0
4028	1	64				PTS_system,_sucrose-specific_IIB_component_(EC2.7.1.69)_/_PTS_system,_sucrose-specific_IIC_component	no	0
495	2	65				PTS_system,_maltose_and_glucose-specific_IICcomponent_(EC_2.7.1.69)_/_PTS_system,_maltose_and	no	0
2597	2	66				PTS_system,_IIc_component mannose	no	0
2598	2	67				PTS_system,_IIb_component mannose	no	0
4220	1	68				PTS_system,_N-acetylgalactosamine-specific_IICcomponent_(EC_2.7.1.69)	no	0
4262	1	68				PTS_system,_N-acetylgalactosamine-_andgalactosamine-specific_IIA_component_(EC_2.7.1.69)	no	0
4437	1	69				PTS_system,_cellobiose-specific_IIA_component(EC_2.7.1.69)	no	0
2507	3			0		Phosphate_transport_ATP-binding_protein_PstB(TC_3.A.1.7.1)	no	0
3166	40			0		Permease_of_the_drug/metabolite_transporter(DMT)_superfamily	core	LGG_00119
2588	40			0		High-affinity_branched-chain_amino_acidtransport_system_permease_protein_LivH_(TC_3.A.1.4.1)	core	LGG_00314

3015	40			0		Branched-chain_amino_acid_transport_systempermease_protein_LivM_(TC_3.A.1.4.1)	core	LGG_00315
2587	40			0		Branched-chain_amino_acid_transport_ATP-bindingprotein_LivG_(TC_3.A.1.4.1)	core	LGG_00316
1074	40			0		Branched-chain_amino_acid_transport_ATP-bindingprotein_LivF_(TC_3.A.1.4.1)	core	LGG_00317
1584	40			0		Substrate-specific_component_YkoE_ofthiamin-regulated_ECF_transporter_for	core	LGG_00363
3295	40			0		Duplicated_ATPase_component_YkoD_of_energizingmodule_of_thiamin-regulated_ECF_transporter_for	core	LGG_00364
2622	39			0		Transmembrane_component_YkoC_of_energizingmodule_of_thiamin-regulated_ECF_transporter_for	no	LGG_00365
3872	39			0		Substrate-specific_component_ThiW_of_predictedthiazole_ECF_transporter	no	LGG_00367
179	29			0		Duplicated_ATPase_component_MtsB_of_energizingmodule_of_methionine-regulated_ECF_transporter	no	LGG_00430
2295	29			0		Transmembrane_component_MtsC_of_energizingmodule_of_methionine-regulated_ECF_transporter	no	LGG_00431
3877	40			0		Cell_division_transporter,_ATP-binding_proteinFtsE_(TC_3.A.5.1.1)	core	LGG_00901

3668	40			0		Phosphate_transport_system_permease_proteinPstC_(TC_3.A.1.7.1)	core	LGG_00907
445	40			0		Phosphate_transport_system_permease_proteinPstA_(TC_3.A.1.7.1)	core	LGG_00908
3084	40			0		Phosphate_transport_ATP-binding_protein_PstB(TC_3.A.1.7.1)	core	LGG_00909
4265	38			0		Phosphate_transport_ATP-binding_protein_PstB(TC_3.A.1.7.1)	core	LGG_00910
316	40			0		Phosphate_transport_system_regulatory_proteinPhoU	core	LGG_00911
3118	40			0		Maltose/maltodextrin_transport_ATP-bindingprotein_MalK_(EC_3.6.3.19)	core	LGG_00954
1645	40			0		Putrescine_transport_ATP-binding_protein_PotA(TC_3.A.1.11.1)	core	LGG_00969
3079	40			0		Glutamate_transport_ATP-binding_protein	core	LGG_01357
1231	40			0		Oligopeptide_transport_system_permease_proteinOppC_(TC_3.A.1.5.1)	core	LGG_01653
3517	40			0		Oligopeptide_transport_ATP-binding_protein_OppF(TC_3.A.1.5.1)	core	LGG_01655
3899	40			0		Oligopeptide_transport_ATP-binding_protein_OppD(TC_3.A.1.5.1)	core	LGG_01656
1596	39			0		Transport_ATP-binding_protein_CydC	core	LGG_01694
2871	40			0		Oligopeptide_transport_ATP-binding_protein_OppF(TC_3.A.1.5.1)	core	LGG_01940

1131	40			0		Oligopeptide_transport_ATP-binding_protein_OppD(TC_3.A.1.5.1)	core	LGG_01941
3031	40			0		Oligopeptide_transport_system_permease_proteinOppB_(TC_3.A.1.5.1)	core	LGG_01944
298	40			0		Cell_division_transporter,_ATP-binding_proteinFtsE_(TC_3.A.5.1.1)	core	LGG_01947
1542	40			0		Transport_ATP-binding_protein_CydC	core	LGG_01948
1200	39			0		Glutamate_transport_membrane-spanning_protein	core	LGG_02006
2420	40			0		Glutamate_transport_membrane-spanning_protein	core	LGG_02007
3235	40			0		Oligopeptide_transport_ATP-binding_protein_OppD(TC_3.A.1.5.1)	core	LGG_02063
1583	40			0		Oligopeptide_transport_system_permease_proteinOppC_(TC_3.A.1.5.1)	core	LGG_02064
3771	40			0		Transport_ATP-binding_protein_CydC	core	LGG_02205
3772	40			0		Transport_ATP-binding_protein_CydD	core	LGG_02206
3231	40			0		Permease_of_the_drug/metabolite_transporter(DMT)_superfamily	core	LGG_02214
1716	40			0		Alkanesulfonates_transport_system_permeaseprotein	core	LGG_02439
3121	40			0		Transmembrane_component_of_general_energizingmodule_of_ECF_transporters	core	LGG_02456

1187	40			0		ATPase_component_of_general_energizing_moduleof_ECF_transporters	core	LGG_02457
792	40			0		ATPase_component_of_general_energizing_moduleof_ECF_transporters	core	LGG_02458
2610	40			0		Lead,_cadmium,_zinc_and_mercury_transportingATPase_(EC_3.6.3.3)_(EC_3.6.3.5)	core	LGG_02803
2692	40			0		Glutamine_transport_system_permease_proteinGlnP_(TC_3.A.1.3.2)	core	LGG_02863
3950	24			1	I	ABC_transporter_ATP-binding_protein	no	0
3141	38			1	I	ComF_operon_protein_A,_DNA_transporter_ATPase	core	LGG_00894
3877	40			1	I	Cell_division_transporter,_ATP-binding_proteinFtsE_(TC_3.A.5.1.1)	core	LGG_00901
2934	40			1	I	Phosphate_ABC_transporter,_periplasmicphosphate-binding_protein_PstS_(TC_3.A.1.7.1)	core	LGG_00906
98	39			2	E	ABC_transporter,_permease_protein_EscB	core	LGG_01776
1917	40			2	E	ABC_transporter,_ATP-binding_protein_EcsA	core	LGG_01777
650	40			3	I	ABC_family_transporter	core	LGG_01080
2330	40			4	I	Manganese_ABC_transporter,_periplasmic-bindingprotein_SitA	core	LGG_02419

3981	40			4	E	Manganese_ABC_transporter,_inner_membranepermease_protein_SitD	core	LGG_02420
384	40			4	E	Manganese_ABC_transporter,_ATP-binding_proteinSitB	core	LGG_02421
2625	40			5	I	Amino_acid_ABC_transporter,_permease_protein	core	LGG_01270
3335	40			5	I	amino_acid_ABC_transporter,_ATP-bindingprotein	core	LGG_01271
1138	40			5	I	amino_acid_ABC_transporter,_amino_acid-bindingprotein	core	LGG_01272
1768	40			6	I	Osmotically_activated_L-carnitine/choline_ABCtransporter,_permease_protein_OpuCD	core	LGG_00075
124	40			6	I	Osmotically_activated_L-carnitine/choline_ABCtransporter,_substrate-binding_protein_OpuCC	core	LGG_00076
1267	40			6	I	Osmotically_activated_L-carnitine/choline_ABCtransporter,_permease_protein_OpuCB	core	LGG_00077
661	40			6	I	Osmotically_activated_L-carnitine/choline_ABCtransporter,_ATP-binding_protein_OpuCA	core	LGG_00078
132	40			7	I	Substrate-specific_component_RibU_of_riboflavinECF_transporter	core	LGG_01383
3287	40			8	I	ABC_transporter,_permease_protein	core	LGG_02920

999	40			9	I	D-serine/D-alanine/glycine_transporter	core	LGG_00996
1584	40			10	I	Substrate-specific_component_YkoE_ofthiamin-regulated_ECF_transporter_for	no	LGG_00363
3295	40			10	I	Duplicated_ATPase_component_YkoD_of_energizingmodule_of_thiamin-regulated_ECF_transporter_for	no	LGG_00364
2622	39			10	I	Transmembrane_component_YkoC_of_energizingmodule_of_thiamin-regulated_ECF_transporter_for	no	LGG_00365
211	40			11	I	Amino_acid_transporter	core	LGG_01632
3898	39			11	I	ABC_transporter,_ATP-binding_and_permeaseprotein	core	LGG_01634
3872	39			12	E	Substrate-specific_component_ThiW_of_predictedthiazole_ECF_transporter	no	LGG_00367
3880	40			13	E	transporter_associated_with_VraSR	core	LGG_01712
3061	40			14	E	amino_acid_transporter	core	LGG_00371
3639	40			15	E	ABC_transporter_ATPase_component	core	LGG_02349
1632	19			16	E	benzoate_MFS_transporter_BenK	no	0
2358	19			16	E	benzoate_MFS_transporter_BenK	no	0
581	40			17	I	ABC_transporter_permease_component	core	LGG_00489

3735	40			17	I	ABC_transporter_ATP-binding_protein	core	LGG_00490
3902	40			18	E	COG0488:_ATPase_components_of_ABC_transporterswith_duplicated_ATPase_domains	core	LGG_00638
2353	40			19	I	ABC_transporter_permease_protein	core	LGG_01426
2354	39			19	I	ABC_transporter_ATP-binding_protein_YvcR	core	LGG_01427
3231	40			20	E	Permease_of_the_drug/metabolite_transporter(DMT)_superfamily	core	LGG_02214
2879	40			21	E	ABC_transporter,_permease_protein	core	LGG_00032
1044	39			22	E	ammonium_transporter_family_protein	core	LGG_00029
431	32			23	E	ABC-2_type_transporter	core	LGG_00025
2646	29			24	E	Substrate-specific_component_MtsA_ofmethionine-regulated_ECF_transporter	no	LGG_00429
179	29			24	I	Duplicated_ATPase_component_MtsB_of_energizingmodule_of_methionine-regulated_ECF_transporter	no	LGG_00430
2295	29			24	I	Transmembrane_component_MtsC_of_energizingmodule_of_methionine-regulated_ECF_transporter	no	LGG_00431
3160	40			24	I	Mg2+/citrate_complex_transporter	core	LGG_01921
534	40			25	E	Bacteriocin_ABC-transporter,_putativecomponent	core	LGG_02385

969	40			25	E	Bacteriocin_ABC-transporter,_ATP-binding_andpermease_component	core	LGG_02386
3748	40			26	I	Glycerol-3-phosphate_ABC_transporter,periplasmic_glycerol-3-phosphate-binding_protein_(TC	core	LGG_02768
2644	40			26	I	Glycerol-3-phosphate_ABC_transporter,_permeaseprotein_UgpE_(TC_3.A.1.1.3)	core	LGG_02769
598	40			26	I	Glycerol-3-phosphate_ABC_transporter,_permeaseprotein_UgpA_(TC_3.A.1.1.3)	core	LGG_02770
890	40			26	I	Glycerol-3-phosphate_ABC_transporter,ATP-binding_protein_UgpC_(TC_3.A.1.1.3)	core	LGG_02771
2893	40			27	E	COG0488:_ATPase_components_of_ABC_transporterswith_duplicated_ATPase_domains	core	LGG_00300
696	40			27	E	YbbL_ABC_transporter_ATP-binding_protein	core	LGG_02565
1382	40			28	I	Zinc_ABC_transporter,_ATP-binding_protein_ZnuC	core	LGG_00285
3314	40			28	I	Zinc_ABC_transporter,_inner_membrane_permeaseprotein_ZnuB	core	LGG_00286
326	40			28	E	Substrate-specific_component_QueT_(COG4708)_ofpredicted_queuosine-regulated_ECF_transporter	core	LGG_02532

2571	40			29	I	Spermidine_Putrescine_ABC_transporter_permeasecomponent_potC_(TC_3.A.1.11.1)	core	LGG_00970
2016	40			29	I	ABC_transporter,_periplasmic_spermidineputrescine-binding_protein_PotD_(TC_3.A.1.11.1)	core	LGG_00972
2335	40			30	I	transporter	core	LGG_00801
880	40			31	I	COG0488:_ATPase_components_of_ABC_transporterswith_duplicated_ATPase_domains	core	LGG_01396
3027	40			31	I	Substrate-specific_component_BioY_of_biotin_ECFtransporter	core	LGG_02223
2866	40			32	E	Substrate-specific_component_FolT_of_folate_ECFtransporter	core	LGG_02255
2614	39			33	E	ABC_transporter,_ATP-binding_protein	core	LGG_01986
1041	40			33	E	Amino_acid_transporter	core	LGG_01987
2484	40			34	E	Ammonium_transporter	core	LGG_00633
32	40			35	E	ABC_transporter_ATPase_component	core	LGG_00622
2485	40			36	E	COG0488:_ATPase_components_of_ABC_transporterswith_duplicated_ATPase_domains	core	LGG_00612
1694	40			37	E	Amino_acid_transporter	core	LGG_00605

2616	39			38	E	Permease_of_the_drug/metabolite_transporter(DMT)_superfamily	core	LGG_00588
3818	40			38	E	Substrate-specific_component_PdxT_of_predictedpyridoxine_ECF_transporter	core	LGG_00594
934	15			39	E	major_facilitator_family_transporter	no	LGG_00465
3017	40			40	I	NrdR-regulated_deoxyribonucleotide_transporter,PnuC-like	core	LGG_01221
60	40			40	I	Oligopeptide_ABC_transporter,_periplasmicoligopeptide-binding_protein_OppA_(TC_3.A.1.5.1)	core	LGG_01224
609	40			41	E	Substrate-specific_component_QueT_(COG4708)_ofpredicted_queuosine-regulated_ECF_transporter	core	LGG_02057
2506	40			42	I	Methionine_ABC_transporter_substrate-bindingprotein	core	LGG_01195
1711	40			42	I	Methionine_ABC_transporter_ATP-binding_protein	core	LGG_01197
1175	40			42	I	Methionine_ABC_transporter_permease_protein	core	LGG_01198
217	40			43	I	Amino_acid_transporter	core	LGG_00824
3290	17			44	I	ABC_transporter_ATP-binding_protein	no	0
2842	35			44	I	Uncharacterized_ABC_transporter,_permeasecomponent	no	LGG_01939

2845	40			44	I	Oligopeptide_ABC_transporter,_periplasmicoligopeptide-binding_protein_OppA_(TC_3.A.1.5.1)	no	LGG_01945
3121	40			45	E	Transmembrane_component_of_general_energizingmodule_of_ECF_transporters	core	LGG_02456
1187	40			45	E	ATPase_component_of_general_energizing_moduleof_ECF_transporters	core	LGG_02457
792	40			45	E	ATPase_component_of_general_energizing_moduleof_ECF_transporters	core	LGG_02458
2346	40			46	I	ABC_transporter_permease_protein	core	LGG_01852
3330	40			46	I	ABC_transporter_substrate-binding_protein	core	LGG_01853
3610	40			47	I	Substrate-specific_component_PanT_of_predictedpantothenate_ECF_transporter	core	LGG_01847
3516	40			47	I	ABC_transporter_ATP-binding_protein	core	LGG_01851
728	40			48	E	Major_myo-inositol_transporter_IolT	core	LGG_00261
3864	40			48	E	Substrate-specific_component_ThiT_of_thiaminECF_transporter	core	LGG_01815
2277	40			49	E	Substrate-specific_component_CbrT_of_predictedcobalamin_ECF_transporter	core	LGG_02294

1015	39			49	E	Additional_lipoprotein_component_of_predictedcobalamin_ECF_transporter	core	LGG_02295
1637	40			50	E	ABC_transporter_ATP-binding_protein_YvcR	core	LGG_00033
272	39			50	E	ABC_transporter_permease_protein	core	LGG_00131
3166	40			51	E	Permease_of_the_drug/metabolite_transporter(DMT)_superfamily	core	LGG_00119
2305	40			52	I	Xyloside_transporter_XynT	core	LGG_00557
265	40			53	I	amino_acid_ABC_transporter,_permease_protein	core	LGG_00573
3980	40			54	I	Niacin_transporter_NiaP	core	LGG_02281
187	40			55	I	ABC_transporter,_permease_protein	core	LGG_00978
188	38			55	I	ABC_transporter,_ATP-binding_protein	core	LGG_00979
2469	40			56	I	Alkanesulfonates_ABC_transporter_ATP-bindingprotein_/_Sulfonate_ABC_transporter,_ATP-binding_subunit	core	LGG_02438
3721	39			57	I	ABC_transporter_ATP-binding_protein	core	LGG_00550
3548	40			58	I	ABC_transporter,_ATP-binding_protein	core	LGG_00961
2742	28			59	I	ABC_transporter,_ATP-binding_protein	no	0

2516	40			60	I	ABC_transporter_ATP-binding_protein	core	LGG_02451
172	40			61	I	putative_ABC_transporter,_ATP-binding_protein	core	LGG_02442
1298	40			62	I	ABC_transporter,_ATP-binding_protein	core	LGG_02245
3030	39			62	I	ABC_transporter_ATP-binding_protein_uup	core	LGG_02248
1499	40			64	I	Putative_ABC_transporter_ATP-binding_protein,spy1790_homolog	core	LGG_01507
2440	39			65	I	ABC_transporter_related	no	LGG_01547
2732	12			66	I	ABC_transporter,_ATP-binding_protein (5genes)	no	0
2196	39			67	I	ABC_transporter,_ATP-binding_protein	core	LGG_00205
3260	18			68	I	ABC_transporter,_ATP-binding_protein	no	0
3474	40			68	I	Oligopeptide_ABC_transporter,_periplasmicoligopeptide-binding_protein_OppA_(TC_3.A.1.5.1)	core	LGG_00201
47	40			69	E	ABC_transporter,_ATP-binding_protein	core	LGG_02187
2746	40			70	E	Putative_ABC_transporter_ATP-binding_protein,spy1790_homolog	core	LGG_02195
3025	40			70	E	Putative_ABC_transporter_(ATP-binding_protein),spy1791_homolog	core	LGG_02196
3888	39			71	I	ABC_transporter,_ATP-binding_protein	core	LGG_01240

1926	39			72	I	ABC_transporter,_ATP-binding_protein	core	LGG_01930
4278	40			73	I	ABC_transporter_ATP-binding/membrane_spanningprotein_-_multidrug_resistance	core	LGG_02936
1143	34			74	I	ABC_transporter_ATP-binding_protein	core	LGG_02623
145	40			75	I	ABC_transporter,_ATP-binding/permease_protein	core	LGG_02626
2392	29			76	I	ABC_transporter,_ATP-binding_protein (5 genes)	no	0
1030	38			77	I	ABC_transporter,_ATP-binding_protein	core	LGG_01796
2834	40			78	I	ABC_transporter_substrate-binding_protein	core	LGG_00687
3124	40			78	I	Methionine_ABC_transporter_ATP-binding_protein	core	LGG_00691
1207	39			79	I	ABC_transporter,_ATP-binding_protein	core	LGG_02318
357	40			80	I	Zinc_ABC_transporter,_periplasmic-bindingprotein_ZnuA	core	LGG_02423
3649	39			81	I	Oligopeptide_ABC_transporter,_periplasmicoligopeptide-binding_protein_OppA_(TC_3.A.1.5.1)	no	LGG_00359
1095	38			82	I	Methionine_ABC_transporter_substrate-bindingprotein	no	LGG_00500
1385	40			83	I	Amino_acid_ABC_transporter,_aminoacid-binding/permease_protein	core	LGG_02816

2734	40			84	I	L-proline_glycine_betaine_binding_ABCtransporter_protein_ProX_(TC_3.A.1.12.1)_/_Osmotic	core	LGG_02617
1158	39			85	I	Oligopeptide_ABC_transporter,_periplasmicoligopeptide-binding_protein_OppA_(TC_3.A.1.5.1)	core	LGG_01652
3300	40			86	I	Methionine_ABC_transporter_substrate-bindingprotein	core	LGG_02926
1386	40			87	I	Glycerol-3-phosphate_ABC_transporter,periplasmic_glycerol-3-phosphate-binding_protein_(TC	core	LGG_02851
3850	13			88	I	oligogalacturonide_transporter	no	0
2912	40			88	I	Oligopeptide_ABC_transporter,_periplasmicoligopeptide-binding_protein_OppA_(TC_3.A.1.5.1)	core	LGG_02801
1529	2			89	I	Oligopeptide_ABC_transporter,_periplasmicoligopeptide-binding_protein_OppA_(TC_3.A.1.5.1)	no	0
1341	40			90	I	Glutamine_ABC_transporter,_periplasmicglutamine-binding_protein_(TC_3.A.1.3.2)	core	LGG_02008
1978	40			91	I	Glycerol-3-phosphate_ABC_transporter,periplasmic_glycerol-3-phosphate-binding_protein_(TC	core	LGG_00951

3940	15			92	I	ABC_transporter,_ATP-binding_protein, mannose/fructose/N-acetylgalactosamine (3 genes)	no	LGG_02612
1011	16			93	I	ABC_transporter,_ATP-binding_protein	no	0
2056	28			93	I	Phospholipid-lipopolysaccharide_ABCtransporter	no	LGG_02336
2139	16			93	I	ABC_transporter, permease	no	
103	12			94	I	ABC_transporter,_ATP-binding_protein	no	LGG_00630
2535	16			95	I	ABC_transporter_ATP-binding_protein	no	LGG_00511
3147	13			96	I	putative_amino_acid_ABC_transporter,ATP-binding_protein	no	LGG_00478
3146	13			96	I	putative_amino_acid_ABC_transporter,periplasmic_amino_acid-binding_protein	no	LGG_00479
28	13			96	I	putative_amino_acid_ABC_transporter,_permeaseprotein	no	LGG_00481
656	23			97	I	ABC_transporter_ATP-binding_protein taurine	no	LGG_00172
72	23			97	I	taurine_transport_system_permease_protein_tauC	no	LGG_00174
1009	15			98	I	ABC_transporter,_permease_protein glutamate	no	0
3074	15			98	I	ABC_transporter,_ATP-binding_protein glutamate	no	0

1405	18			99	I	putative_ABC_transporter_ATP-binding_protein	no	0
231	3			100	I	Polar_amino_acid_ABC_uptake_transportermembrane-spanning_protein	no	0
233	3			100	I	ABC_transporter_ATPase_component	no	0
235	3			100	I	Glutamine_ABC_transporter_substrate-bindingprotein	no	0
2480	3			100	I	ABC_transporter_permease_component	no	0
450	2			101	I	Maltose/maltodextrin_ABC_transporter,_permeaseprotein_MalF	no	0
2315	2			101	I	Maltose/maltodextrin_ABC_transporter,_substratebinding_periplasmic_protein_ MalE	no	0
2313	6			101	I	Multiple_sugar_ABC_transporter,_ATP-bindingprotein	no	0
4466	1			102	I	ABC_transporter,_ATP-binding_protein	no	0
4202	1			103	I	ABC_transporter_ATP-binding_protein	no	0
2990	3			104	I	Sugar_transporter [IS transfer from zeae, with it's transporter and hydrolase]	no	0
2607	4			105	I	ABC_transporter_ATP-binding_protein	no	0
302	30			106	I	Ribose_ABC_transport_system,_high_affinitypermease_RbsD_(TC_3.A.1.2.1)	no	0

2292	30			106	I	Ribose_ABC_transport_system,_ATP-bindingprotein_RbsA_(TC_3.A.1.2.1)	no	0
3365	30			106	I	Ribose_ABC_transport_system,_permease_proteinRbsC_(TC_3.A.1.2.1)	no	0
1434	40			106	I	Ribose_ABC_transport_system,_periplasmicribose-binding_protein_RbsB_(TC_3.A.1.2.1)	core	0
1092	40			106	I	Ribose_operon_repressor	core	LGG_00372
4333	1					transport_protein	no	0
2561	2					Negative_transcriptional_regulator-coppertransport_operon	no	0
1371	4					Probable_cadmium-transporting_ATPase_(EC3.6.3.3)	no	0
3937	6					transport_protein	no	0
2396	8					Manganese_transport_protein_MntH	no	0
3064	16					sodium-dependent_transporter_(huNaDC-1)	no	0
1204	18					sugar_transport_protein	no	0
3834	19					Transport_protein_SgaT,_putative	no	0
1044	39					ammonium_transporter_family_protein	core	LGG_00029
3519	40					Putative_melibiose:Na(+)_transport_protein	core	LGG_00058

2859	40					Lead,_cadmium,_zinc_and_mercury_transportingATPase_(EC_3.6.3.3)_(EC_3.6.3.5)	core	LGG_00082
3164	40					Lead,_cadmium,_zinc_and_mercury_transportingATPase_(EC_3.6.3.3)_(EC_3.6.3.5)	core	LGG_00135
2533	39					Transport_protein_SgaT,_putative	core	LGG_00194
728	40					Major_myo-inositol_transporter_IolT	core	LGG_00261
2923	40					Branched-chain_amino_acid_transport_systemcarrier_protein	core	LGG_00292
3061	40					amino_acid_transporter	core	LGG_00371
7	7					FIG006427:_Putative_transport_system_permeaseprotein	no	LGG_00423
2646	29					Substrate-specific_component_MtsA_ofmethionine-regulated_ECF_transporter	no	LGG_00429
113	14					cation-transporting_ATPase,_E1-E2_family	no	LGG_00456
934	15					major_facilitator_family_transporter	no	LGG_00465
2029	40					Probable_transport_protein	core	LGG_00555
2305	40					Xyloside_transporter_XynT	core	LGG_00557
1964	11					transport_protein	no	LGG_00565

1189	40					Branched-chain_amino_acid_transport_systemcarrier_protein	core	LGG_00575
2616	39					Permease_of_the_drug/metabolite_transporter(DMT)_superfamily	core	LGG_00588
3818	40					Substrate-specific_component_PdxT_of_predictedpyridoxine_ECF_transporter	core	LGG_00594
1694	40					Amino_acid_transporter	core	LGG_00605
3905	40					Histidine_transport_protein_(permease)	core	LGG_00616
2484	40					Ammonium_transporter	core	LGG_00633
2824	40					Peptide_transport_system_permease_protein_sapB(TC_3.A.1.5.5)	core	LGG_00635
3644	39					cation-transporting_ATPase	core	LGG_00693
2975	40					Magnesium_and_cobalt_transport_protein_CorA	core	LGG_00724
676	40					Cation-transporting_ATPase	core	LGG_00735
2335	40					transporter	core	LGG_00801
217	40					Amino_acid_transporter	core	LGG_00824
2411	40					Magnesium_and_cobalt_transport_protein_corA	core	LGG_00875
3199	40					FIG000605:_protein_co-occurring_with_transportsystems_(COG1739)	core	LGG_00893

3141	38					ComF_operon_protein_A,_DNA_transporter_ATPase	core	LGG_00894
754	39					Transcriptional_regulator,_MerR_family,_nearpolyamine_transporter	core	LGG_00968
976	40					cation-transporting_ATPase,_E1-E2_family	core	LGG_00976
999	40					D-serine/D-alanine/glycine_transporter	core	LGG_00996
3017	40					NrdR-regulated_deoxyribonucleotide_transporter,PnuC-like	core	LGG_01221
2102	40					transport	core	LGG_01294
3531	40					ATPase,_P-type_(transporting),_HAD_superfamily,subfamily_IC	core	LGG_01319
1058	40					Late_competence_protein_ComEC,_DNA_transport	core	LGG_01335
132	40					Substrate-specific_component_RibU_of_riboflavinECF_transporter	core	LGG_01383
3846	39					4-oxalocrotonate_tautomerase_(EC_5.3.2.-)	core	LGG_01506
211	40					Amino_acid_transporter	core	LGG_01632
3880	40					transporter_associated_with_VraSR	core	LGG_01712
2562	40					Lead,_cadmium,_zinc_and_mercury_transportingATPase_(EC_3.6.3.3)_(EC_3.6.3.5)	core	LGG_01801

3186	40					Negative_transcriptional_regulator-coppertransport_operon	core	LGG_01802
3864	40					Substrate-specific_component_ThiT_of_thiaminECF_transporter	core	LGG_01815
3610	40					Substrate-specific_component_PanT_of_predictedpantothenate_ECF_transporter	core	LGG_01847
96	40					Calcium-transporting_ATPase_(EC_3.6.3.8)/EC_number=3.6.3.8	core	LGG_01855
2475	40					Ferrous_iron_transport_protein_B	core	LGG_01878
1336	39					Fe2+_transport_system_protein_A	core	LGG_01879
3160	40					Mg2+/citrate_complex_transporter	core	LGG_01921
3538	40					Oligopeptide_transport_system_permease_proteinOppC_(TC_3.A.1.5.1)	core	LGG_01943
3860	38					Manganese_transport_protein_MntH	core	LGG_01981
1041	40					Amino_acid_transporter	core	LGG_01987
609	40					Substrate-specific_component_QueT_(COG4708)_ofpredicted_queuosine-regulated_ECF_transporter	core	LGG_02057
3756	40					Oligopeptide_transport_system_permease_proteinOppB_(TC_3.A.1.5.1)	core	LGG_02065

3027	40					Substrate-specific_component_BioY_of_biotin_ECFtransporter	core	LGG_02223
2866	40					Substrate-specific_component_FolT_of_folate_ECFtransporter	core	LGG_02255
2064	40					transport_protein	core	LGG_02278
3980	40					Niacin_transporter_NiaP	core	LGG_02281
2190	40					Mg(2+)_transport_ATPase,_P-type_(EC_3.6.3.2)/EC_number=3.6.3.2	core	LGG_02290
2277	40					Substrate-specific_component_CbrT_of_predictedcobalamin_ECF_transporter	core	LGG_02294
1015	39					Additional_lipoprotein_component_of_predictedcobalamin_ECF_transporter	core	LGG_02295
3478	40					Manganese_transport_protein_MntH	core	LGG_02411
1236	40					Cation_transport_ATPase	core	LGG_02412
326	40					Substrate-specific_component_QueT_(COG4708)_ofpredicted_queuosine-regulated_ECF_transporter	core	LGG_02532
1153	40					transport_protein_(putative)	core	LGG_02839

Supplementary Table.8. KEGG - KAAS created metabolic maps summary.

KEGG Map	Metabolism overview	pan	core	clade1	clade2	clade3	clade4	clade5	clade6	clade7	clade8	Variable
10	Glycolysis / Gluconeogenesis	26	24	24	25	25	24	25	24	25	25	
20	Citrate cycle (TCA cycle)	11	10	11	11	11	10	11	10	11	11	
30	Pentose phosphate pathway	23	17	19	20	20	21	20	21	20	20	
40	Pentose and glucuronate interconversions	22	10	16	15	15	12	15	15	15	15	
51	Fructose and mannose metabolism	33	20	27	29	29	28	29	29	29	29	
52	Galactose metabolism	31	18	21	28	30	23	28	31	28	30	*
53	Ascorbate and aldarate metabolism	5	4	4	5	5	4	5	5	5	5	
61	Fatty acid biosynthesis	11	11	11	11	11	11	11	11	11	11	
71	Fatty acid metabolism	2	2	2	2	2	2	2	2	2	2	
72	Synthesis and degradation of ketone bodies	3	2	2	2	2	3	2	2	2	2	
2010	ABC transporters	74	57	68	63	62	65	63	62	63	62	*
2020	Two-component system	27	24	27	26	26	25	26	25	26	26	

2030	Bacterial chemotaxis	1	0	1	1	1	1	1	1	1	1	
2060	Phosphotransferase system (PTS)	43	26	30	42	43	30	42	38	42	42	*
3010	Ribosome	52	52	52	52	52	52	52	52	52	52	
3013	RNA transport	2	2	2	2	2	2	2	2	2	2	
3018	RNA degradation	8	8	8	8	8	8	8	8	8	8	
3020	RNA polymerase	5	5	5	5	5	5	5	5	5	5	
3030	DNA replication	14	14	14	14	14	14	14	14	14	14	
3060	Protein export	10	10	10	10	10	10	10	10	10	10	
3070	Bacterial secretion system	9	8	9	8	9	8	8	9	8	9	
3410	Base excision repair	11	11	11	11	11	11	11	11	11	11	
3420	Nucleotide excision repair	7	7	7	7	7	7	7	7	7	7	
3430	Mismatch repair	17	16	16	17	16	16	17	16	17	16	
3440	Homologous recombination	19	19	19	19	19	19	19	19	19	19	
130	Ubiquinone and other terpenoid-quinone biosynthesis	6	1	3	3	3	4	3	2	3	3	*

190	Oxidative phosphorylation	13	12	13	13	13	12	13	12	13	13	
195	Photosynthesis	8	8	8	8	8	8	8	8	8	8	
230	Purine metabolism	47	47	47	47	47	46	46	46	47	46	
240	Pyrimidine metabolism	41	41	40	42	42	40	41	41	42	41	
250	Alanine, aspartate and glutamate metabolism	18	18	18	18	18	18	18	18	18	18	
253	Tetracycline biosynthesis	4	4	4	4	4	4	4	4	4	4	
260	Glycine, serine and threonine metabolism	16	12	14	15	16	13	15	15	15	16	*
270	Cysteine and methionine metabolism	16	14	16	14	16	14	14	16	14	16	
280	Valine, leucine and isoleucine degradation	9	6	6	6	6	6	6	6	6	6	*
290	Valine, leucine and isoleucine biosynthesis	2	2	2	2	2	2	2	2	2	2	
300	Lysine biosynthesis	13	13	13	13	13	13	13	13	13	13	
310	Lysine degradation	1	1	1	1	1	1	1	1	1	1	
311	Penicillin and cephalosporin biosynthesis	1	1	1	1	1	1	1	1	1	1	

312	beta-Lactam resistance	1	1	1	1	1	1	1	1	1	1	
330	Arginine and proline metabolism	14	12	13	13	13	13	13	14	13	13	
340	Histidine metabolism	11	11	11	11	11	11	11	11	11	11	
350	Tyrosine metabolism	5	5	5	5	5	5	5	5	5	5	
360	Phenylalanine metabolism	3	3	3	3	3	3	3	3	3	3	
361	Chlorocyclohexane and chlorobenzene degradation	1	1	1	1	1	1	1	1	1	1	
362	Benzoate degradation	4	4	4	4	4	4	4	4	4	4	
363	Bisphenol degradation	1	1	1	1	1	1	1	1	1	1	
380	Tryptophan metabolism	2	2	2	2	2	2	2	2	2	2	
400	Phenylalanine, tyrosine and tryptophan biosynthesis	8	7	7	7	7	8	7	7	7	7	
401	Novobiocin biosynthesis	1	1	1	1	1	1	1	1	1	1	
410	beta-Alanine metabolism	1	1	1	1	1	1	1	1	1	1	
430	Taurine and hypotaurine metabolism	2	2	2	2	2	2	2	2	2	2	
450	Selenocompound metabolism	7	7	7	7	7	7	7	7	7	7	

460	Cyanoamino acid metabolism	3	3	3	3	3	3	3	3	3	3	
471	D-Glutamine and D-glutamate metabolism	2	2	2	2	2	2	2	2	2	2	
473	D-Alanine metabolism	4	4	4	4	4	4	4	4	4	4	
480	Glutathione metabolism	7	7	7	7	7	7	7	7	7	7	
500	Starch and sucrose metabolism	23	21	22	21	21	22	21	21	21	21	
510	N-Glycan biosynthesis	1	1	1	1	1	1	1	1	1	1	
511	Other glycan degradation	4	1	2	2	3	2	2	4	2	3	
520	Amino sugar and nucleotide sugar metabolism	28	21	27	23	25	26	23	25	23	26	
521	Streptomycin biosynthesis	8	4	5	8	8	8	8	8	8	8	*
523	Polyketide sugar unit biosynthesis	4	0	1	4	4	4	4	4	4	4	*
524	Butirosin and neomycin biosynthesis	1	1	1	1	1	1	1	1	1	1	
531	Glycosaminoglycan degradation	1	1	1	1	1	1	1	1	1	1	
540	Lipopolysaccharide biosynthesis	2	0	2	0	2	0	0	2	0	2	
550	Peptidoglycan biosynthesis	15	15	15	15	15	15	15	15	15	15	

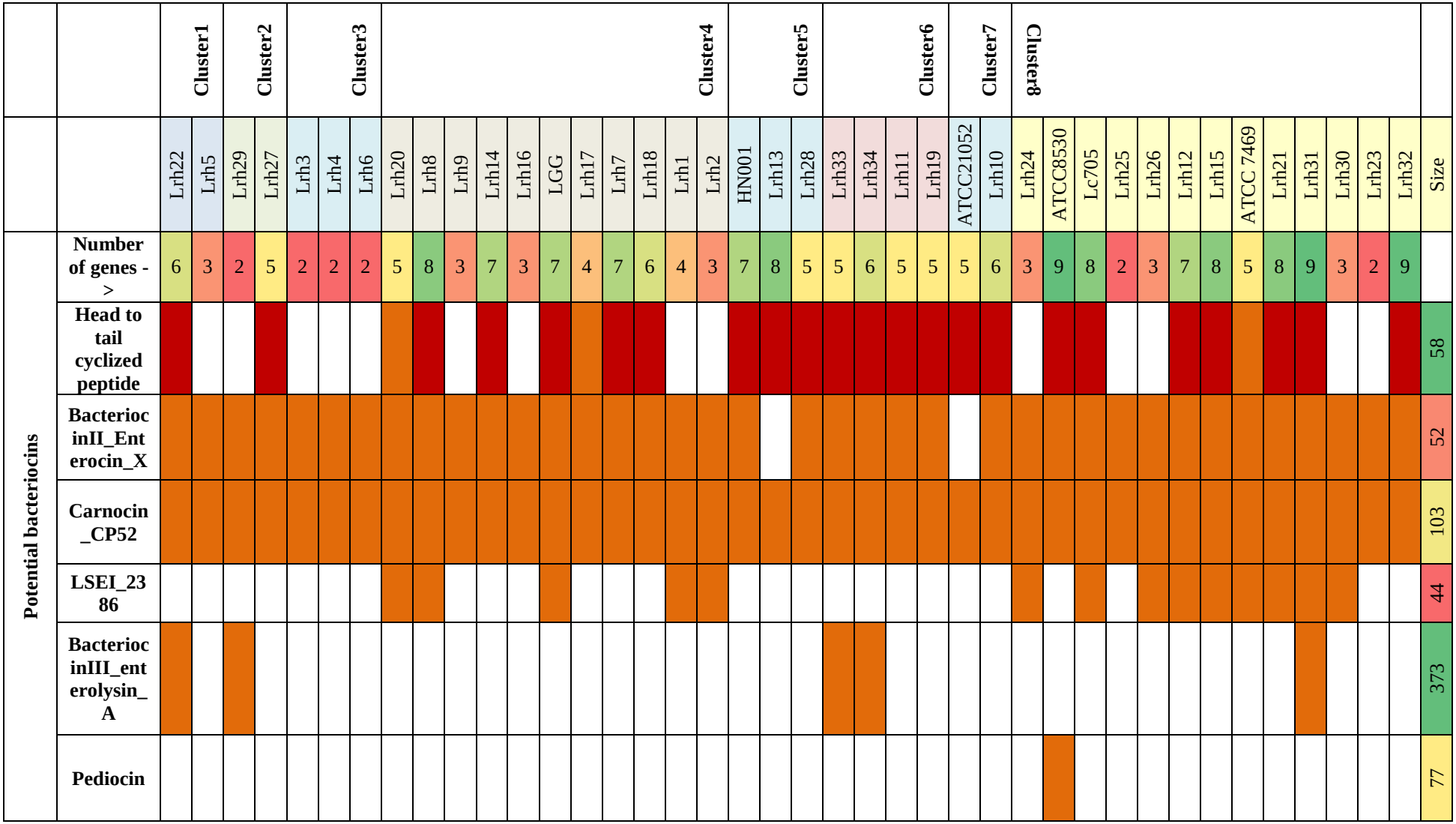
561	Glycerolipid metabolism	9	9	9	9	9	9	9	9	9	9	
562	Inositol phosphate metabolism	10	10	10	10	10	10	10	10	10	10	
564	Glycerophospholipid metabolism	9	9	9	9	9	9	9	9	9	9	
590	Arachidonic acid metabolism	1	1	1	1	1	1	1	1	1	1	
591	Linoleic acid metabolism	1	1	1	1	1	1	1	1	1	1	
600	Sphingolipid metabolism	3	1	2	2	3	1	2	3	2	3	*
603	Glycosphingolipid biosynthesis - globo series	1	1	1	1	1	1	1	1	1	1	
620	Pyruvate metabolism	25	24	24	24	24	24	24	25	24	24	*
621	Dioxin degradation	2	2	2	2	2	2	2	2	2	2	
622	Xylene degradation	3	3	3	3	3	3	3	3	3	3	
623	Toluene degradation	2	1	2	2	2	1	2	1	2	2	*
624	Polycyclic aromatic hydrocarbon degradation	1	1	1	1	1	1	1	1	1	1	
625	Chloroalkane and chloroalkene degradation	3	3	3	3	3	3	3	3	3	3	

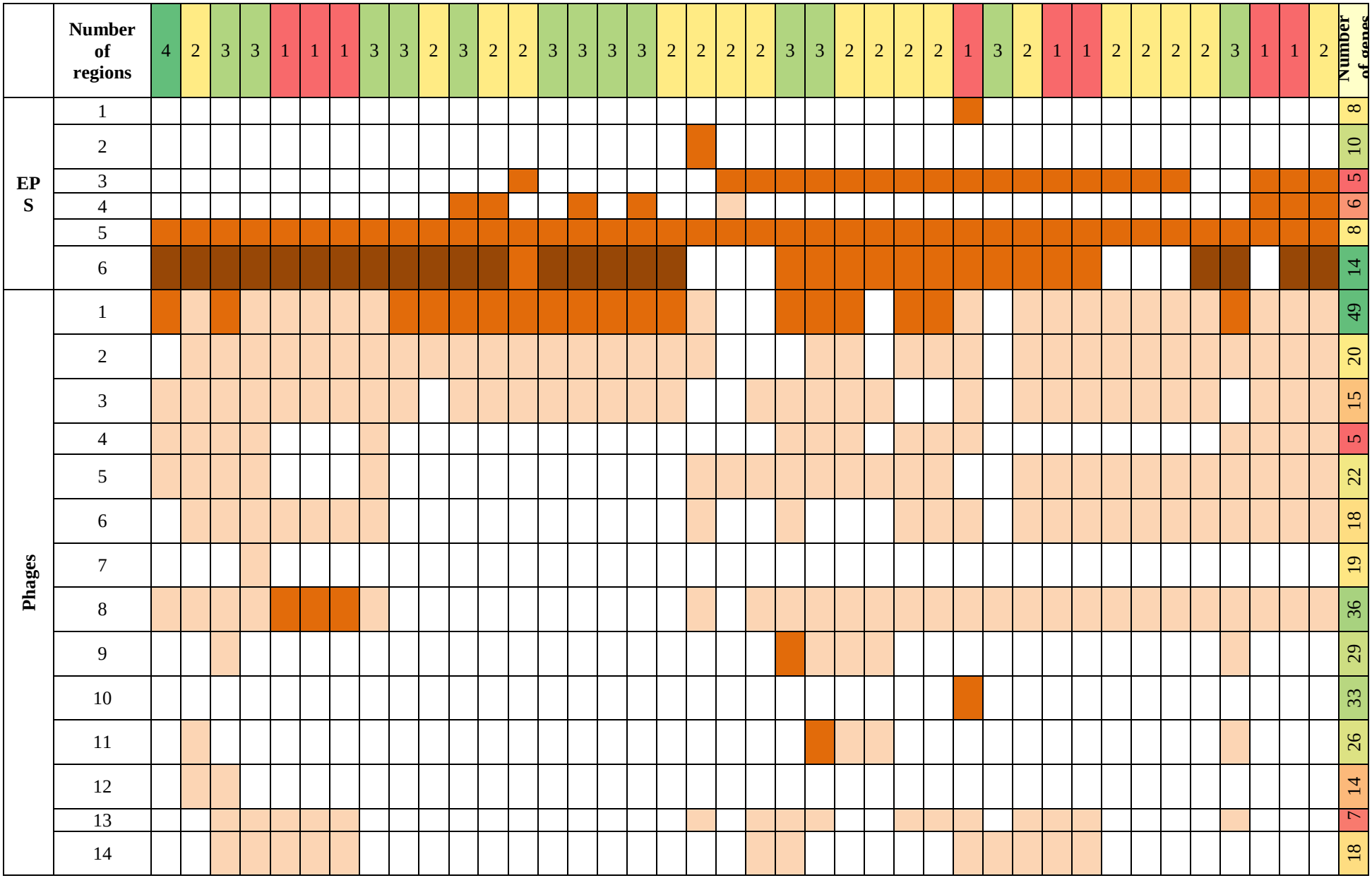
626	Naphthalene degradation	1	1	1	1	1	1	1	1	1	1	
627	Aminobenzoate degradation	3	3	3	3	3	3	3	3	3	3	
630	Glyoxylate and dicarboxylate metabolism	7	5	7	7	7	6	7	6	7	7	
640	Propanoate metabolism	11	10	10	10	10	11	10	10	10	10	
643	Styrene degradation	1	1	1	1	1	1	1	1	1	1	
650	Butanoate metabolism	13	10	11	11	11	10	11	10	11	11	
660	C5-Branched dibasic acid metabolism	2	2	2	2	2	2	2	2	2	2	
670	One carbon pool by folate	10	10	10	10	10	10	10	10	10	10	
680	Methane metabolism	14	9	10	12	12	9	12	12	12	12	*
710	Carbon fixation in photosynthetic organisms	12	10	10	10	10	11	10	10	10	10	
720	Carbon fixation pathways in prokaryotes	13	12	13	13	13	12	13	12	13	13	
730	Thiamine metabolism	8	6	8	8	8	8	8	8	8	8	
740	Riboflavin metabolism	1	1	1	1	1	1	1	1	1	1	
750	Vitamin B6 metabolism	3	2	3	2	2	2	2	2	2	2	

760	Nicotinate and nicotinamide metabolism	7	6	7	6	6	6	6	6	6	6	
770	Pantothenate and CoA biosynthesis	8	7	7	7	8	7	7	7	7	8	
780	Biotin metabolism	4	4	4	4	4	4	4	4	4	4	
785	Lipoic acid metabolism	1	1	1	1	1	1	1	1	1	1	
790	Folate biosynthesis	2	2	2	2	2	2	2	2	2	2	
860	Porphyrin and chlorophyll metabolism	3	3	3	3	3	3	3	3	3	3	
900	Terpenoid backbone biosynthesis	12	11	11	11	11	11	11	12	11	11	
908	Zeatin biosynthesis	1	1	1	1	1	1	1	1	1	1	
910	Nitrogen metabolism	8	5	5	5	5	5	5	5	5	5	
920	Sulfur metabolism	5	5	9	8	6	8	8	6	8	6	*
930	Caprolactam degradation	1	1	1	1	1	1	1	1	1	1	
940	Phenylpropanoid biosynthesis	1	1	1	1	1	1	1	1	1	1	
944	Flavone and flavonol biosynthesis	1	1	1	1	1	1	1	1	1	1	
960	Tropane, piperidine and pyridine alkaloid biosynthesis	1	1	1	1	1	1	1	1	1	1	

970	Aminoacyl-tRNA biosynthesis	26	26	26	26	26	26	26	26	26	26	
983	Drug metabolism - other enzymes	8	7	7	8	8	7	8	8	8	8	
1040	Biosynthesis of unsaturated fatty acids	2	2	2	2	2	2	2	2	2	2	
1051	Biosynthesis of ansamycins	1	0	0	0	0	1	0	0	0	0	
1054	Nonribosomal peptide structures	1	1	1	1	1	1	1	1	1	1	
1055	Biosynthesis of vancomycin group antibiotics	1	0	1	1	1	1	1	1	1	1	

Supplementary Table.9. Bacteriocins, EPS and phages. A. Presence and number of bacteriocins identified using BAGEL. Legend: orange = one copy of the gene /strain and red >1 gene/ strain. B. Presence and number of EPS clusters and phages identified. Strains are organized by genetic clade. Legend: light orange = one copy of the gene /strain, orange = between 1 and 10 genes/ strain and dark orange >10 genes/ strain. C.

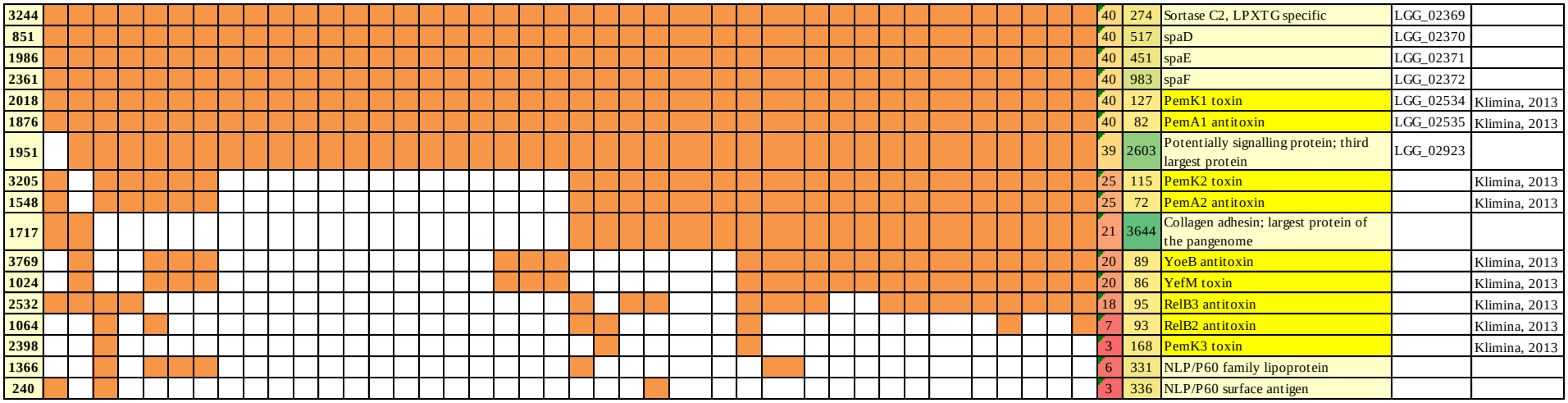



Number of regions
4 2 3 3 1 1 1 3 3 2 3 2 2 3 3 3 3 2 2 2 2 3 3 2 2 2 2 1 3 2 1 1 2 2 2 2 3 1 1 2
Number of genes
EPS
1 8
2 10
3 5
4 6
5 8
6 14
Phages
1 49
2 20
3 15
4 5
5 22
6 18
7 19
8 36
9 29
10 33
11 26
12 14
13 7
14 18

Supplementary Table.10. Selected *L. rhamnosus* genes and gene clusters associated with an effect on human cells by *in vitro* studies (A) and proteins containing an LPxTG domain (B) identified in the *L. rhamnosus* pangenome and their domain composition.

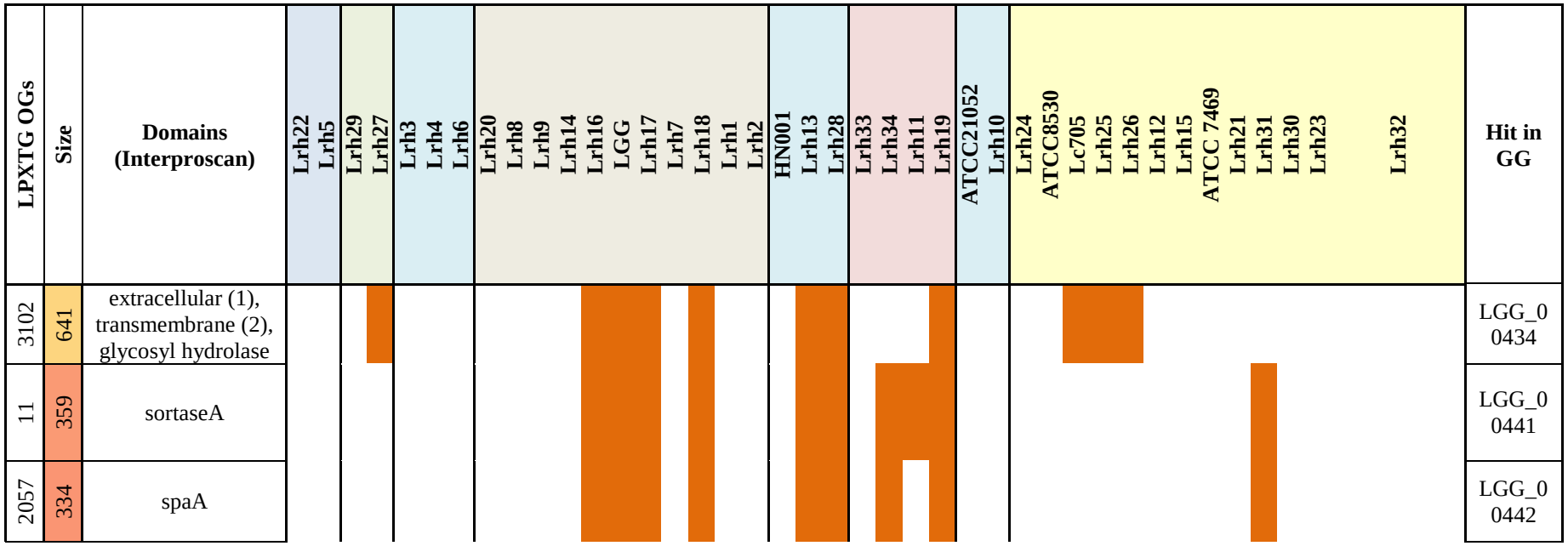
A.

OG	Lrh22	Lrh5	Lrh29	Lrh27	Lrh3	Lrh4	Lrh6	Lrh20	Lrh8	Lrh9	Lrh14	Lrh16	LGG	Lrh17	Lrh7	Lrh18	Lrh1	Lrh2	HN001	Lrh13	Lrh28	Lrh33	Lrh34	Lrh11	Lrh19	ATCC21052	Lrh10	Lrh24	ATCC8530	Lc705	Lrh25	Lrh26	Lrh12	Lrh15	ATCC7469	Lrh21	Lrh31	Lrh30	Lrh23	Lrh32	Presence	Size	RAST annotation	LGG genes	Ref.
2839																																								40	577	Fibronectin/fibrinogen-binding protein	LGG_01450	
3736																																								40	416	NLP/p40 Cell wall-associated hydrolase	LGG_00031	
2068																																								40	466	Nox, NADH peroxidase (EC 1.11.1.1)	LGG_00212	
3369																																								40	513	NLP/p75 Cell wall-associated hydrolase	LGG_00324	
11																																								11	359	Sortase C1, LPXTG specific	LGG_00441	
2057																																								10	334	spaA	LGG_00442	
1037																																								10	268	spaB	LGG_00443	
216																																								11	898	spaC	LGG_00444	
3662																																								40	90	RelE1 toxin	LGG_00493	Klimina, 2013
1096																																								40	72	RelB3 toxin	LGG_00523	Klimina, 2013
861																																								40	157	LuxS Autoinducer-2 production protein (EC 4.4.1.21)	LGG_00750	
3346																																								36	49	dltX D-Ala-teichoic acid biosynthesis	LGG_00776	
1835																																								39	506	dltA D-alanine--poly(phosphoribitol) ligase	LGG_00777	
156																																								40	405	dltB D-alanyl-lipoteichoic acid biosynthesis	LGG_00778	
1970																																								40	81	dltC D-alanine-poly(phosphoribitol) ligase (EC 6.1.1.13)	LGG_00779	
3528																																								40	423	dltD Poly(glycerophosphate chain) D-alanine transfer	LGG_00780	
3219																																								40	93	RelE3 antitoxin	LGG_00829	Klimina, 2013
1263																																								28	334	spc operon oxidoreductase, Gfo/Idh/MocA family	LGG_01582	Gagic 2013
3997																																								28	431	spc operon membrane protein	LGG_01588	
269																																								27	572	spcB docking protein	LGG_01589	
3049																																								27	79	spcC transmembrane protein	LGG_01590	
1330																																								29	913	spcD citoplasmic protein	LGG_01591	
2113																																								29	3390	spcA immunoglobulin like	LGG_01592	
1989																																								13	259	spcA/B class II aldolase/adducin domain protein		
2891																																								40	742	clpA ATP-dependent protease	LGG_01823	
1922																																								39	2357	mabA and fourth largest protein	LGG_01865	
3839																																								40	399	NLP/p60 Surface antigen	LGG_02016	
1655																																								40	233	Sortase A2	LGG_02143	
1104																																								37	1985	PrtP lactocepin fifth largest protein	LGG_02274	Schillde, 2013
3245																																								40	438	MBD cell wall surface anchor family protein	LGG_02337	
3686																																								40	371	Pili retraction protein pilT	LGG_02339	



3244	40	274	Sortase C2, LPXTG specific	LGG_02369	
851	40	517	spaD	LGG_02370	
1986	40	451	spaE	LGG_02371	
2361	40	983	spaF	LGG_02372	
2018	40	127	PemK1 toxin	LGG_02534	Klimina, 2013
1876	40	82	PemA1 antitoxin	LGG_02535	Klimina, 2013
1951	39	2603	Potentially signalling protein; third largest protein	LGG_02923	
3205	25	115	PemK2 toxin		Klimina, 2013
1548	25	72	PemA2 antitoxin		Klimina, 2013
1717	21	3644	Collagen adhesin; largest protein of the pangenome		
3769	20	89	YoeB antitoxin		Klimina, 2013
1024	20	86	YefM toxin		Klimina, 2013
2532	18	95	RelB3 antitoxin		Klimina, 2013
1064	7	93	RelB2 antitoxin		Klimina, 2013
2398	3	168	PemK3 toxin		Klimina, 2013
1366	6	331	NLP/P60 family lipoprotein		
240	3	336	NLP/P60 surface antigen		

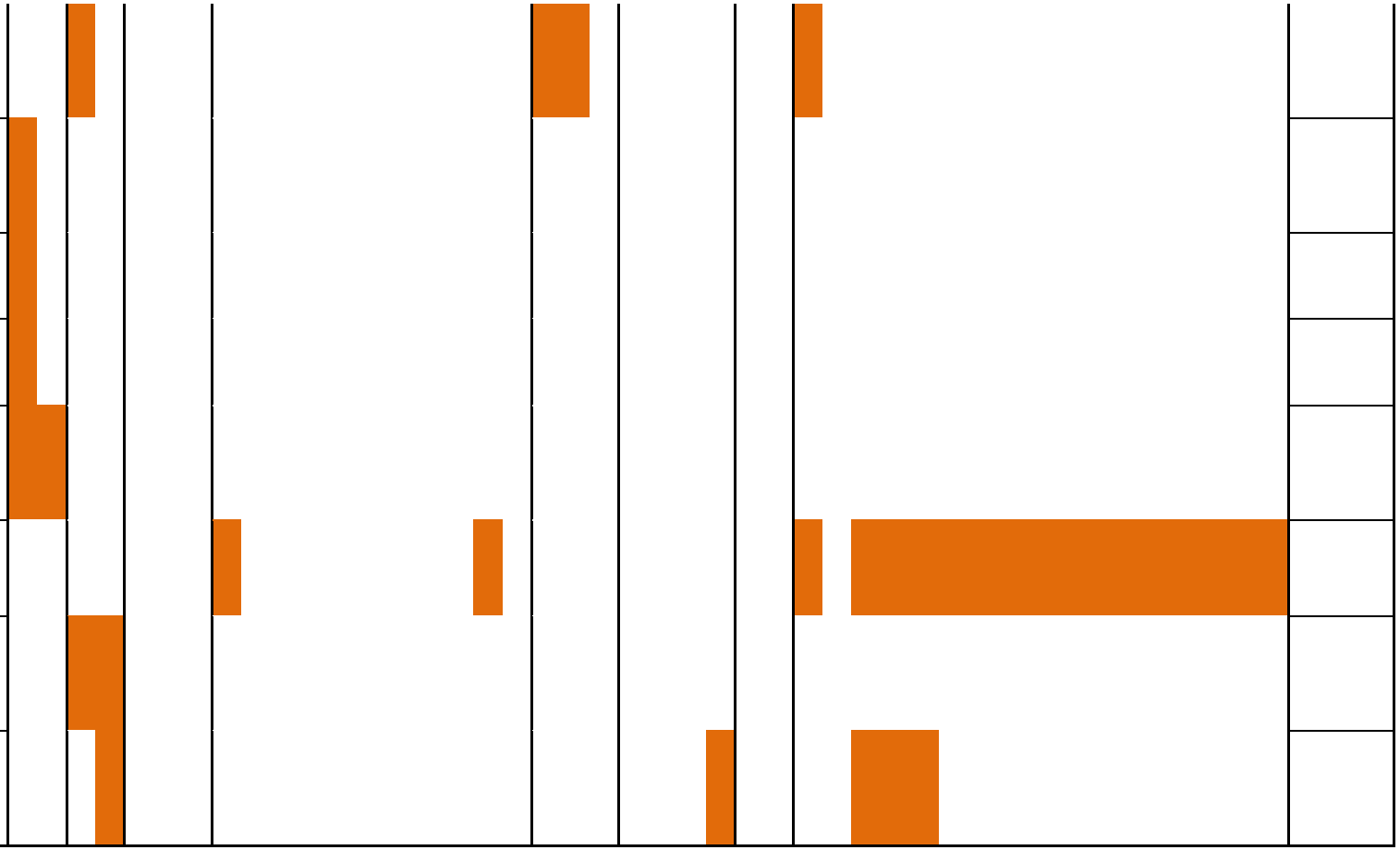
B.



LPXTG OGs	Size	Domains (Interproscan)	Hit in GG
3102	641	extracellular (1), transmembrane (2), glycosyl hydrolase	LGG_00434
11	359	sortaseA	LGG_00441
2057	334	spaA	LGG_00442


1037
268
spaB
LGG_0
0443
216
898
spaC
LGG_0
0444
471
118
extracellular (1),
transmembrane (2)
LGG_0
0578
2307
124
transmembrane (1)
LGG_0
0584
3582
319
extracellular (1),
transmembrane (1)
LGG_0
0721
269
572
Ig-like (2),
Cysteine-rich (1)
LGG_0
1589
2113
3390
SerAla repeat (9)
LGG_0
1592
1922
2357
DUF1542 (26),
extracellular (1),
transmembrane (1)
LGG_0
1865
707
1565
Phosphoesterase,
extracellular (1),
transmembrane (1)
LGG_0
2087
1104
1985
transmembrane (2),
Subtilisin serine
protease family
(S8)
LGG_0
2274
371
1561
extracellular (1),
transmembrane (2)
LGG_0
2283
3245
438
Cell-wall surface
anchor repeat (4),
extracellular (1),
transmembrane (2)
LGG_0
2337
3244
274
sortaseA
LGG_0
2369
851
517
spaD
LGG_0
2370

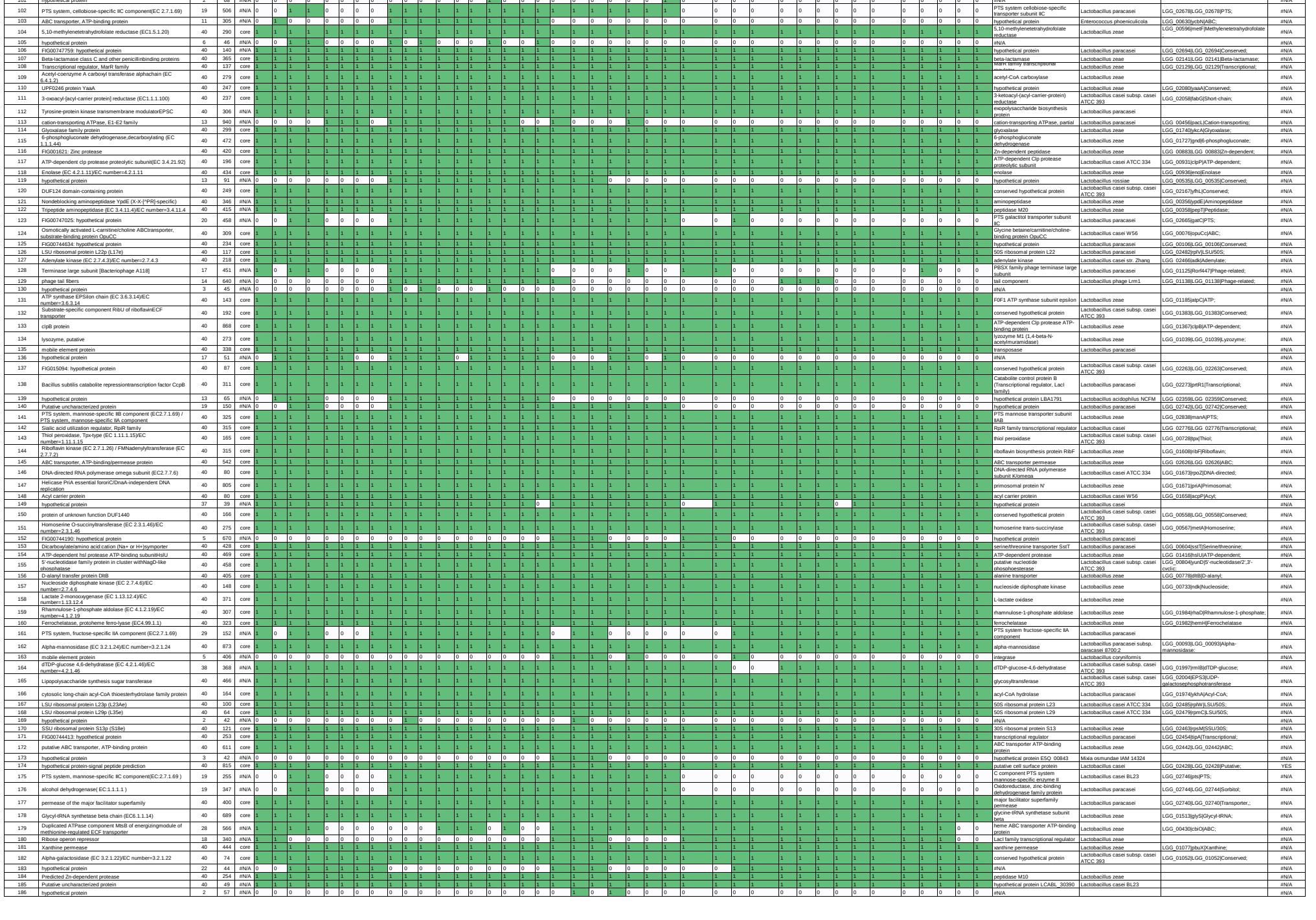

1986
451
spaE
LGG_02371
2361
983
spaF
LGG_02372
174
815
SEC10/PgrA surface exclusion (1), extracellular (1), transmembrane (2)
LGG_02428
2540
1498
Subtilisin serine protease (3), extracellular (1), transmembrane (2)
LGG_02734
1951
2603
Leucine rich repeats (6), extracellular (1), transmembrane (2)
LGG_02923
3312
955
Coils Coil (6), extracellular (1), transmembrane (2)
1778
1019
Glycoside hydrolase, extracellular (1), transmembrane (2)
2369
605
MucBP (3), extracellular (1), transmembrane (1)
1717
3644
collagen binding (16), extracellular (1), transmembrane (2)
1361
1270
collagen binding (6), extracellular (1), transmembrane (2)
2868
133
extracellular (1), transmembrane (1)
2405
893
extracellular (1), transmembrane (1)

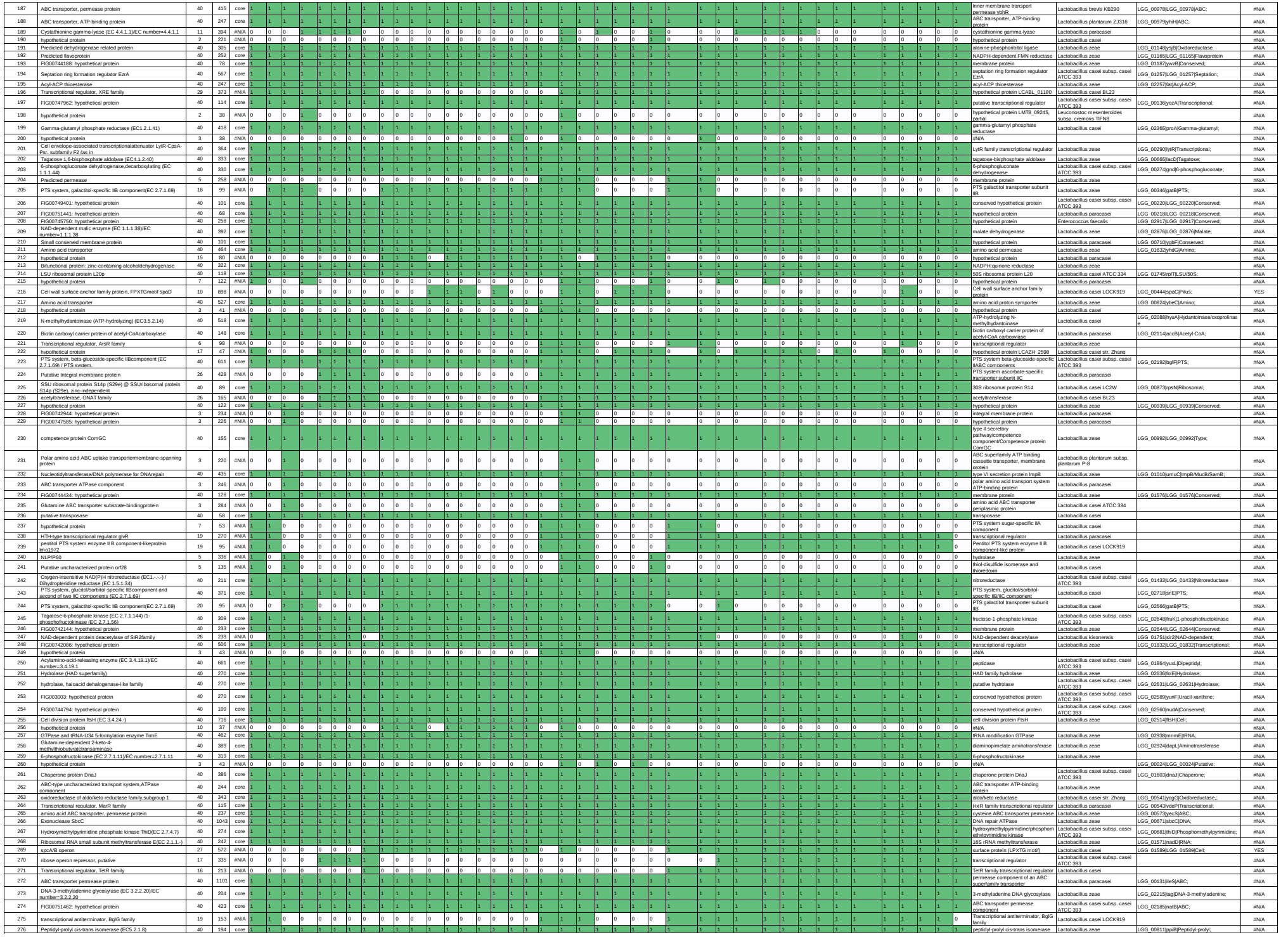
2397	138	collagen binding (1), extracellular (1), transmembrane (1)	
358	746	collagen binding (4), extracellular (1), transmembrane (1)	
4194	228	sortase	
4132	302	extracellular (1), transmembrane (1)	
470	1044	transmembrane (2), Alginate lyase, Sugar-binding Domain (2)	
2135	1089	extracellular (1), transmembrane (1)	
2492	804	Cysteine-rich (6), coils coil (6), extracellular (1), transmembrane (1)	
4103	338	collagen binding (2), extracellular (1), transmembrane (1)	

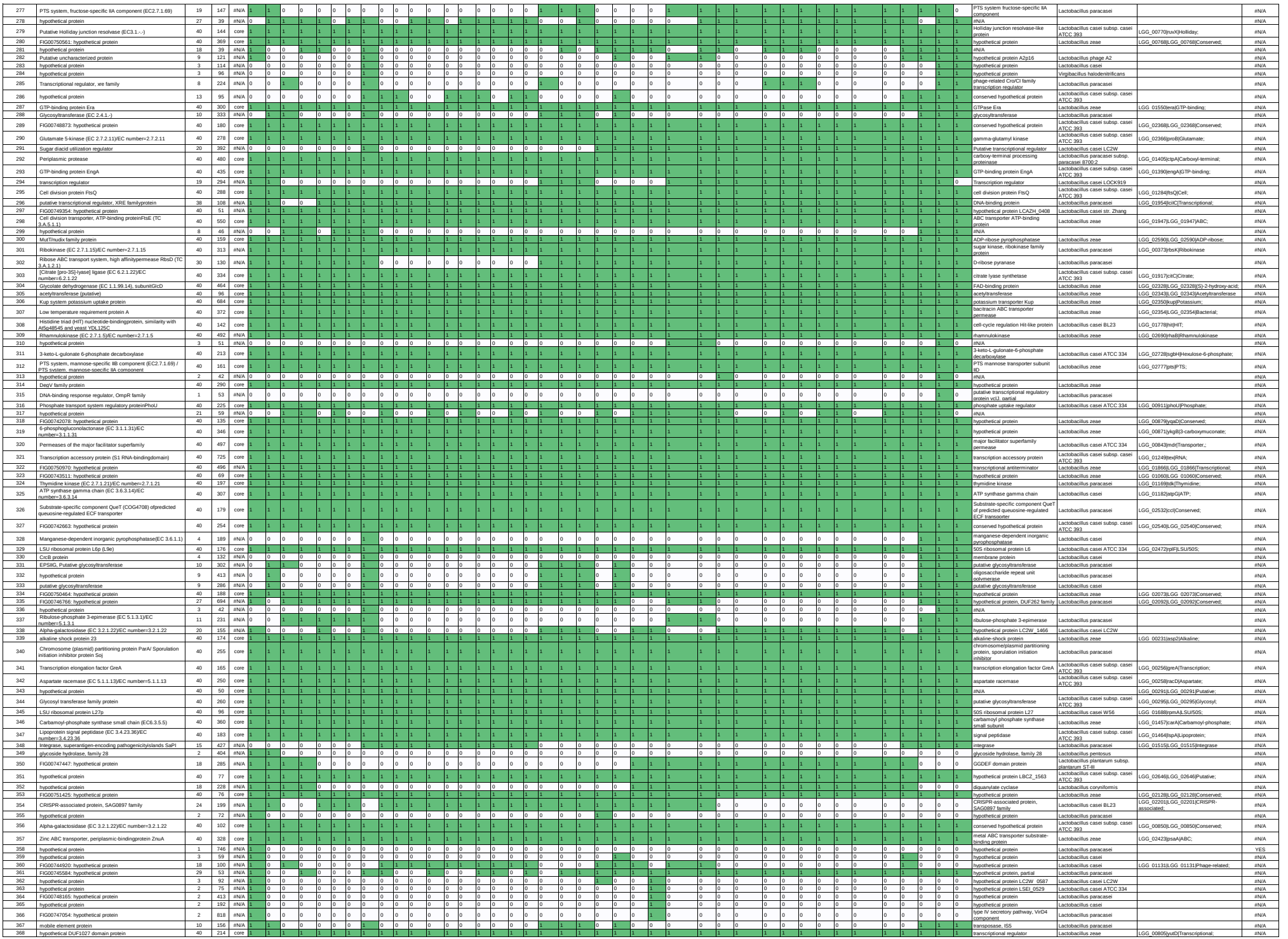
Supplementary Table.11. Summary of the two component regulatory systems in the *L. rhamnosus* genomes.

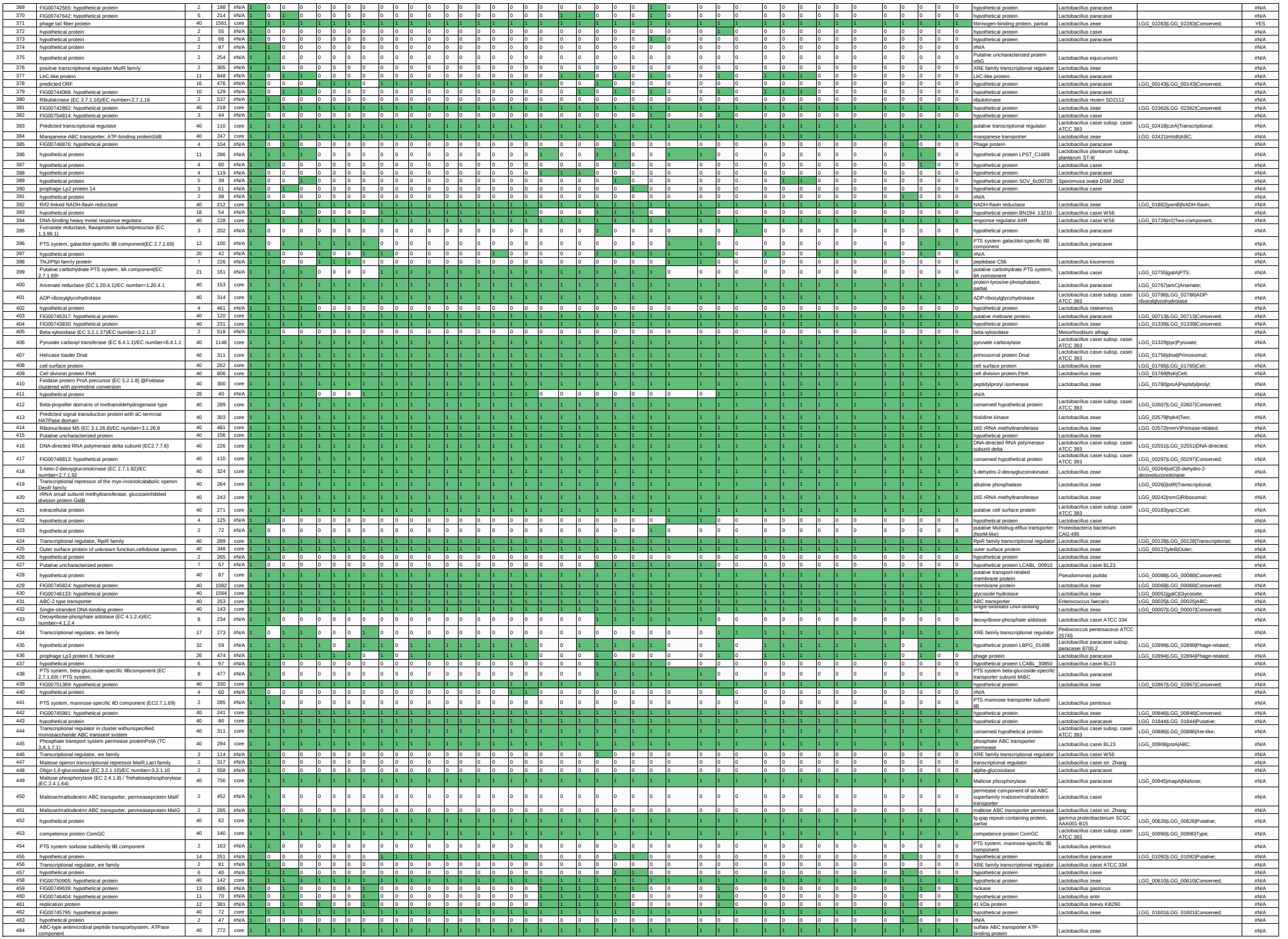
OG-annotation	Presence	OG	Size	Core	Lrh22	Lrh5	Lrh29	Lrh27	Lrh3	Lrh4	Lrh6	Lrh20	Lrh8	Lrh9	Lrh14	Lrh16	LGG	Lrh17	Lrh7	Lrh18	Lrh1	Lrh2	HN001	Lrh13	Lrh28	Lrh33	Lrh34	Lrh11	Lrh19	ATCC27	Lrh10	Lrh24	ATCC83	Lc705	Lrh25	Lrh26	Lrh12	Lrh15	ATCC74	Lrh21	Lrh31	Lrh30	Lrh23	Lrh32	Protein hit in nr database	Species hit in nr database	Hit in GG
orth1968; hypothetical_protein	11	1968	266	#N/A																																								hypothetical protein	*Bacillus fordii*	LGG_00628
orth3110; Two-component_sensor_kinase_ycbM(_EC:2.7.3.-_)	11	3110	303	#N/A																																								hypothetical protein	*Enterococcus phoeniculicola*	LGG_00629
orth103; ABC_transporter,_ATP-binding_protein	11	103	305	#N/A																																								hypothetical protein	*Enterococcus phoeniculicola*	LGG_00630
orth633; hypothetical_protein	11	633	228	#N/A																																								hypothetical protein	*Enterococcus phoeniculicola*	LGG_00631
orth1258; Two-component_response_regulator	25	1258	215	#N/A																																								transcriptional regulator	*Lactobacillus paracasei*	
orth2074; FIG00747001:_hypothetical_protein	25	2074	336	#N/A																																								signal transduction histidine kinase	*Lactobacillus casei str. Zhang*	
orth3950; ABC_transporter_ATP-binding_protein	25	3950	221	#N/A																																								ATPase component of an ABC superfamily antimicrobial peptide transporter	*Lactobacillus paracasei*	
orth1257; FIG00748333:_hypothetical_protein	25	1257	789	#N/A																																								lipoprotein release ABC-type transport system	*Lactobacillus paracasei subsp. paracasei 8700:2*	
orth1634; FIG00744054:_hypothetical_protein	40	1634	355	core																																								histidine kinase	*Oenococcus oeni*	LGG_00155
orth1773; DNA-binding_response_regulator,_OmpR_family	40	1773	224	core																																								transcriptional regulator	*Clostridium pasteurianum*	LGG_00156
orth3083; DNA-binding_response_regulator,_OmpR_family	40	3083	228	core																																								Response regulator	*Lactobacillus casei BL23*	LGG_00252
orth1879; Osmosensitive	40	1879	396	core																																								sensor histidine kinase	*Lactobacillus zeae*	LGG_00253
orth3413; DNA-binding_response_regulator,_OmpR_family(Rec-wHTH_domains)	40	3413	231	core																																								OmpR family DNA-binding response regulator	*Lactobacillus casei str. Zhang*	LGG_00548
orth2820; Signal_transduction_histidine_kinase	40	2820	374	core																																								two-component system, signal transduction histidine kinase	*Lactobacillus paracasei*	LGG_00549
orth18; Two_component_system_response_regulator_CiaR	40	18	231	core																																								transcriptional regulator	*Lactobacillus zeae*	LGG_00700
orth2710; FIG00745678:_hypothetical_protein	40	2710	430	core																																								signal transduction histidine kinase	*Lactobacillus zeae*	LGG_00701
orth2844; Alkaline_phosphatase_synthesis_transcriptional regulatory_protein_PhoP	40	2844	234	core																																								PhoP family transcriptional regulator	*Lactobacillus zeae*	LGG_00904
orth3140; Phosphate_regulon_sensor_protein_PhoR_(SphS)(EC_2.7.13.3)	40	3140	555	core																																								signal transduction histidine kinase	*Lactobacillus zeae*	LGG_00905
orth511; DNA-binding_response_regulator,_OmpR_family	40	511	243	core																																								PhoP family transcriptional regulator	*Lactobacillus zeae*	LGG_01003
orth2092; GO:0000155	40	2092	519	core																																								signal transduction histidine kinase	*Lactobacillus zeae*	LGG_01004
orth1683; Signal_transduction_histidine_kinase	40	1683	301	core																																								signal transduction histidine kinase	*Lactobacillus zeae*	LGG_01225
orth61; putative_response_regulator	40	61	227	core																																									putative response regulator	*Lactobacillus casei*	LGG_01226
orth1318; sensor_histidine_kinase	40	1318	546	core																																									signal transduction histidine kinase	*Lactobacillus paracasei*	LGG_01237
orth895; Signal_transduction_histidine_kinase	40	895	548	core																																									Sensor histidine kinase, putative	*Lactobacillus paracasei*	LGG_01429
orth2045; Two_component_transcriptional_regulator_VraR	40	2045	221	core																																									hypothetical protein BN194_18490	*Lactobacillus casei W56*	LGG_01710
orth2213; Sensor_histidine_kinase_VraS	40	2213	350	core																																									signal transduction histidine kinase	*Lactobacillus zeae*	LGG_01711
orth3880; transporter_associated_with_VraSR	40	3880	234	core																																									membrane protein	*Lactobacillus zeae*	LGG_01712
orth1673; Two-component_system_histidine_kinase	40	1673	527	core																																										sensor histidine kinase	*Lactobacillus zeae*	LGG_01725
orth394; DNA-binding_heavy_metal_response_regulator	40	394	228	core																																									response regulator ArlR	*Lactobacillus casei W56*	LGG_01726
orth2040; sensor_histidine_kinase	40	2040	374	core																																									two-component sensor kinase	*Lactobacillus casei subsp. casei ATCC 393*	LGG_01799
orth640; Response_regulator_of_the_LytR/AlgR_family	40	640	234	core																																									two-component response regulator	*Lactobacillus casei subsp. casei ATCC 393*	LGG_02578
orth413; Predicted_signal_transduction_protein_with_aC-terminal_HATPase_domain	40	413	303	core																																									histidine kinase	*Lactobacillus zeae*	LGG_02579
orth892; Two-component_sensor_kinase_SA14-24	40	892	639	core																																									sensor histidine kinase	*Lactobacillus zeae*	LGG_02812
orth2240; Two-component_response_regulator_SA14-24	40	2240	246	core																																									DNA-binding response regulator	*Lactobacillus casei ATCC 334*	LGG_02813
orth2417; Two-component_sensor_histidine_kinase,_malate(EC_2.7.3.-)	40	2417	517	core																																									two-component sensor histidine kinase	*Lactobacillus casei*	LGG_02878

OG number	Annotation	Presence	Size	Core	Lrh22	Lrh5	Lrh29	Lrh27	Lrh3	Lrh4	Lrh6	Lrh20	Lrh8	Lrh9	Lrh14	Lrh16	LGG	Lrh17	Lrh7	Lrh18	Lrh1	Lrh2	HN001	Lrh13	Lrh28	Lrh33	Lrh34	Lrh11	Lrh19	ATCC21052	Lrh10	Lrh24	ATCC8530	Lc705	Lrh25	Lrh26	Lrh12	Lrh15	ATCC 7469	Lrh21	Lrh31	Lrh30	Lrh23	Lrh32	Hit in nr database - Protein annotation	Hit from nr database - Species	LGG gene	LPXTG genes
1	RNA-binding protein Jag	40	255	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	RNA-binding protein	Lactobacillus zeae	LGG_02940|LGG_02940|RNA-binding;	#N/A
2	hypothetical protein	14	108	#N/A	0	0	1	1	0	0	0	0	1	1	1	1	1	1	1	1	1	1	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	hypothetical protein LCAZH_1221	Lactobacillus casei str. Zhang	LGG_00079|LGG_00079|Conserved;	#N/A
3	N-acetylmannosamine kinase (EC 2.7.1.60)/EC number=2.7.1.60	40	291	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	ROK family transcriptional regulator	Lactobacillus zeae	LGG_00096|glk|ROK;	#N/A
4	Aspartate-semialdehyde dehydrogenase (EC1.2.1.11)	40	352	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	aspartate-semialdehyde dehydrogenase	Lactobacillus zeae	LGG_00115|asd|Aspartate-semialdehyde;	#N/A
5	Fructose-bisphosphate aldolase class II (EC4.1.2.13)	29	297	#N/A	0	0	1	1	0	0	0	1	0	0	0	1	1	1	0	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	class II fructose-bisphosphate aldolase family protein	Lactobacillus casei	LGG_00414|gatY|Fructose-bisphosphate;	#N/A
6	PTS system, mannose-specific IIB component (EC2.7.1.69)	6	158	#N/A	0	0	1	1	0	0	0	0	0	0	0	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	PTS system mannose-specific transporter subunit IIB	Lactobacillus paracasei	LGG_00415|pts|PTS;	#N/A
7	FIG006427: Putative transport system permeaseprotein	6	431	#N/A	0	0	1	1	0	0	0	0	0	0	0	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	hypothetical protein LSEI_0379	Lactobacillus casei ATCC 334	LGG_00423|LGG_00423|Conserved;	#N/A
8	Protein containing an Alanine Racemase Domain	6	390	#N/A	0	0	1	1	0	0	0	0	0	0	0	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	amino acid racemase	Lactobacillus paracasei	LGG_00426|yhfX|Amino;	#N/A
9	O-succinylbenzoate synthase (EC 4.2.1.113)/ EC number=4.2.1.113	10	377	#N/A	0	0	1	1	0	0	0	0	0	0	0	1	1	1	0	1	0	0	0	0	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	O-succinylbenzoate-CoA synthase	Lactobacillus casei	LGG_00433|menC|N-acylamino;	#N/A
10	hypothetical protein	1	52	#N/A	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	hypothetical protein, partial	Lactobacillus paracasei	LGG_00439|LGG_00439|Putative;	#N/A
11	Sortase C1, LPXTG specific	10	359	#N/A	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	1	0	0	0	1	1	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	LPXTG specific sortase A	Lactobacillus casei LOCK919	LGG_00441|srtC1|Sortase;	YES
12	hypothetical protein	2	37	#N/A	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	#N/A			#N/A
13	Ribonucleotide reduction protein NrdI	40	202	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	flavoprotein	Lactobacillus zeae	LGG_00743|nrdI|Ribonucleotide;	#N/A
14	Predicted galactitol operon regulator (Transcriptional antiterminator), BglG family / PTS	8	149	#N/A	0	0	1	1	0	0	0	0	0	1	1	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	BglG family transcriptional antiterminator, partial	Lactobacillus paracasei	LGG_00400|ulaA|PTS;	#N/A
15	Transaldolase (EC 2.2.1.2)/EC number=2.2.1.2	8	222	#N/A	0	0	1	1	0	0	0	0	0	1	1	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	fructose-6-phosphate aldolase	Enterococcus faecalis	LGG_00398|tal|Transaldolase	#N/A
16	Galactose-1-phosphate uridylyltransferase (EC2.7.7.10)	40	485	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	galactose-1-phosphate uridyltransferase	Lactobacillus casei subsp. casei ATCC 393	LGG_00655|galT|Galactose-1-phosphate;	#N/A
17	FIG00745652: hypothetical protein	33	68	#N/A	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	1	1	1	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	hypothetical protein	Lactobacillus zeae	LGG_00661|LGG_00661|Conserved;	#N/A
18	Two component system response regulator CiaR	40	231	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	transcriptional regulator	Lactobacillus zeae	LGG_00700|phoP1|Two-component;	#N/A
19	hypothetical protein	7	37	#N/A	1	0	0	0	0	0	0	0	1	0	1	1	1	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	#N/A			#N/A
20	Transcriptional regulator/sugar kinase	40	298	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	Transcriptional regulator/sugar kinase	Lactobacillus paracasei	LGG_02736|LGG_02736|Transcriptional;	#N/A
21	DNA polymerase I (EC 2.7.7.7)/EC number=2.7.7.7	40	881	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	DNA polymerase I	Lactobacillus casei subsp. casei ATCC 393		#N/A
22	L-lactate dehydrogenase (EC 1.1.1.27)/EC number=1.1.1.27	40	296	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	lactate dehydrogenase	Lactobacillus zeae	LGG_01324|LGG_01324|Malate/lactate;	#N/A
23	COG0536: GTP-binding protein Obg	40	428	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	putative GTPase	Lactobacillus casei subsp. casei ATCC 393	LGG_01362|obg|GTP-binding;	#N/A
24	hypothetical protein	35	42	#N/A	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	1	1	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	#N/A			#N/A
25	hypothetical protein	24	50	#N/A	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	1	0	1	1	1	0	0	0	0	0	#N/A			#N/A
26	hypothetical protein	14	233	#N/A	1	1	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	hypothetical protein	Lactobacillus paracasei	LGG_02611|LGG_02611|Conserved;	#N/A
27	FIG00743589: hypothetical protein	40	92	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	hypothetical protein LCABL_27790	Lactobacillus casei BL23	LGG_02597|LGG_02597|Conserved;	#N/A
28	putative amino acid ABC transporter, permeaseprotein	12	228	#N/A	0	0	1	1	0	0	0	0	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	amino acid ABC transporter permease	Lactobacillus paracasei	LGG_00481|LGG_00481|ABC;	#N/A
29	FIG00745049: hypothetical protein	12	222	#N/A	0	0	1	1	0	0	0	0	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	amino acid ABC transporter permease	Lactobacillus casei BL23	LGG_00480|LGG_00480|ABC;	#N/A
30	FIG00742398: hypothetical protein	40	221	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	hypothetical protein	Lactobacillus zeae	LGG_02528|dsg|CBS;	#N/A
31	NADP-specific glutamate dehydrogenase (EC1.4.1.4)	40	447	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	glutamate dehydrogenase	Lactobacillus zeae	LGG_00607|gdhA|Glutamate;	#N/A
32	ABC transporter ATPase component	40	297	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	Nod factor export ATP-binding protein I	Lactobacillus paracasei	LGG_00622|cylA|ABC;	#N/A
33	ATP-dependent DNA helicase RecG (EC 3.6.1.-)	40	679	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	ATP-dependent DNA helicase	Lactobacillus zeae	LGG_01660|recG|ATP-dependent;	#N/A
34	Ribosomal RNA small subunit methyltransferase B(EC 2.1.1.-)	40	446	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	16S rRNA methyltransferase	Lactobacillus zeae	LGG_01669|sun|tRNA;	#N/A
35	FIG00743351: hypothetical protein	40	508	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	transcriptional regulator	Lactobacillus zeae	LGG_01895|LGG_01895|Transcriptional;	#N/A
36	cell surface hydrolase (putative)	40	323	#N/A	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	alpha/beta hydrolase	Lactobacillus zeae		#N/A
37	Transcriptional regulator of rhamnoseutilization, AraC family	40	340	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	bacterial regulatory helix-turn-helix s, AraC family protein	Lactobacillus paracasei	LGG_02692|LGG_02692|Transcriptional;	#N/A
38	ATP-dependent DNA helicase UvrD/PcrA	6	595	#N/A	0	0	0	0	0	0	0	0	0	1	1	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	hypothetical protein	Lactobacillus paracasei	LGG_00449|LGG_00449|UvrD/REP;	#N/A
39	FIG00746435: hypothetical protein	14	458	#N/A	0	0	0	0	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	putative membrane protein	Lactobacillus paracasei	LGG_01994|LGG_01994|Conserved;	#N/A
40	FIG000233: metal-dependent hydrolase	40	152	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	rRNA maturation factor	Lactobacillus zeae	LGG_01553|LGG_01553|Metalloprotease	#N/A
41	hypothetical protein	10	331	#N/A	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	hypothetical protein	Lactobacillus paracasei	LGG_01247|LGG_01247|Putative;	#N/A
42	Transketolase, N-terminal section (EC 2.2.1.1)/EC number=2.2.1.1	8	285	#N/A	0	0	1	1	0	0	0	0	0	1	1	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	putative transketolase	Lactobacillus paracasei	LGG_00405|tktN|Transketolase	#N/A
43	predicted ORF	13	75	#N/A	0	0	0	0	1	1	1	0	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	putative transposase	Lactobacillus paracasei	LGG_00235|is7|Transposase	#N/A
44	Putative parvulin type peptidyl-prolylisomerase, similarity with PrsA foldase	40	300	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	peptidylprolyl isomerase	Lactobacillus zeae	LGG_02275|prtM|Endopeptidase;	#N/A
45	FIG00750119: hypothetical protein	40	214	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	protease	Lactobacillus zeae	LGG_02241|LGG_02241|Metal-dependent;	#N/A
46	FIG00747177: hypothetical protein	40	214	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	hypothetical protein	Lactobacillus paracasei	LGG_02188|LGG_02188|Conserved;	#N/A
47	ABC transporter, ATP-binding protein	40	288	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	multidrug ABC transporter ATPase	Lactobacillus casei BL23	LGG_02187|CcmA|ABC;	#N/A
48	Transcriptional regulator, GntR family	40	131	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	GntR family transcriptional regulator	Lactobacillus paracasei	LGG_02186|gntR|Transcriptional;	#N/A
49	Putative prenyltransferase, contains1,4-dihydroxy-2-naphthoate octaprenyltransferase domain	40	304	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1,4-dihydroxy-2-naphthoate prenyltransferase	Lactobacillus zeae	LGG_02181|menA|1,4-dihydroxy-2-naphthoate;	#N/A
50	ThiJ/PfpI family	13	242	#N/A	0	0	0	0	1	1	1	0	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	glutamine amidotransferase	Leuconostoc citreum	LGG_02177|LGG_02177|Putative;	#N/A
51	FIG00742542: hypothetical protein	40	187	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	MFS transporter	Lactobacillus zeae	LGG_01842|LGG_01842|5-methylcytosine-specific;	#N/A
52	Acetyltransferase, GNAT family	40	154	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	putative acetyltransferase	Lactobacillus casei subsp. casei ATCC 393	LGG_01849|LGG_01849|Acetyltransferase,;	#N/A
53	DedA protein	40	217	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	membrane protein	Lactobacillus zeae	LGG_00808|dedA|DedA	#N/A
54	hypothetical protein	20	40	#N/A	0	0	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	#N/A			#N/A
55	FIG00742877: hypothetical protein	40	69	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	conserved hypothetical protein	Lactobacillus casei subsp. casei ATCC 393	LGG_02357|LGG_02357|CsbD-like;	#N/A
56	FIG00748552: hypothetical protein	40	189	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	hypothetical protein	Lactobacillus casei	LGG_02330|LGG_02330|Conserved;	#N/A
57	Putative prenyltransferase, contains1,4-dihydroxy-2-naphthoate octaprenyltransferase domain	40	312	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	putative 1,4-dihydroxy-2-naphthoate octaprenyltransferase	Lactobacillus casei	LGG_02303|ubiA|1,4-Dihydroxy-2-naphtoate;	#N/A
58	hypothetical protein	14	73	#N/A	1	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	hypothetical protein	Lactobacillus paracasei	LGG_01091|LGG_01091|Putative;	#N/A
59	Putative uncharacterized protein	17	97	#N/A	0	1	0	0	1	1	1	0	1	0	0	1	0	0	0	1	0	0	1	0	0	1	1	1	0	1	1	0	0	1	0	0	0	0	1	1	1	0	0	0	Phage protein	Lactobacillus paracasei	LGG_01141|LGG_01141|Phage-related;	#N/A
60	Oligopeptide ABC transporter, periplasmicoligopeptide-binding protein OppA (TC 3.A.1.5.1)	40	546	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	periplasmic-binding component of an ABC superfamily oligopeptide transporter	Lactobacillus paracasei	LGG_01224|oppA|ABC;	#N/A
61	putative response regulator	40	227	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	putative response regulator	Lactobacillus casei	LGG_01226|rr1|Two-component;	#N/A
62	hypothetical protein	4	43	#N/A	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	#N/A			#N/A
63	Transcriptional regulator	40	280	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	RpiR family transcriptional regulator	Lactobacillus zeae	LGG_00273|gntR|Transcriptional;	#N/A
64	Alpha-L-Rha alpha-1,3-L-rhamnosyltransferase(EC 2.4.1.-)	18	339	#N/A	0	0	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	putative rhamnosyltransferase	Lactobacillus paracasei	LGG_00281|welC|alpha-L-Rha;	#N/A
65	FIG00745165: hypothetical protein	40	256	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	hypothetical protein LCABL_23090	Lactobacillus casei BL23	LGG_02130|LGG_02130|Conserved;	#N/A
66	Transcriptional regulator of fatty acidbiosynthesis FabT	40	149	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	marR family transcriptional regulator	Lactobacillus casei W56		#N/A
67	phage infection protein	40	910	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	phage infection protein	Lactobacillus zeae	LGG_02102|pip|Phage-related;	#N/A
68	putative type II restriction enzyme methylasesubunit	11	740	#N/A	0	0	0	0	0	0	0	1	1	1	1	1	1	0	1	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	hypothetical protein	Lactobacillus paracasei	LGG_02095|LGG_02095|Adenine-specific;	#N/A
69	probable monooxygenase	40	369	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	monooxygenase	Lactobacillus zeae	LGG_00030|LGG_00030|Monooxygenase	#N/A
70	Alpha-L-Rha alpha-1,3-L-rhamnosyltransferase(EC 2.4.1.-)	10	311	#N/A	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Alpha-L-Rha alpha-1,3-L-rhamnosyltransferase	Lactobacillus casei LOCK919	LGG_02046|welH|alpha-L-Rha;	#N/A
71	Manganese-dependent inorganic pyrophosphatase(EC 3.6.1.1)	40	310	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	inorganic pyrophosphatase	Lactobacillus casei subsp. casei ATCC 393	LGG_01424|ppaC|Manganese-dependent;	#N/A
72	taurine transport system permease protein tauC	22	259	#N/A	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	taurine transport system permease tauC	Lactobacillus paracasei subsp. paracasei 8700:2	LGG_00174|tauC|Taurine;	#N/A
73	unknown	40	191	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	alanine acetyltransferase	Lactobacillus zeae	LGG_00180|yoaA|Alanine;	#N/A
74	2-haloacid dehalogenase (putative)	40	218	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	HAD family hydrolase	Lactobacillus brevis	LGG_00232|LGG_00232|Hydrolase;	#N/A
75	Putative stomatin/prohibitin-family membraneprotease subunit YbbK	40	310	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	hypothetical protein	Lactobacillus zeae	LGG_01037|hflC|Spfh;	#N/A
76	Putative pheromone cAM373 precursor lipoproteinCamS	40	383	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	conserved hypothetical protein	Lactobacillus casei subsp. casei ATCC 393	LGG_01017|camS|Lipoprotein	#N/A
77	ATP-dependent DNA helicase UvrD/PcrA	40	750	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	ATP-dependent DNA helicase PcrA	Lactobacillus zeae	LGG_01015|pcrA|ATP-dependent;	#N/A
78	Beta-glucosidase (EC 3.2.1.21)/EC number=3.2.1.21	19	485	#N/A	0	0	1	1	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	beta-glucosidase	Lactobacillus casei	LGG_02672|bglA|Beta-glucosidase;	#N/A
79	Transcriptional regulator, GntR family	24	245	#N/A	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	GntR family transcriptional regulator	Lactobacillus casei	LGG_02651|LGG_02651|Transcriptional;	#N/A
80	Glycerol kinase (EC 2.7.1.30)/EC number=2.7.1.30	40	510	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	glycerol kinase	Lactobacillus paracasei	LGG_00644|glpK|Glycerol;	#N/A
81	FIG00745449: hypothetical protein	40	157	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	hypothetical protein	Lactobacillus zeae	LGG_02645|LGG_02645|Conserved;	#N/A
82	hypothetical protein	7	47	#N/A	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	1	0	0	0	1	0	0	#N/A			#N/A
83	FIG005686: hypothetical protein	40	191	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	hypothetical protein	Lactobacillus zeae	LGG_01485|ypsA|Conserved;	#N/A
84	ATP-dependent nuclease, subunit A	40	1236	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	ATP-dependent helicase	Lactobacillus zeae	LGG_01496|rexA|ATP-dependent;	#N/A
85	putative tail component	20	126	#N/A	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	1	1	1	1	putative tail component	Lactobacillus casei	LGG_01529|LGG_01529|Phage-related;	#N/A
86	PTS system, cellobiose-specific IIC component(EC 2.7.1.69)	40	454	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	PTS cellobiose transporter subunit IIC	Lactobacillus paracasei	LGG_00162|celC|PTS;	#N/A
87	Glutaredoxin related protein	40	105	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	glutaredoxin	Lactobacillus zeae	LGG_01875|LGG_01875|Glutaredoxin;	#N/A
88	hypothetical protein	40	140	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	hypothetical protein	Lactobacillus zeae	LGG_01871|LGG_01871|Conserved;	#N/A
89	FIG00746427: hypothetical protein	40	193	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	conserved hypothetical protein	Lactobacillus casei subsp. casei ATCC 393		#N/A
90	FIG00747155: hypothetical protein	40	46	#N/A	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	#N/A			#N/A
91	FIG00751827: hypothetical protein	40	63	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	hypothetical protein	Lactobacillus zeae	LGG_00975|LGG_00975|Conserved;	#N/A
92	Multi antimicrobial extrusion protein(Na(+)/drug antiporter), MATE family of MDR efflux pumps	40	451	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	multidrug efflux pump	Lactobacillus casei subsp. casei ATCC 393		#N/A
93	hypothetical protein	16	738	#N/A	0	0	1	1	0	0	0	0	1	1	1	1	1	1	1	1	1	1	0	0	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	putative transcriptional regulator	Lactobacillus paracasei	LGG_00474|LGG_00474|Putative;	#N/A
94	Predicted Integral membrane protein	40	291	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	integral membrane protein	Lactobacillus paracasei	LGG_02550|LGG_02550|Conserved;	#N/A
95	hypothetical protein	39	93	#N/A	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	glutamate synthase	Gluconobacter morbifer		#N/A
96	Calcium-transporting ATPase (EC 3.6.3.8)/EC number=3.6.3.8	40	887	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	ATPase	Lactobacillus zeae	LGG_01855|pacL|Cation-transporting;	#N/A









| | | |

| | | |

| | | |

| | | |

| | | |

| | | |

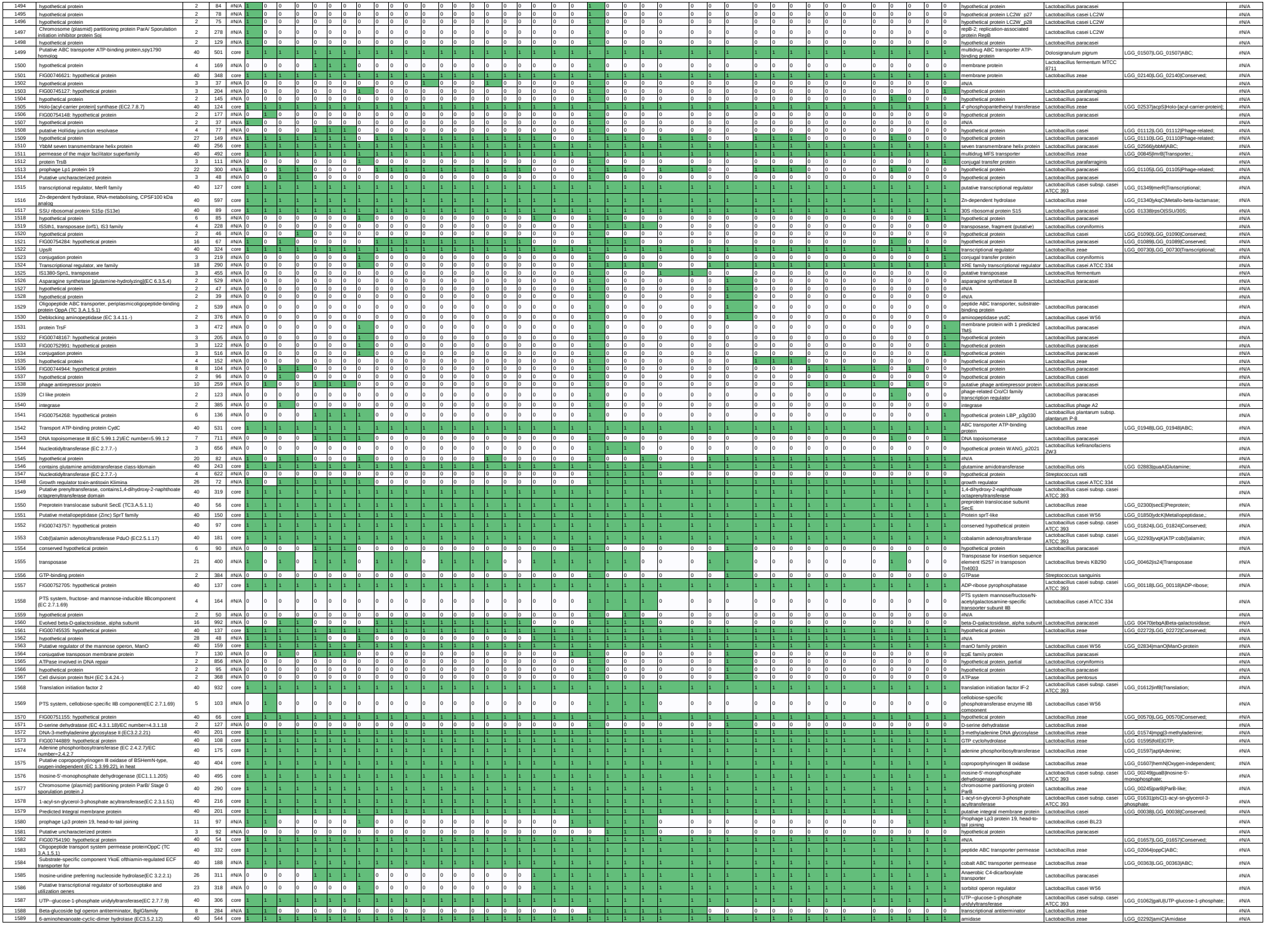
| | | |

| | | |

| | | |

| | | |

| | | |

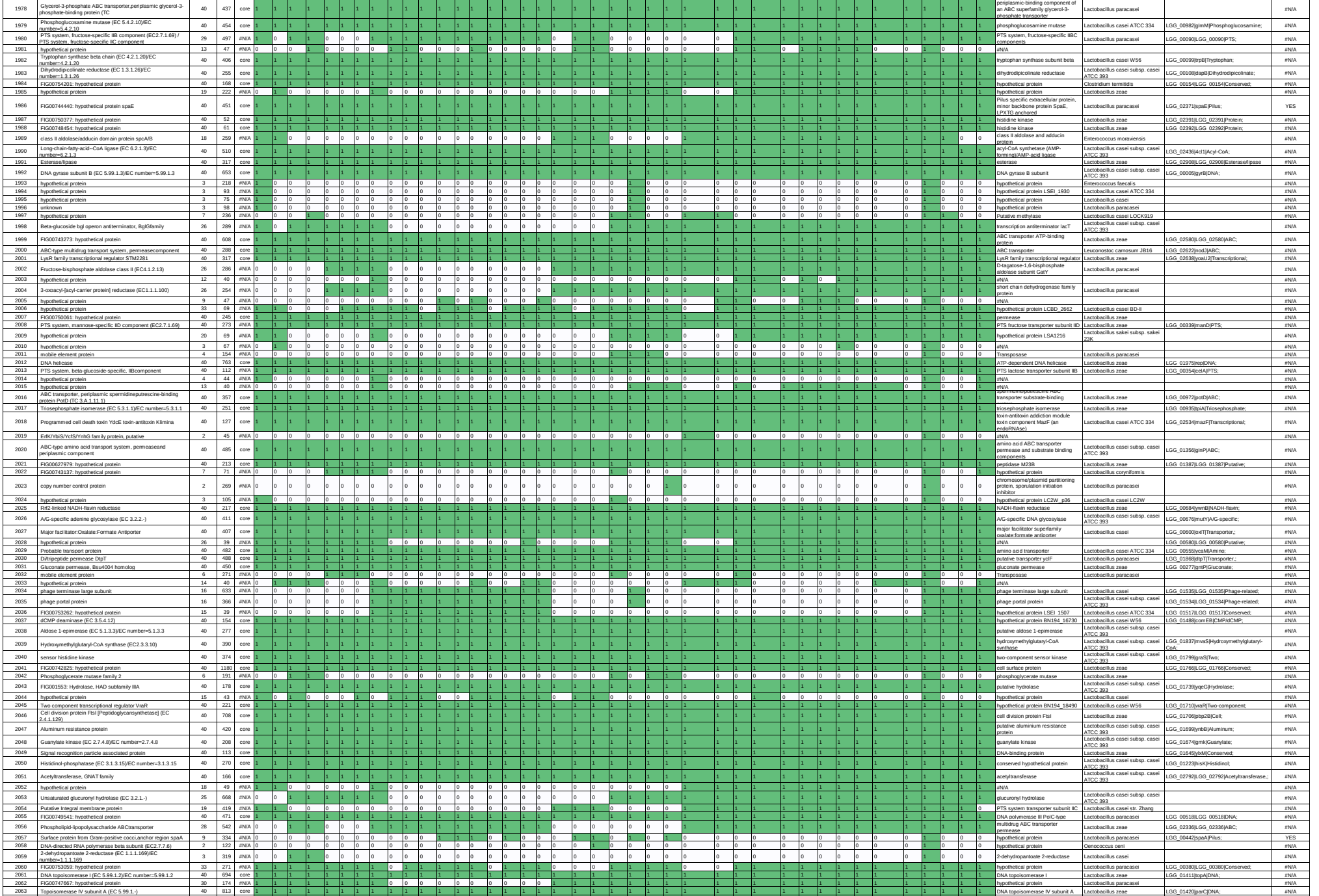


| | | |

| | | |

| | | |

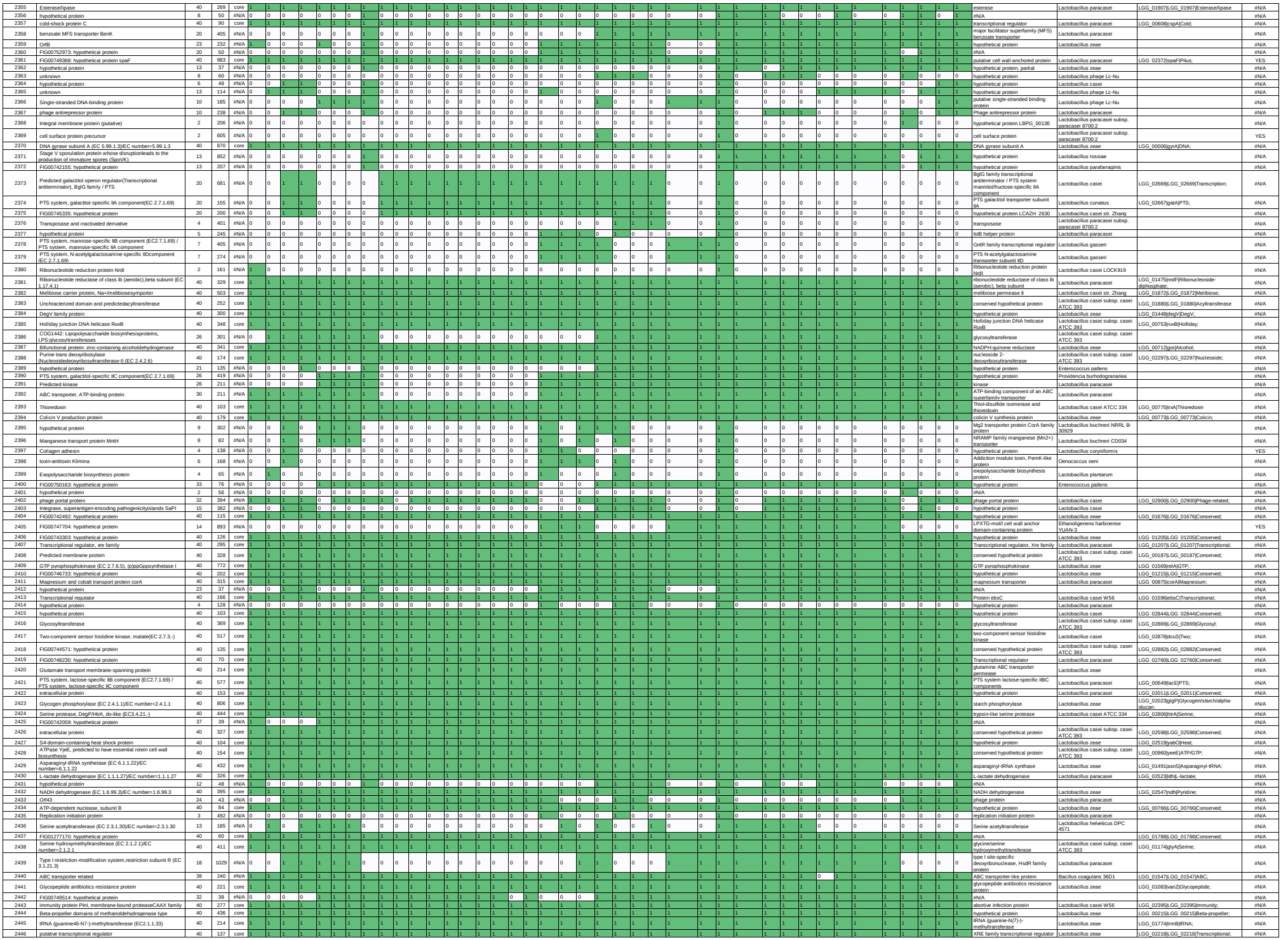
| | | |

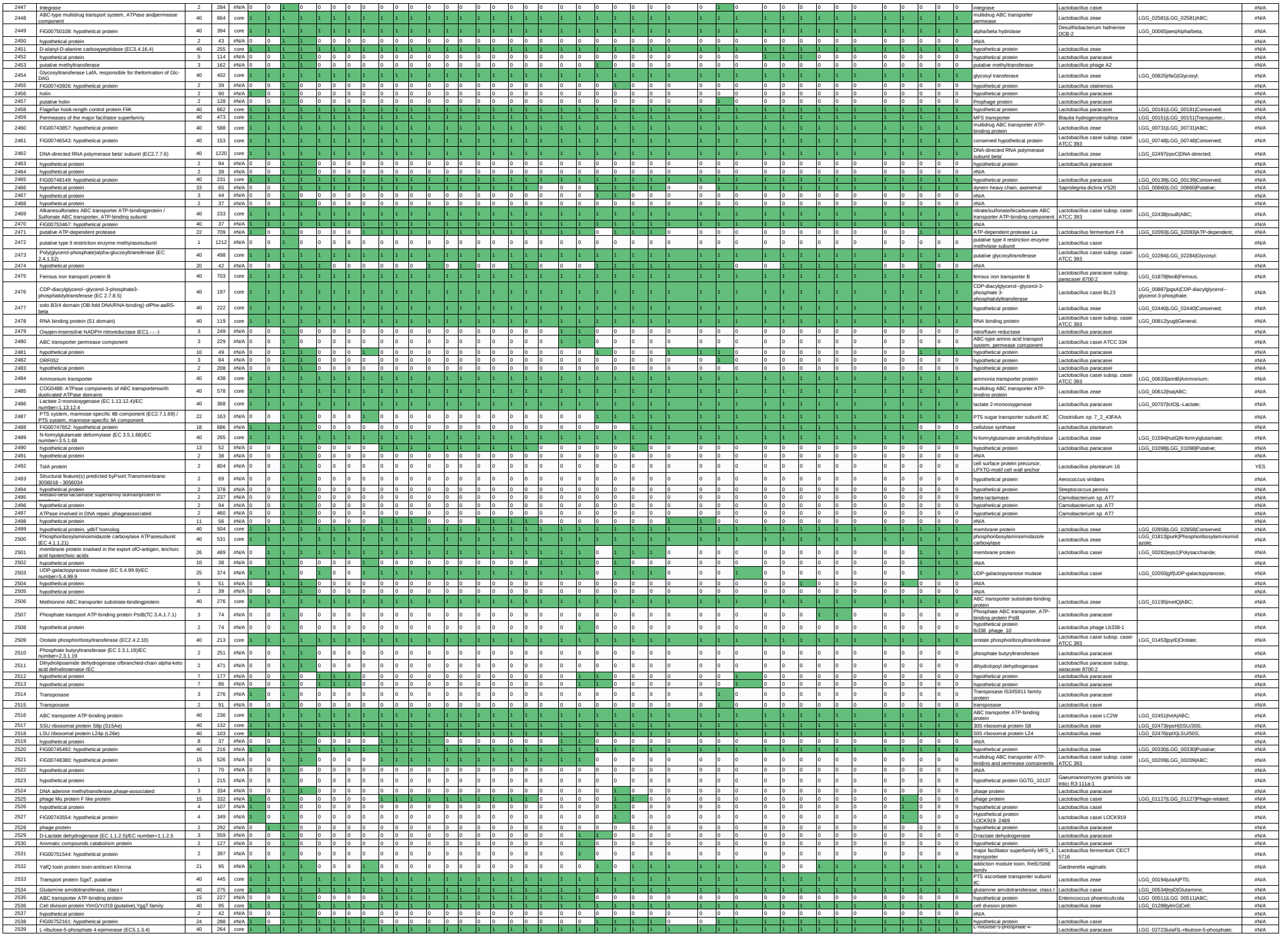


| | | |

| | | |

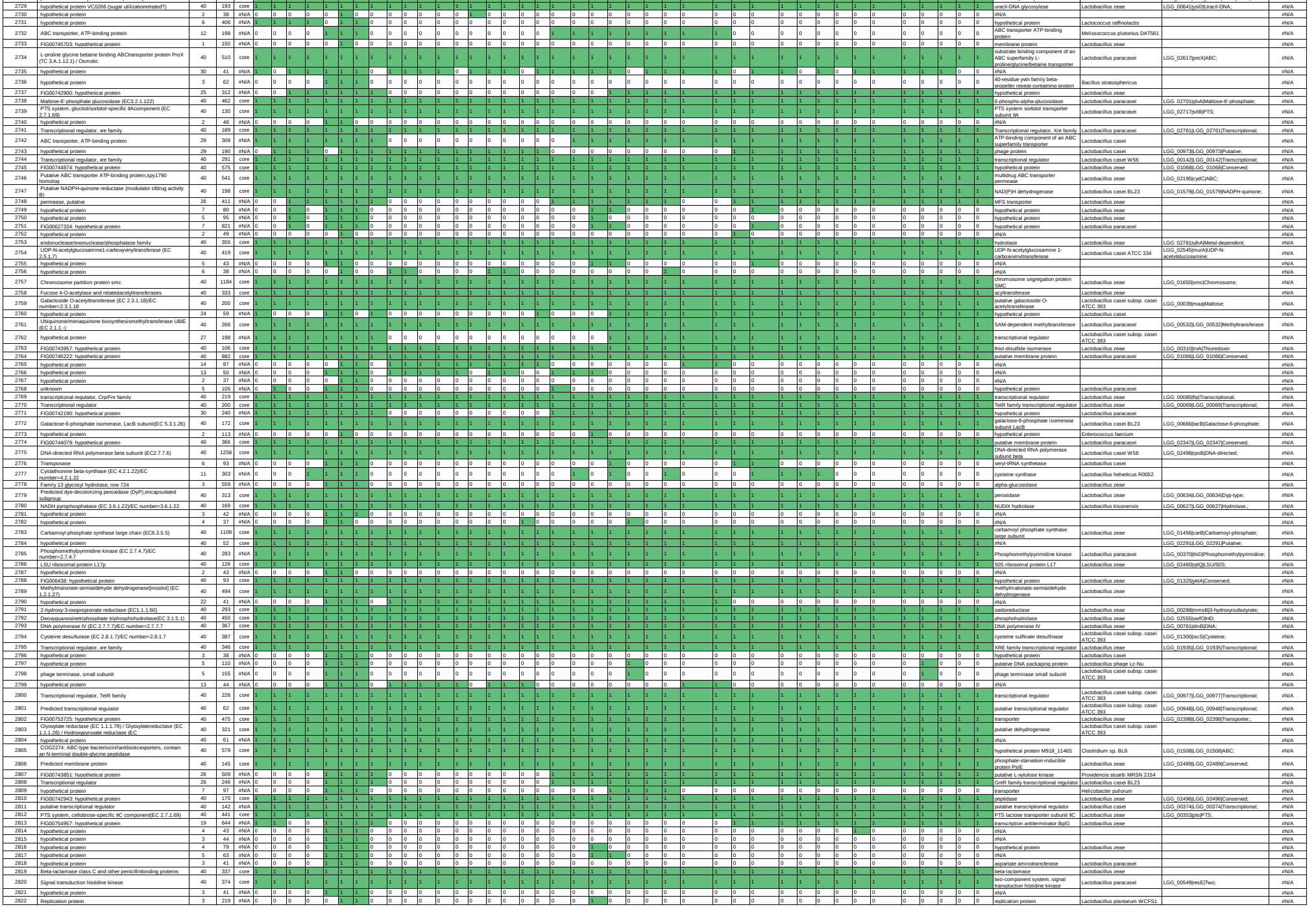
| | | |





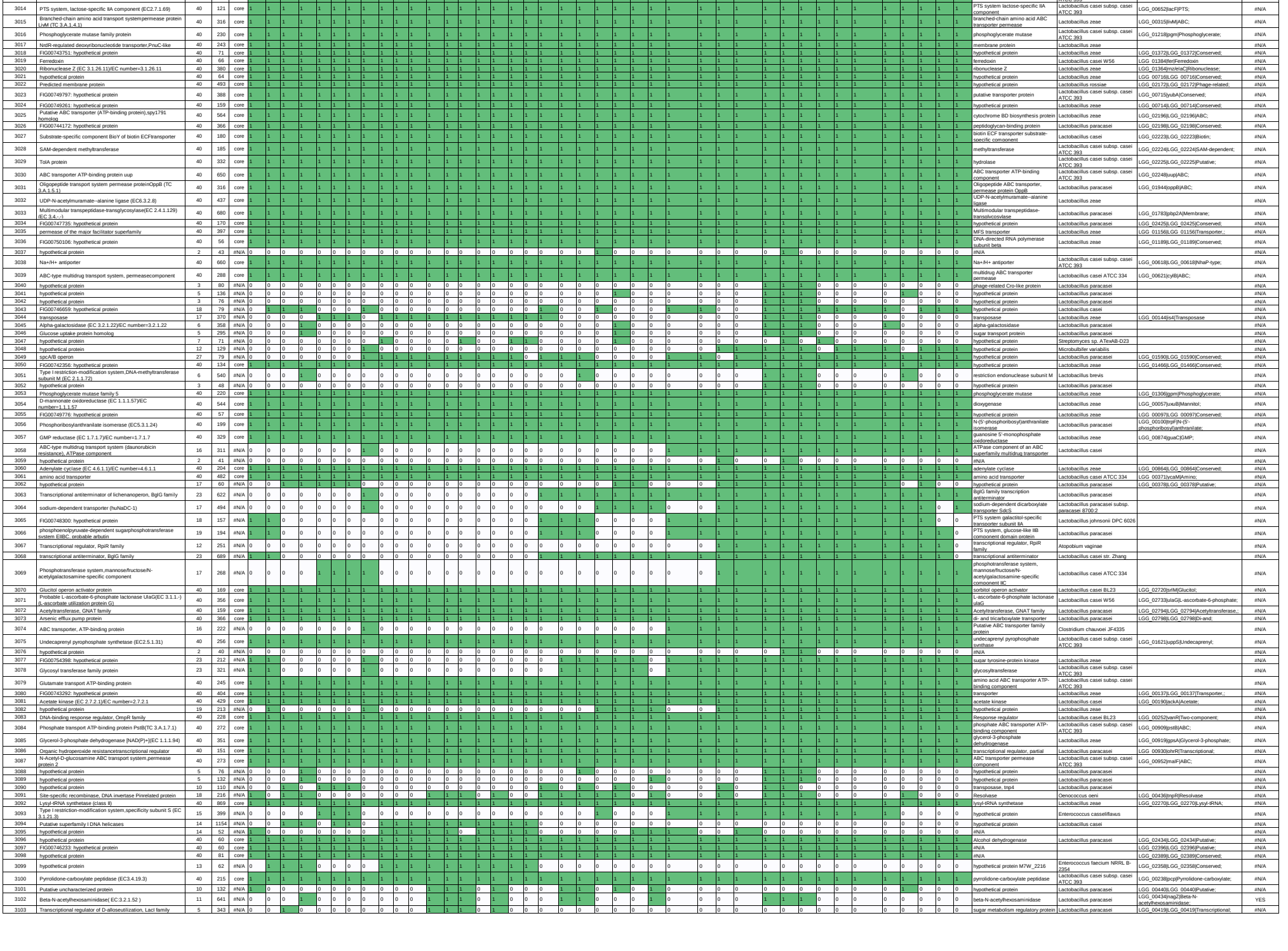
| | | |

| | | |



| | | |

| | | |



| | | |

| | | |

3287	ABC transporter, permease protein	40	622	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	peptide ABC transporter permease	Lactobacillus zeae	LGG_02920|salX|ABC;	#N/A
3288	periplasmic component of efflux system	40	356	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	transporter	Lactobacillus zeae	LGG_02919|LGG_02919|Transporter,;	#N/A
3289	FIG00746506: hypothetical protein	40	86	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	hypothetical protein	Lactobacillus zeae	LGG_02915|LGG_02915|Conserved;	#N/A
3290	ABC transporter ATP-binding protein	18	235	#N/A	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	peptide ABC transporter ATPase	Lactobacillus casei ATCC 334		#N/A
3291	FIG00744428: hypothetical protein	40	87	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	hypothetical protein	Lactobacillus paracasei	LGG_01946|LGG_01946|Conserved;	#N/A
3292	Acetylornithinedeacetylase/Succinyl-diaminopimelate desuccinylase and	40	442	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	acetylornithine deacetylase	Lactobacillus zeae	LGG_02800|dapE|Peptidase;	#N/A
3293	FIG00746526: hypothetical protein	40	110	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	hypothetical protein LCAZH_2769	Lactobacillus casei str. Zhang		#N/A
3294	FIG00747649: hypothetical protein	2	83	#N/A	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	hypothetical protein LGAS_0929	Lactobacillus gasseri ATCC 33323		#N/A
3295	Duplicated ATPase component YkoD of energizingmodule of thiamin-regulated ECF transporter for	40	469	#N/A	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	cobalt ABC transporter ATP-binding protein	Lactobacillus zeae	LGG_00364|cbiO|ABC;	#N/A
3296	FIG00752772: hypothetical protein	18	109	#N/A	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	PTS system galactitol-specific enzyme IIB component family protein	Streptococcus criceti		#N/A
3297	PTS system, mannose-specific IIA component (EC2.7.1.69)	17	138	#N/A	0	0	0	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	PTS system mannose/fructose-specific transporter subunit IIA	Lactobacillus paracasei		#N/A
3298	Conjugated bile salt hydrolase related amidase	40	338	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	putative conjugated bile salt hydrolase	Lactobacillus casei subsp. casei ATCC 393	LGG_00501|LGG_00501|Choloylglycine;	#N/A
3299	FIG00751110: hypothetical protein	23	198	#N/A	1	0	1	1	1	0	0	1	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	hypothetical protein	Lactobacillus murinus		#N/A
3300	Methionine ABC transporter substrate-bindingprotein	40	279	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	dioxygenase	Lactobacillus zeae	LGG_02926|metQ|NLPA;	#N/A
3301	Zn-dependent hydrolase, RNA-metabolising, CPSF73 kDa analog	40	640	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	putative hydrolase	Lactobacillus casei subsp. casei ATCC 393	LGG_01311|ykqC|Metallo-beta-lactamase;	#N/A
3302	Protein of unknown function DUF1447	40	123	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	hypothetical protein	Lactobacillus zeae	LGG_01312|LGG_01312|Conserved;	#N/A
3303	GTP-binding protein TypA/BipA	40	614	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	GTP-binding protein	Lactobacillus zeae	LGG_01327|typA|GTP-binding;	#N/A
3304	Ribosomal RNA small subunit methyltransferase D(EC 2.1.1.-)	40	188	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	rRNA methyltransferase	Lactobacillus zeae	LGG_01331|rsmD|Methyltransferase	#N/A
3305	Cell division trigger factor (EC 5.2.1.8)/EC number=5.2.1.8	40	446	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	trigger factor	Lactobacillus zeae	LGG_01351|tig|Trigger;	#N/A
3306	hypothetical protein	40	185	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	hypothetical protein	Lactobacillus zeae	LGG_01369|LGG_01369|Conserved;	#N/A
3307	Segregation and condensation protein A	40	241	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	segregation and condensation protein A	Lactobacillus zeae	LGG_01380|scpA|Segregation;	#N/A
3308	FIG00751031: hypothetical protein	2	329	#N/A	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	hypothetical protein	Lactobacillus paracasei		#N/A
3309	Glycosyl transferase, group 1	23	378	#N/A	1	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	glycosyltransferase	Lactobacillus casei subsp. casei ATCC 393		#N/A
3310	Heat shock protein GrpE	40	197	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	GrpE family heat shock protein	Lactobacillus paracasei	LGG_01605|grpE|GrpE;	#N/A
3311	Ribosome-binding factor A	40	146	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	ribosome-binding factor A	Lactobacillus zeae	LGG_01186|LGG_01186|Putative;	#N/A
3312	FIG00751977: hypothetical protein	19	955	#N/A	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	hypothetical protein	Lactobacillus paracasei		YES
3313	5-formyltetrahydrofolate cyclo-ligase (EC6.3.3.2)	40	190	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	5-formyltetrahydrofolate cyclo-ligase	Lactobacillus casei subsp. casei ATCC 393	LGG_00722|fthC|5-formyltetrahydrofolate;	#N/A
3314	Zinc ABC transporter, inner membrane permeaseprotein ZnuB	40	273	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	permease component of an ABC superfamily zinc transporter	Lactobacillus casei	LGG_00286|znuB|ABC;	#N/A
3315	FIG00750567: hypothetical protein	40	174	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	hypothetical protein	Lactobacillus zeae	LGG_00680|LGG_00680|Conserved;	#N/A
3316	FIG00753120: hypothetical protein	25	37	#N/A	1	0	1	1	1	1	1	1	1	0	0	0	1	0	0	0	0	0	0	1	1	0	0	0	0	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	#N/A			#N/A
3317	N-acetylmuramoyl-L-alanine amidase (EC3.5.1.28)	40	440	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	N-acetylmuramoyl-L-alanine amidase	Lactobacillus zeae	LGG_01566|ami|N-acetylmuramoyl-L-alanine;	#N/A
3318	hypothetical protein	22	37	#N/A	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	0	1	0	0	0	0	0	0	0	0	1	1	1	1	1	0	1	1	1	0	1	1	1	#N/A			#N/A
3319	Superfamily II DNA and RNA helicase	14	1170	#N/A	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	DNA helicase	Lactobacillus coryniformis		#N/A
3320	Alpha-galactosidase (EC 3.2.1.22)/EC number=3.2.1.22	40	836	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	alpha-galactosidase	Lactobacillus zeae	LGG_02079|melA|Alpha-galactosidase;	#N/A
3321	Type I restriction-modification system,specificity subunit S (EC 3.1.21.3)	14	407	#N/A	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	type I restriction-modification system specificity subunit	Lactobacillus paracasei		#N/A
3322	hypothetical protein	11	47	#N/A	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	0	0	0	0	#N/A			#N/A
3323	3-oxoacyl-[acyl-carrier-protein] synthase,KASII (EC 2.3.1.41)	40	408	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	3-oxoacyl-ACP synthase	Lactobacillus zeae		#N/A
3324	Teichoic acid glycosylation protein	40	136	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	membrane protein	Lactobacillus zeae	LGG_02144|gtcA|Cell;	#N/A
3325	FIG00746461: hypothetical protein	40	162	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	hypothetical protein	Lactobacillus zeae	LGG_02212|LGG_02212|Conserved;	#N/A
3326	modification methylase (Cytosine-specificmethyltransferase	14	415	#N/A	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	DNA-cytosine methyltransferase	Lactobacillus coryniformis		#N/A
3327	DNA replication intiation control protein YabA	40	112	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	conserved hypothetical protein	Lactobacillus casei subsp. casei ATCC 393	LGG_02259|LGG_02259|Initiation-control;	#N/A
3328	hypothetical protein	40	121	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	hypothetical protein	Lactobacillus paracasei	LGG_02268|LGG_02268|Conserved;	#N/A
3329	FIG00747933: hypothetical protein	34	156	#N/A	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	hypothetical protein	Enterococcus casseliflavus	LGG_01881|LGG_01881|Conserved;	#N/A
3330	ABC transporter substrate-binding protein	40	331	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	ABC transporter substrate binding component	Lactobacillus casei subsp. casei ATCC 393	LGG_01853|ABC-SBP|ABC;	#N/A
3331	Indole-3-glycerol phosphate synthase (EC4.1.1.48)	40	259	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	Indole-3-glycerol phosphate synthase	Lactobacillus casei ATCC 334	LGG_00101|trpC|Indole-3-glycerol;	#N/A
3332	Carboxylesterase (EC 3.1.1.1)/EC number=3.1.1.1	40	213	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	putative esterase	Lactobacillus casei subsp. casei ATCC 393	LGG_01742|yahD|Phospholipase/Carboxylesterase	#N/A
3333	Acylphosphate phosphohydrolase (EC 3.6.1.7),putative	40	93	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	acylphosphatase	Lactobacillus zeae	LGG_01723|acyP|Acylphosphatase	#N/A
3334	Translation elongation factor P	40	187	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	elongation factor P	Lactobacillus zeae	LGG_01686|efp|Translation;	#N/A
3335	amino acid ABC transporter, ATP-bindingprotein	40	212	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	amino acid ABC transporter ATP-binding component	Lactobacillus casei subsp. casei ATCC 393	LGG_01271|glnQ|ABC;	#N/A
3336	ATP synthase B chain (EC 3.6.3.14)/EC number=3.6.3.14	40	162	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	F0F1 ATP synthase subunit B	Lactobacillus zeae	LGG_01179|atpF|ATP;	#N/A
3337	FIG00754372: hypothetical protein	33	61	#N/A	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	#N/A		LGG_01154|LGG_01154|Putative;	#N/A
3338	2-dehydro-3-deoxyphosphogluconate aldolase (EC4.1.2.14) / 2-dehydro-3-deoxygluconate kinase (EC	40	558	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2-keto-3-deoxy-6-phosphogluconate aldolase and 2-dehydro-3-deoxygluconokinase	Lactobacillus casei ATCC 334	LGG_00060|kdfK|2-dehydro-3-deoxygluconokinase	#N/A
3339	PTS system,hyaluronate-oligosaccharide-specific IIB component (EC	25	166	#N/A	0	0	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	PTS N-acetylgalactosamine transporter subunit IIB	Lactobacillus zeae		#N/A
3340	Na+:H+ antiporter	40	704	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	sodium:proton antiporter	Lactobacillus zeae	LGG_00967|LGG_00967|NhaP-type;	#N/A
3341	hypothetical protein	3	40	#N/A	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	1	0	#N/A			#N/A
3342	homolog offig|393130.3.peg.2627 surface antigen	40	496	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	conserved hypothetical protein	Lactobacillus casei subsp. casei ATCC 393	LGG_00914|LGG_00914|Conserved;	#N/A
3343	Septum site-determining protein MinC	40	218	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	cell division inhibitor	Lactobacillus zeae	LGG_01268|minC|Cell;	#N/A
3344	FIG00746758: hypothetical protein	16	241	#N/A	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	ABC transporter-like protein	Rhodopseudomonas palustris BisA53		#N/A
3345	CBS domain protein, lmo1865 homolog	40	185	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	hypothetical protein	Lactobacillus zeae	LGG_02352|yqzB|CBS;	#N/A
3346	D-Ala-teichoic acid biosynthesis protein(putative)	40	49	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	cytochrome C553	Lactobacillus zeae	LGG_00776|dltX|D-Ala-teichoic;	#N/A
3347	FIG006126: DNA helicase,restriction/modification system component YeeB	4	650	#N/A	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	0	0	0	0	hypothetical protein	Lactobacillus paracasei		#N/A
3348	YeeC-like protein	4	402	#N/A	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	0	0	0	0	hypothetical protein	Lactobacillus paracasei		#N/A
3349	FIG00750005: hypothetical protein	40	110	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	hypothetical protein LBPG_02310	Lactobacillus paracasei subsp. paracasei 8700:2	LGG_02399|LGG_02399|Conserved;	#N/A
3350	replication protein B	3	182	#N/A	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	replication protein B	Lactobacillus paracasei		#N/A
3351	Hypoxanthine-guanine phosphoribosyltransferase(EC 2.4.2.8)	40	182	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	hypoxanthine phosphoribosyltransferase	Lactobacillus zeae	LGG_02515|hpt|Hypoxanthine;	#N/A
3352	Cardiolipin synthetase (EC 2.7.8.-)	40	477	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	cardiolipin synthetase	Lactobacillus zeae	LGG_01274|cls|Cardiolipin;	#N/A
3353	CAAX amino terminal protease family	40	141	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	hypothetical protein	Lactobacillus paracasei	LGG_00019|LGG_00019|Putative;	#N/A
3354	transcriptional regulator, MerR family	12	135	#N/A	0	0	1	1	0	0	0	0	1	1	1	0	0	1	1	1	1	1	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	transcriptional regulator	Lactobacillus brevis ATCC 367		#N/A
3355	Beta-lactamase class C and other penicillinbinding proteins	40	334	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	beta-lactamase	Lactobacillus casei subsp. casei ATCC 393	LGG_01167|ampC|Beta-lactamase;	#N/A
3356	hypothetical protein	5	61	#N/A	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	N-acetylmuramoyl-L-alanine amidase	Lactobacillus casei BL23		#N/A
3357	Glycosyltransferase	40	387	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	glycosyltransferase	Lactobacillus zeae	LGG_01147|LGG_01147|Glycosyl;	#N/A
3358	hypothetical protein	25	37	#N/A	1	0	0	1	0	0	0	1	0	0	0	1	0	1	0	1	0	0	1	1	1	0	1	1	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	#N/A			#N/A
3359	hypothetical protein	7	228	#N/A	1	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	hypothetical protein	Lactobacillus zeae		#N/A
3360	Tryptophanyl-tRNA synthetase (EC 6.1.1.2)/EC number=6.1.1.2	40	338	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	tryptophanyl-tRNA synthetase	Lactobacillus casei subsp. casei ATCC 393	LGG_02609|trpS|Tryptophanyl-tRNA;	#N/A
3361	hypothetical protein	5	54	#N/A	0	0	1	1	0	0	0	0	0	0	0	1	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	#N/A		LGG_00432|LGG_00432|Putative;	#N/A
3362	Aspartate aminotransferase family	6	368	#N/A	0	0	1	1	0	0	0	0	0	0	0	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	cystathionine beta-lyase family protein	Lactobacillus paracasei	LGG_00425|yhfS|Pyridoxal;	#N/A
3363	FIG00748993: hypothetical protein	40	45	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	#N/A		LGG_02029|LGG_02029|Putative;	#N/A
3364	FIG00747013: hypothetical protein	30	226	#N/A	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	hypothetical protein LCBD_0342	Lactobacillus casei BD-II		#N/A
3365	Ribose ABC transport system, permease proteinRbsC (TC 3.A.1.2.1)	30	327	#N/A	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	ribose ABC transporter permease	Lactobacillus zeae		#N/A
3366	Transcriptional antiterminator of lichenanoperon, BglG family	40	602	#N/A	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	transcriptional antiterminator	Lactobacillus zeae	LGG_00357|LGG_00357|Transcription;	#N/A
3367	PTS system, galactitol-specific IIA component(EC 2.7.1.69)	18	168	#N/A	0	1	1	1	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	PTS galactitol transporter subunit IIA	Lactobacillus zeae	LGG_00345|gatA|PTS;	#N/A
3368	FIG00746095: hypothetical protein	8	96	#N/A	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	1	0	1	1	1	0	0	0	0	0	0	0	0	hypothetical protein	Lactobacillus paracasei		#N/A
3369	Cell wall-associated hydrolase NLP/p75	40	513	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	Putative secreted protein	Lactobacillus paracasei	LGG_00324|p75|Cell;	#N/A
3370	FIG00744856: hypothetical protein	40	226	#N/A	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	conserved hypothetical protein	Lactobacillus casei subsp. casei ATCC 393	LGG_00663|LGG_00663|Conserved;	#N/A
3371	putative minor capsid protein	2	494	#N/A	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	phage minor capsid protein	Lactobacillus paracasei		#N/A
3372	Terminase large subunit	2	435	#N/A	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	phage terminase	Streptococcus agalactiae		#N/A
3373	hypothetical protein	2	239	#N/A	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	hypothetical protein LEGAS_0827	Leuconostoc gasicomitatum LMG 18811		#N/A
3374	hypothetical protein	2	227	#N/A	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	hypothetical protein	Lactococcus garvieae		#N/A
3375	phage transcriptional activator ArpU	18	140	#N/A	0	0	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	phage transcriptional regulator	Lactobacillus paracasei subsp. paracasei JCM 8130	LGG_01967|LGG_01967|Conserved;	#N/A
3376	hypothetical protein	2	38	#N/A	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	#N/A			#N/A
3377	hypothetical protein	2	72	#N/A	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	hypothetical protein	Allofustis seminis		#N/A
3378	hypothetical protein	3	76	#N/A	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	hypothetical protein Lrm1_gp45	Lactobacillus phage Lrm1		#N/A
3379	Methyltransferase	2	274	#N/A	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	D12 class N6 adenine-specific DNA methyltransferase family protein	Carnobacterium maltaromaticum LMA28		#N/A
3380	ORF039	3	126	#N/A	1	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	hypothetical protein BN194_12830	Lactobacillus casei W56		#N/A

| | | |

| | | |

| | | |

| | | |

3746	hypothetical protein	15	48	#N/A	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	0	0	0	1	1	0	1	1	0	1	1	1	1	1	1	hypothetical protein LBCZ_1448	Lactobacillus casei subsp. casei ATCC 393		#N/A
3747	FIG00744395: hypothetical protein	40	83	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	hypothetical protein Lbuc_0263	Lactobacillus buchneri NRRL B-30929	LGG_01073|LGG_01073|Putative;	#N/A
3748	Glycerol-3-phosphate ABC transporter,periplasmic glycerol-3-phosphate-binding protein (TC	40	457	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	sugar ABC transporter periplasmic protein	Lactobacillus casei ATCC 334	LGG_02768|ugpB|ABC;	#N/A
3749	Phosphoribosylaminoimidazole carboxylase ATPasesubunit (EC 4.1.1.21)	40	369	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	phosphoribosylaminoimidazole carboxylase	Lactobacillus zeae	LGG_01078|purK|Phosphoribosylaminoimidazole;	#N/A
3750	ABC-type multidrug transport system, ATPase andpermease components	21	537	#N/A	0	0	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	ABC transporter, atpase and permease components	Lactobacillus crispatus ST1		#N/A
3751	Aryl-alcohol dehydrogenase related enzyme	40	309	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	Protein IolS	Lactobacillus paracasei	LGG_02833|iolS|Aldo/keto;	#N/A
3752	hypothetical protein	23	229	#N/A	0	0	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	rhomboid family S54 peptidase	Lactobacillus casei		#N/A
3753	Iron-sulfur cluster assembly protein SufB	40	478	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	Fe-S cluster assembly protein SufB	Lactobacillus zeae	LGG_01203|sufB|Fe-S;	#N/A
3754	Deoxyadenosine kinase (EC 2.7.1.76) /Deoxyguanosine kinase (EC 2.7.1.113)	40	242	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	deoxynucleoside kinase	Lactobacillus casei subsp. casei ATCC 393	LGG_01222|dgk|Deoxynucleoside;	#N/A
3755	hypothetical protein	18	49	#N/A	0	0	0	0	1	1	1	0	1	0	0	0	0	0	1	1	0	1	1	1	1	0	0	0	0	0	0	1	1	1	1	1	1	1	1	0	0	0	0	0	#N/A			#N/A
3756	Oligopeptide transport system permease proteinOppB (TC 3.A.1.5.1)	40	307	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	oligopeptide transport system permease OppB	Lactobacillus casei W56	LGG_02065|oppB|ABC;	#N/A
3757	hypothetical protein	40	65	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	#N/A		LGG_02929|LGG_02929|Putative;	#N/A
3758	Manganese-dependent protein-tyrosinephosphatase (EC 3.1.3.48)	40	255	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	tyrosine protein phosphatase	Lactobacillus zeae	LGG_02036|wzb|Capsular;	#N/A
3759	hypothetical protein	10	363	#N/A	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	1	1	1	1	1	1	1	0	0	0	0	hypothetical membrane protein	Lactobacillus plantarum		#N/A
3760	dTDP-rhamnosyl transferase RfbF (EC 2.-.-.-)	10	272	#N/A	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	1	1	1	1	1	1	1	0	0	0	0	dTDP-rhamnosyl transferase RfbF	Lactobacillus casei		#N/A
3761	membrane protein involved in the export ofO-antigen, teichoic acid lipoteichoic acids	12	463	#N/A	0	0	0	1	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	1	1	1	1	1	1	1	0	0	0	0	hypothetical protein	Lactobacillus parabrevis		#N/A
3762	Flavodoxin	40	251	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	flavodoxin	Lactobacillus casei subsp. casei ATCC 393	LGG_00221|fldA|Flavodoxin	#N/A
3763	hypothetical protein	14	41	#N/A	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	abortive phage infection protein	Lactobacillus pentosus		#N/A
3764	FIG00748281: hypothetical protein	40	377	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	conserved hypothetical protein	Lactobacillus casei subsp. casei ATCC 393	LGG_00208|LGG_00208|Conserved;	#N/A
3765	Phenylalanyl-tRNA synthetase alpha chain (EC6.1.1.20)	40	356	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	phenylalanyl-tRNA synthetase subunit alpha	Lactobacillus casei ATCC 334	LGG_01718|pheS|Phenylalanyl-tRNA;	#N/A
3766	FIG00748062: hypothetical protein	40	95	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	hypothetical protein	Lactobacillus paracasei	LGG_01971|LGG_01971|Conserved;	#N/A
3767	hypothetical protein	10	47	#N/A	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	1	0	0	0	1	0	1	1	1	0	0	0	1	0	1	0	#N/A			#N/A
3768	RecD-like DNA helicase YrrC	40	853	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	exodeoxyribonuclease V alpha subunit	Lactobacillus casei subsp. casei ATCC 393		#N/A
3769	Antitoxin of toxin-antitoxin stability system	21	89	#N/A	0	1	0	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	prevent-host-death family protein	Lactobacillus paracasei		#N/A
3770	contains glycosyl transferase family 8 domain	40	318	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	lipopolysaccharide biosynthesis glycosyltransferase	Lactobacillus casei subsp. casei ATCC 393	LGG_00999|arbY|Lipopolysaccharide;	#N/A
3771	Transport ATP-binding protein CydC	40	575	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	cytochrome D ABC transporter ATP-binding and permease components	Lactobacillus casei subsp. casei ATCC 393	LGG_02205|cydD|ABC;	#N/A
3772	Transport ATP-binding protein CydD	40	580	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	cysteine ABC transporter ATP-binding protein	Lactobacillus zeae	LGG_02206|cydC|ABC;	#N/A
3773	FIG00750795: hypothetical protein	36	71	#N/A	1	1	1	1	1	1	1	1	0	1	1	1	1	1	0	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	hypothetical protein	Lactobacillus casei	LGG_00389|LGG_00389|Putative;	#N/A
3774	Lj928 prophage protein	5	490	#N/A	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	1	1	1	0	0	0	0	hypothetical protein	Bacillus acidiproducens		#N/A
3775	Prophage Lp1 protein 51	5	237	#N/A	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	1	1	1	0	0	0	0	hypothetical protein	Bacillus cereus		#N/A
3776	phage tail length tape-measure protein	11	1630	#N/A	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	1	1	0	0	0	0	0	1	1	1	1	0	1	0	0	Phage tail length tape-measure protein	Lactobacillus paracasei		#N/A
3777	phage capsid and scaffold	15	128	#N/A	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	1	0	0	1	1	1	1	1	1	1	0	1	1	1	phage-related head-to-tail joining protein	Lactobacillus paracasei		#N/A
3778	putative sugar phosphotransferase component IIB	26	103	#N/A	0	0	0	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	Pentitol phosphotransferase enzyme II, B component	Lactobacillus casei BL23		#N/A
3779	phage terminase, large subunit	18	570	#N/A	0	1	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	1	1	0	0	1	1	1	1	1	1	1	0	1	1	1	large terminase subunit	Lactobacillus phage Lrm1		#N/A
3780	unknown	12	69	#N/A	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	1	1	0	0	0	0	0	1	1	1	1	1	1	0	0	hypothetical protein	Lactobacillus phage Lc-Nu		#N/A
3781	hypothetical protein	2	42	#N/A	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	#N/A			#N/A
3782	PTS system, trehalose-specific IIB component(EC 2.7.1.69) / PTS system, trehalose-specific IIC	40	666	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	beta-glucosides PTS, EIIABC	Lactobacillus casei BL23	LGG_00603|treB|PTS;	#N/A
3783	polysaccharide biosynthesis protein CpsM	23	246	#N/A	1	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	glycosyltransferase	Lactobacillus casei subsp. casei ATCC 393		#N/A
3784	Putative O-unit flippase	23	479	#N/A	1	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	exopolysaccharide synthesis protein	Lactobacillus casei subsp. casei ATCC 393		#N/A
3785	Peroxide stress regulator PerR, FUR family	40	146	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	peptide ABC transporter substrate-binding protein	Lactobacillus paracasei	LGG_00831|perR|Transcriptional;	#N/A
3786	hypothetical protein	34	47	#N/A	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	#N/A			#N/A
3787	D-3-phosphoglycerate dehydrogenase (EC1.1.1.95)	17	325	#N/A	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2-hydroxyacid dehydrogenase	Lactobacillus salivarius		#N/A
3788	Protein export cytoplasm protein SecA2 ATPase	31	1023	#N/A	0	0	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	preprotein translocase subunit SecA	Lactobacillus casei str. Zhang	LGG_01063|LGG_01063|Conserved;	#N/A
3789	Glycerol uptake facilitator protein	40	235	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	glycerol transporter	Lactobacillus zeae	LGG_00642|glpF|Glycerol;	#N/A
3790	FIG00745092: hypothetical protein	40	99	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	hypothetical protein	Lactobacillus zeae	LGG_00711|LGG_00711|Conserved;	#N/A
3791	FIG00744595: hypothetical protein	40	157	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	hypothetical protein BN194_07820	Lactobacillus casei W56	LGG_00702|LGG_00702|Conserved;	#N/A
3792	Phospho-N-acetylmuramoyl-pentapeptide-transferase (EC 2.7.8.13)	40	321	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	phospho-N-acetylmuramoyl-pentapeptide-transferase	Lactobacillus zeae	LGG_01281|mraY|Phospho-N-acetylmuramoyl-pentapeptide-;	#N/A
3793	hypothetical protein	13	54	#N/A	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	1	1	1	1	1	1	1	1	0	0	1	0	0	#N/A			#N/A
3794	Putative inner membrane protein	40	231	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	hypothetical protein	Lactobacillus paracasei	LGG_02710|LGG_02710|Conserved;	#N/A
3795	hypothetical protein	21	63	#N/A	0	0	0	0	0	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	1	0	0	0	1	1	0	0	1	1	1	0	#N/A			#N/A
3796	FIG00744451: hypothetical protein	40	49	#N/A	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	hypothetical protein LSEI_2434	Lactobacillus casei ATCC 334	LGG_02427|LGG_02427|Conserved;	#N/A
3797	hypothetical protein	19	58	#N/A	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	#N/A			#N/A
3798	LSU ribosomal protein L1p (L10Ae)	40	229	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	50S ribosomal protein L1	Lactobacillus zeae	LGG_02288|rplA|LSU/50S;	#N/A
3799	Biotin-protein ligase (EC 6.3.4.15) / Biotinoperon repressor	40	248	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	biotin operon repressor	Lactobacillus casei subsp. casei ATCC 393		#N/A
3800	Predicted multitransmembrane protein	40	245	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	conserved hypothetical protein	Lactobacillus casei subsp. casei ATCC 393	LGG_00802|LGG_00802|Conserved;	#N/A
3801	Poly(glycerol-phosphate)alpha-glucosyltransferase (EC 2.4.1.52)	40	538	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	glycosyl transferases group 1 family protein	Lactobacillus paracasei	LGG_00349|yohH|Poly(Glycerol-phosphate);	#N/A
3802	FIG146262: hypothetical protein	40	266	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	UPF0135 protein, partial	Lactobacillus paracasei	LGG_01509|LGG_01509|NIF3-related;	#N/A
3803	FIG00745626: hypothetical protein	40	447	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	conserved hypothetical protein	Lactobacillus casei subsp. casei ATCC 393	LGG_00861|LGG_00861|Conserved;	#N/A
3804	NUDIX family hydrolase	40	166	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	NUDIX family hydrolase	Lactobacillus paracasei	LGG_00855|LGG_00855|Phosphohydrolase,;	#N/A
3805	FIG00750064: hypothetical protein	12	833	#N/A	0	0	1	1	0	0	0	0	1	1	1	0	1	0	1	1	0	0	1	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	phage tail assembly	Lactobacillus casei	LGG_01139|LGG_01139|Phage-related;	#N/A
3806	putative head-tail joining protein	18	109	#N/A	0	1	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	1	1	0	0	1	1	1	1	1	1	1	0	1	1	1	phage-related head-tail joining protein	Lactobacillus paracasei		#N/A
3807	hypothetical protein	3	72	#N/A	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	hypothetical protein	Lactobacillus paracasei		#N/A
3808	hypothetical protein	9	72	#N/A	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	1	1	1	1	1	1	1	0	0	0	0	peptidase	Lactobacillus casei subsp. casei ATCC 393		#N/A
3809	FIG00750939: hypothetical protein	40	512	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	hypothetical protein	Lactobacillus zeae	LGG_01245|LGG_01245|Conserved;	#N/A
3810	hypothetical protein	14	805	#N/A	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	hypothetical protein	Lactobacillus coryniformis		#N/A
3811	Redox-sensitive transcriptional regulator(AT-rich DNA-binding protein)	40	221	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	redox-sensing transcriptional repressor Rex	Lactobacillus casei subsp. casei ATCC 393	LGG_01260|rex|Redox-sensing;	#N/A
3812	Universal stress protein family	40	166	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	universal stress protein UspA	Lactobacillus zeae	LGG_01208|usp|Universal;	#N/A
3813	L-proline glycine betaine ABC transport systempermease protein ProV (TC 3.A.1.12.1)	40	319	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	ABC transporter ATP-binding protein	Lactobacillus zeae		#N/A
3814	prophage Lp3 protein 18	33	523	#N/A	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	0	0	1	1	1	1	1	0	0	1	0	1	1	1	1	1	1	1	0	1	1	1	phage protein	Lactobacillus casei	LGG_02901|LGG_02901|Phage-related;	#N/A
3815	Maltodextrin glucosidase (EC 3.2.1.20)/EC number=3.2.1.20	17	749	#N/A	0	0	0	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	alpha-glucosidase	Lactobacillus casei subsp. casei ATCC 393		#N/A
3816	hypothetical protein	18	37	#N/A	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	#N/A			#N/A
3817	hypothetical protein	40	82	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	hypothetical protein LCAZH_0137	Lactobacillus casei str. Zhang	LGG_02790|LGG_02790|Conserved;	#N/A
3818	Substrate-specific component PdxT of predictedpyridoxine ECF transporter	40	182	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	membrane protein	Lactobacillus zeae	LGG_00594|LGG_00594|Putative;	#N/A
3819	FIG145533: Methyltransferase (EC 2.1.1.-)	40	245	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	SAM-dependent methyltransferase	Lactobacillus zeae	LGG_01733|yqeM|Methyltransferase	#N/A
3820	tRNA pseudouridine synthase B (EC 4.2.1.70)/EC number=4.2.1.70	40	301	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	tRNA pseudouridine synthase B	Lactobacillus zeae	LGG_01609|truB|tRNA;	#N/A
3821	hypothetical protein	8	45	#N/A	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	1	1	1	hypothetical protein	Lactobacillus paracasei		#N/A
3822	NAD-dependent protein deacetylase of SIR2family	40	230	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	NAD-dependent deacetylase	Lactobacillus casei subsp. casei ATCC 393	LGG_02593|npdA|NAD-dependent;	#N/A
3823	Transcriptional regulator, MarR family	40	180	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	MarR family transcriptional regulator	Lactobacillus zeae	LGG_02127|LGG_02127|Transcriptional;	#N/A
3824	hypothetical protein	2	37	#N/A	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	#N/A			#N/A
3825	hypothetical protein	2	43	#N/A	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	#N/A			#N/A
3826	FIG00749021: hypothetical protein	40	164	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	hypothetical protein	Lactobacillus zeae	LGG_02232|LGG_02232|Conserved;	#N/A
3827	hypothetical protein	8	37	#N/A	1	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	1	0	0	0	1	0	0	1	0	0	1	1	0	0	0	0	0	0	0	0	#N/A			#N/A
3828	FIG00747050: hypothetical protein	40	419	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	ABC transporter permease	Lactobacillus zeae	LGG_00962|natB|ABC;	#N/A
3829	hypothetical protein	5	49	#N/A	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	1	0	1	0	0	0	0	#N/A			#N/A
3830	hypothetical protein	6	50	#N/A	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	1	0	0	0	1	0	1	0	0	0	0	0	0	#N/A			#N/A
3831	transcription regulator	19	569	#N/A	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	transcription regulator	Lactobacillus paracasei		#N/A
3832	Cellobiose phosphotransferase system YdjC-likeprotein	12	260	#N/A	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	0	hypothetical protein	Lactococcus garvieae		#N/A
3833	putative hydrolase YitU	19	267	#N/A	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	Hydrolase, HAD superfamily	Lactobacillus casei LOCK919		#N/A
3834	Transport protein SgaT, putative	19	437	#N/A	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	Putative integral membrane protein	Lactobacillus casei LOCK919		#N/A
3835	RecA protein	40	395	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	protein recA	Lactobacillus casei BL23	LGG_00889|recA|Protein;	#N/A
3836	Ribosomal-protein-S5p-alanineacetyltransferase	40	242	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	acetyltransferase	Lactobacillus casei subsp. casei ATCC 393	LGG_01901|LGG_01901|Ribosomal-protein-S5-alanine;	#N/A
3837	hypothetical protein	3	46	#N/A	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	#N/A			#N/A
3838	FIG00752295: hypothetical protein	23	92	#N/A	1	0	1	1	1	1	1	0	1	0	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	1	0	1	1	1	1	1	1	1	0	0	0	0	hypothetical protein	Lactobacillus zeae	LGG_02887|LGG_02887|Putative;	#N/A

| | | |

| | | |

| | | |

| | | |

4221	hypothetical protein	1	188	#N/A	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	uncharacterized protein BN782_00673	Eubacterium sp. CAG:786		#N/A
4222	hypothetical protein	1	187	#N/A	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	hypothetical protein LBCZ_2524	Lactobacillus casei subsp. casei ATCC 393		#N/A
4223	hypothetical protein	1	186	#N/A	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	hypothetical protein	Lactobacillus pentosus		#N/A
4224	hypothetical protein	1	184	#N/A	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	hypothetical protein	Lactobacillus casei		#N/A
4225	FIG00754149: hypothetical protein	1	181	#N/A	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Hypothetical protein LOCK919_2121	Lactobacillus casei LOCK919		#N/A
4226	phage capsid and scaffold	1	180	#N/A	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	putative scaffold protein	Lactobacillus casei		#N/A
4227	FIG00748045: hypothetical protein	1	178	#N/A	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	hypothetical protein	Lactobacillus casei		#N/A
4228	phage tail fiber protein	2	177	#N/A	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	L-aspartate oxidase	Pseudomonas syringae		#N/A
4229	Transposase	1	176	#N/A	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	ISSha1-like transposase	Enterococcus faecalis 62		#N/A
4230	Alpha/beta superfamily hydrolase	2	175	#N/A	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	predicted esterase protein-like	Lactobacillus paracasei subsp. paracasei		#N/A
4231	hypothetical protein	1	174	#N/A	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	hypothetical protein	Lactobacillus paracasei		#N/A
4232	FIG00750704: hypothetical protein	1	173	#N/A	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	hypothetical protein LBPG_01288	Lactobacillus paracasei subsp. paracasei 8700:2		#N/A
4233	FIG00745582: hypothetical protein	40	172	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	hypothetical protein	Lactobacillus paracasei	LGG_00576|LGG_00576|Conserved;	#N/A
4234	hypothetical protein	1	172	#N/A	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	hypothetical protein	Lactobacillus casei		#N/A
4235	hypothetical protein	1	172	#N/A	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	hypothetical protein	Lactobacillus buchneri		#N/A
4236	hypothetical protein	1	171	#N/A	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	hypothetical protein	Lactobacillus murinus		#N/A
4237	hypothetical protein	1	171	#N/A	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	conserved hypothetical protein	Lactobacillus casei subsp. casei ATCC 393		#N/A
4238	hypothetical protein	1	168	#N/A	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	hypothetical protein LSEI_1963	Lactobacillus casei ATCC 334		#N/A
4239	hypothetical membrane protein	1	166	#N/A	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	membrane protein	Lactobacillus zeae		#N/A
4240	hypothetical	1	163	#N/A	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	hypothetical protein	Lactobacillus paracasei		#N/A
4241	hypothetical protein	1	159	#N/A	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	hypothetical protein LBPG_00200	Lactobacillus paracasei subsp. paracasei 8700:2		#N/A
4242	PTS system, galactitol-specific IIA component(EC 2.7.1.69)	1	159	#N/A	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	PTS system galactitol-specific IIA component	Lactobacillus paracasei		#N/A
4243	major capsid protein gpP	1	158	#N/A	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	phage major capsid protein	Lactobacillus paracasei		#N/A
4244	hypothetical protein	1	158	#N/A	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	hypothetical protein LCBD_p16	Lactobacillus casei BD-II		#N/A
4245	Teichoic acid glycosylation protein	1	157	#N/A	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	putative teichoic acid glycosylation protein	Lactobacillus paracasei		#N/A
4246	phage terminase small subunit	1	157	#N/A	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	phage terminase small subunit	Lactobacillus paracasei		#N/A
4247	hypothetical protein	1	154	#N/A	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	hypothetical protein	Lactobacillus paracasei		#N/A
4248	hypothetical protein	1	153	#N/A	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	hypothetical protein N134_05705	Lactobacillus reuteri TD1		#N/A
4249	phage terminase, small subunit	1	151	#N/A	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	hypothetical protein	Lactobacillus paracasei		#N/A
4250	hypothetical protein	2	150	#N/A	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	hypothetical protein	Lactobacillus paracasei		#N/A
4251	hypothetical protein	1	150	#N/A	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	hypothetical protein	Lactobacillus paracasei		#N/A
4252	hypothetical protein	1	149	#N/A	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	hypothetical protein	Lactobacillus zeae		#N/A
4253	prophage pi1 protein 32	1	149	#N/A	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	Phage endonuclease	Lactobacillus paracasei		#N/A
4254	hypothetical protein	1	148	#N/A	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Mob	Lactobacillus paracasei		#N/A
4255	conserved hypothetical protein	1	146	#N/A	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	hypothetical protein	Lactobacillus paracasei		#N/A
4256	hypothetical protein	1	144	#N/A	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	predicted protein	Lactobacillus fermentum		#N/A
4257	hypothetical protein	1	143	#N/A	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	hypothetical protein	Lactobacillus paracasei		#N/A
4258	transcriptional regulator, XRE family	1	143	#N/A	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	XRE family transcriptional regulator	Lactobacillus paracasei		#N/A
4259	N-acetylglucosamine/galactosamine PTS, EIIA(EC:2.7.1.69)	1	142	#N/A	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	PTS system,N-acetylglucosamine/galactosamine-specific EIIA component	Lactobacillus paracasei		#N/A
4260	FIG00632156: hypothetical protein	1	141	#N/A	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	conserved hypothetical protein	Enterococcus italicus		#N/A
4261	Putative uncharacterized protein	1	141	#N/A	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	hypothetical protein LCABL_09970	Lactobacillus casei BL23		#N/A
4262	PTS system, N-acetylgalactosamine- andgalactosamine-specific IIA component (EC 2.7.1.69)	1	140	#N/A	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	PTS sugar transporter	Lactobacillus gasseri		#N/A
4263	hypothetical protein	1	139	#N/A	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	hypothetical protein	Lactobacillus paracasei		#N/A
4264	hypothetical protein	1	139	#N/A	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	hypothetical protein LSEI_0492	Lactobacillus casei ATCC 334		#N/A
4265	Phosphate transport ATP-binding protein PstB(TC 3.A.1.7.1)	40	138	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	phosphate ABC transporter ATP-binding protein	Lactobacillus zeae	LGG_00910|pstB|ABC;	#N/A
4266	hypothetical protein	1	137	#N/A	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	hypothetical protein	Lactobacillus paracasei		#N/A
4267	hypothetical protein	1	136	#N/A	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	hypothetical protein	Lactobacillus pentosus		#N/A
4268	orf15	1	136	#N/A	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	hypothetical phage protein	Lactobacillus casei subsp. casei ATCC 393		#N/A
4269	hypothetical protein	1	135	#N/A	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	hypothetical protein	Gracilimonas tropica		#N/A
4270	mobile element protein	7	134	#N/A	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	1	1	1	0	0	0	0	0	1	0	0	transposase, IS5, partial	Lactobacillus paracasei		#N/A
4271	Chaperone protein DnaJ	1	133	#N/A	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	molecular chaperone DnaJ	Lactobacillus zeae		#N/A
4272	hypothetical protein	1	132	#N/A	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	hypothetical protein	Lactobacillus casei		#N/A
4273	minor capsid protein	1	130	#N/A	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	phage minor capsid protein	Lactobacillus paracasei		#N/A
4274	hypothetical protein	1	129	#N/A	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	hypothetical protein	Lactobacillus paracasei		#N/A
4275	hypothetical protein	1	129	#N/A	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	hypothetical protein	Lactobacillus paracasei		#N/A
4276	nucleoid DNA-binding protein	1	129	#N/A	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	histone family protein DNA-binding protein	Lactobacillus paracasei		#N/A
4277	ORF42	1	128	#N/A	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	hypothetical protein	Lactobacillus paracasei		#N/A
4278	ABC transporter ATP-binding/membrane spanningprotein - multidrug resistance	40	128	core	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	ABC transporter permease	Lactobacillus zeae	LGG_02936|LGG_02936|ABC;	#N/A
4279	hypothetical protein	1	127	#N/A	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	hypothetical protein	Lactobacillus paracasei		#N/A
4280	phage protein	1	127	#N/A	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Phage protein	Lactobacillus casei LOCK919		#N/A
4281	prophage LambdaBa04, Gp54	1	126	#N/A	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	hypothetical protein	Lactobacillus paracasei		#N/A
4282	hypothetical protein	1	126	#N/A	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	hypothetical protein	Oenococcus oeni		#N/A
4283	hypothetical protein	1	124	#N/A	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	PREDICTED: laminin subunit alpha-1	Ictidomys tridecemlineatus		#N/A
4284	putative head to tail joining	1	124	#N/A	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	head to tail joining protein	Lactobacillus paracasei		#N/A
4285	FIG00743773: hypothetical protein	1	124	#N/A	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	hypothetical protein LBPG_00819	Lactobacillus paracasei subsp. paracasei 8700:2		#N/A
4286	Putative uncharacterized protein	1	124	#N/A	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	hypothetical protein BN194_10780	Lactobacillus casei W56		#N/A
4287	Transcriptional regulator of sugar metabolism	19	122	#N/A	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	Lactose phosphotransferase system repressor	Lactobacillus casei LOCK919		#N/A
4288	hypothetical protein	1	122	#N/A	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	hypothetical protein	Lactobacillus paracasei		#N/A
4289	phage protein	1	121	#N/A	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	phage protein	Lactobacillus casei		#N/A
4290	hypothetical protein	1	121	#N/A	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	hypothetical protein	Lactobacillus paracasei		#N/A
4291	ORF042	1	120	#N/A	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	DNA packaging protein	Lactobacillus paracasei		#N/A
4292	hypothetical protein	1	120	#N/A	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	hypothetical protein	Lactobacillus kisonensis		#N/A
4293	mobilization protein	1	117	#N/A	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	mobilization protein	Lactobacillus buchneri		#N/A
4294	hypothetical protein	1	116	#N/A	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	hypothetical protein	Lactobacillus paracasei		#N/A
4295	hypothetical protein	1	115	#N/A	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	hypothetical protein	Lactobacillus paracasei		#N/A
4296	hypothetical protein	1	114	#N/A	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	#N/A			#N/A
4297	hypothetical protein	1	113	#N/A	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Hypothetical protein LOCK919_2110	Lactobacillus casei LOCK919		#N/A
4298	hypothetical protein	1	113	#N/A	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	AraC family transcriptional regulator	Lactobacillus helveticus R0052		#N/A
4299	N-acetylmuramic acid 6-phosphate etherase (EC4.2.-.-)	1	112	#N/A	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	N-acetylmuramic acid-6-phosphate etherase	Lactobacillus zeae		#N/A
4300	PapB	1	112	#N/A	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	PapB (putative pediocin Ach/PA-1 immunity protein)	Lactobacillus pentosus IG1		#N/A
4301	ATP-dependent protease (EC 3.4.21.-)	1	112	#N/A	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	MobC	Lactobacillus plantarum		#N/A
4302	hypothetical protein	1	112	#N/A	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	hypothetical protein N692_08915	Lactobacillus plantarum EGD-AQ4		#N/A
4303	Transcriptional regulator, xre family	1	111	#N/A	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	XRE family transcriptional regulator	Lactobacillus brevis ATCC 367		#N/A
4304	Arsenical resistance operon trans-actingrepressor ArsD	1	111	#N/A	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	arsenical resistance operon trans-acting repressor	Lactobacillus casei subsp. casei ATCC 393		#N/A
4305	hypothetical protein	1	110	#N/A	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	transcriptional regulatory family protein	Clostridium difficile CD69		#N/A
4306	hypothetical protein	1	110	#N/A	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	hypothetical protein lb338_phage_175	Lactobacillus phage Lb338-1		#N/A
4307	Putative uncharacterized protein	1	109	#N/A	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Phage protein	Lactobacillus paracasei		#N/A
4308	Non-specific DNA-binding protein Dps /Iron-binding ferritin-like antioxidant protein /	1	109	#N/A	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	non-heme iron-containing ferritin	Lactobacillus paracasei		#N/A
4309	hypothetical protein	1	109	#N/A	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	hypothetical protein	Lactobacillus rennini		#N/A
4310	ORF-11	1	108	#N/A	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	hypothetical protein	Lactobacillus paracasei		#N/A
4311	mobilization protein	1	108	#N/A	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	mobilization protein	Lactobacillus paracasei		#N/A
4312	phage protein	1	108	#N/A	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	phage protein	Lactobacillus casei BL23		#N/A
4313	Arsenical resistance operon repressor	1	107	#N/A	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	transcriptional regulator, ArsR family	Lactobacillus buchneri NRRL B-30929		#N/A
4314	killer suppression protein HigA, putative	1	107	#N/A	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Plasmid maintenance system killer protein	Lactobacillus plantarum ZJ316		#N/A
4315	hypothetical protein	1	106	#N/A	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	toxin-antitoxin system, antitoxin, xre/HigA/VapI family	Lactobacillus pentosus		#N/A
4316	Thioredoxin	1	106	#N/A	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	thiol-disulfide isomerase	Lactobacillus pentosus		#N/A
4317	membrane protein involved in the export ofO-antigen, teichoic acid lipoteichoic acids	10	104	#N/A	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Membrane protein involved in the export of O-antigen, teichoic acid lipoteichoic acids	Lactobacillus casei LOCK919	LGG_02049|wzx|Polysaccharide;	#N/A

| | | |

| | | |

| | | |

| | | |