CORA Cork Open Research Archive
Cartlann Taighde Oscailte Chorcaí

Title	Genome-scale analyses of health-promoting bacteria: probiogenomics
Authors	Ventura, Marco;O'Flaherty, Sarah;Claesson, Marcus J.;Turroni, Francesca;Klaenhammer, Todd R.;van Sinderen, Douwe;O'Toole, Paul W.
Publication date	2009-01-01
Original Citation	Ventura M., O'Flaherty S., Claesson M. J., Turroni F., Klaenhammer T.R., van Sinderen D., O'Toole P.W. 2009. Genome-scale analyses of health-promoting bacteria: probiogenomics. Nature Reviews Microbiology 7, pp61-71 doi:10.1038/nrmicro2047
Type of publication	Article (peer-reviewed)
Link to publisher's version	10.1038/nrmicro2047
Rights	© 2009 Macmillan Publishers Limited. All rights reserved - http://www.nature.com/authors/editorial_policies/license.html
Download date	2026-10-09 11:19:12
Item downloaded from	https://hdl.handle.net/10468/118

UCC
University College Cork, Ireland
Coláiste na hOllscoile Corcaigh

ucc
Coláiste na hOllscoile Corcaigh, Éire
University College Cork, Ireland

Ventura M., O'Flaherty S., Claesson M. J., Turroni F., Klaenhammer T.R., van Sinderen D., O'Toole P. W. 2009. Genome-scale analyses of health-promoting bacteria: probiogenomics. Nature Reviews Microbiology 7, pp61-71 doi:10.1038/nrmicro2047

Deposited on: 16 Feb 2010

CORA Cork Open Research Archive http://cora.ucc.ie

Genome-scale analyses of health-promoting bacteria: probiogenomics

Key words: genomics, functional genomics, probiotic bacteria, intestinal tract, microbiota

Marco Ventura[*], Sarah O'Flaherty[‡], Marcus J. Claesson[§], Francesca Turroni[*], Todd R. Klaenhammer[‡], Douwe van Sinderen[§], and Paul W. O'Toole[§]

[*]Department of Genetics, Biology of Microorganisms, Anthropology and Evolution, University of Parma, Italy, [‡]Department of Food, Bioprocessing and Nutrition Sciences, North Carolina State University, Raleigh, North Carolina 27695, [§]Alimentary Pharmabiotic Centre and Department of Microbiology, University College Cork, Western Road, Cork, Ireland

Correspondence to P.W.O

e.mail: pwotoole@ucc.ie

Abstract

The human body is colonized by an enormous population of bacteria (microbiota) that outnumbers the human somatic and germ cells and provides the host with additional coding capacity and metabolic activities. Among the human gut microbiota are health-promoting indigenous species, also referred to as probiotic bacteria, which are commonly consumed as live dietary supplements. Although there is a growing list of health benefits provided by the consumption of probiotics, their precise mechanisms of action remain largely unknown. Recent genomics based studies (probiogenomics) are starting to provide insights into the ways probiotic bacteria sense and adapt to the gastrointestinal tract environment. In this review, we will discuss the application of probiogenomics in the elucidation of the molecular basis of probiosis using the well recognized model probiotic bacteria *Bifidobacterium* and *Lactobacillus* as examples.

The availability of the sequence of the human genome has paved the way for a better understanding of the genetic basis for many aspects of human health and disease. However, fully understanding the human genotype, and its relationship with health and disease susceptibility, requires better information explaining how environmental and developmental factors interact with the genome to influence health status. Human beings are colonized by, or transiently harbour, a wide, complex and dynamic collection of bacteria that outnumber the human somatic and germ cells, and that collectively represent significantly more genetic variety than the genome of their host[1]. However, at the present time, the components of the human microbiota remain poorly identified and characterized. Recent culture-independent studies of the microbiota of the human gastrointestinal tract (GIT) have identified more than 1000 phylotypes, representing over 7000 strains and belonging to eight major phyla[1-4] (see also[5] for an overview).

It has been suggested that the composition of the gut microbiota is the result of selective pressure imposed by the host, and further modulated by competition between constituent bacterial members[6]. The interactions between various bacteria and the human host can be categorized as a continuum ranging from symbiosis to commensalism and through to pathogenesis, where the two former relationships can be grouped as mutualism (Fig. 1). In the human gut environment, the adaptive co-evolution of humans and bacteria may lead to the development of commensal relationships, where neither partner is disadvantaged, or symbiotic relationships where unique metabolic activities or other benefits are provided. The intestinal microbiota contributes to host nutrition[1, 7, 8] and it impacts on intestinal cell proliferation and differentiation, pH, the development of the immune system and innate and acquired response to pathogens[1, 9, 10].

Alterations in the composition of the intestinal microbiota have recently been linked to a variety of conditions ranging from Inflammatory Bowel Disease to allergy and obesity[6, 11-14]. Among the variable constituents of the microbiota are health-promoting indigenous species (or autochthonous microbiota), also known as probiotic bacteria, which are commonly consumed as

live dietary supplements[15]. The mechanisms by which probiotic micro-organisms beneficially affect human health (reviewed in[16, 17]) are typically divided into a number of general categories, including 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 mucosal binding sites[16, 18]. Although there is suggestive evidence for each of these functional claims, the molecular mechanisms remain largely unknown.

Genomics offers the possibility of accelerating research into probiotic bacteria. In recent years, genome sequencing of gut commensals and symbionts has come to the fore, currently represented by the development of a novel scientific discipline, called probiogenomics[19], which aims to provide insights into the diversity and evolution of commensal/probiotic bacteria and to reveal the molecular basis for their health-promoting activities. The integration of probiogenomics and functional genomic information with data on host gene expression in the human gut will expand our understanding of the roles of (probiotic) microbiota, microbe-microbe and host-microbe interactions. These "omics" approaches allow the simultaneous analysis of very large numbers of genes or proteins[20]. Probiogenomics is thus one strand of gut systems microbiology. Significantly, when studied in combination with host genome variation, probiogenomics offers a comprehensive systems model, even at individual subject level.

Here we address current developments in analyzing the genome sequences of probiotic bacteria and how these data can be integrated in a global view using omics approaches in order to elucidate genome evolution and genetic adaptation of these bacteria to the human gut ecological niche. We consider the well recognized model probiotic bacteria *Bifidobacterium* spp. and *Lactobacillus* spp. which are phylogenetically distant (although well-characterized; Fig 1), have distinguishing properties, and different depths of biological characterization.

Bifidobacteria genomics

The genus *Bifidobacterium* is relatively small, with 30 species, and a low level of phylogenetic and genomic diversity[21]. Bifidobacteria were originally isolated from a breast-fed infant[23] and since then, 30 species have been isolated from the GIT contents of mammals, birds and insects[19]. Those bifidobacteria that may be isolated from the human intestine have attracted the interest of genomic research due to their probiotic properties. However, of the bifidobacterial taxa described to date, genomes of only three strains, which belong to the *B. longum* and *B. adolescentis* groups, have been sequenced to completion (Table 1). The availability of genome sequences provided a genetic basis for the observation that bifidobacteria are extensively prototrophic, indicating that these bacteria are well adapted to grow in an environment such as the human colon, which is poor in certain growth substrates (e.g. vitamins, amino acids and nucleotides)[24]. In fact, bifidobacterial genome sequences available to date revealed that these organisms harbour genes for the synthesis of at least 19 amino acids and they encode all enzymes needed for the biosynthesis of pyrimidine and purine nucleotides, as well as those required for the synthesis of the B vitamins, folic acid, thiamine and nicotinate ([25]; Ventura et al., unpublished data; Leahy and D. van Sinderen, unpublished data). Annotation and pathway prediction revealed the presence of all the required genetic information to shunt many monosaccharides or disaccharides into the fructose-6-phosphate pathway[24].

Adaptation to the human gut.

The amount and types of "non-digestible" saccharides in the diet (some of which are referred to as prebiotics) has a major influence on the numbers and metabolic activities of different groups of bacteria within the enteric microbiota[26]. The range of polysaccharide substrates that arrive in the intestine is extremely broad[27]. This diversity of carbon substrates potentially generates a vast array of ecological roles and niches that may be exploited by gut bacteria. Although some members of the gut microbiota can switch rapidly between different substrates (e.g. derived from diet or of host origin), others (e.g. those associated with insoluble substrates) are much more specialized[28]. In this context, bifidobacteria have a presumed ecological advantage due to their capacity to metabolize

116 complex sugars derived from the diet as well as from the host[29]. Genome annotation confirms that
117 genes required for the breakdown of complex sugars are abundant in sequenced bifidobacterial
118 genomes[19]. Over 8% of annotated bifidobacterial genes encode enzymes involved in carbohydrate
119 metabolism. These include various glycosyl hydrolases (GH) for utilization of diverse, but in most
120 cases un-identified, plant-derived dietary fibers or complex carbohydrate structures. Most of the
121 bifidobacterial GHs are predicted to be intracellular including those that are thought to hydrolyze
122 arabinogalactans and arabinoxylans, or starch and related polysaccharides[25, 30, 31]. The genes for
123 these GHs are associated with genetic loci for the uptake of structurally diverse sugar substrates. In
124 fact, about 5% of the total bifidobacterial gene content is dedicated to sugar internalization, through
125 ABC transporters, permeases, and proton symporters rather than phosphoenolpyruvate-
126 phosphotransferase systems (PEP-PTSs)[25, 32, 33]. Bifidobacteria utilize a kind of docking station to
127 sequester and capture high molecular weight carbohydrates molecules (e.g., xylose- and arabinose-
128 containing polysaccharides; Fig. 2) and bind these to their cell surface[30, 33], presumably to avoid
129 losing them to nearby competitors. This is reminiscent of a putative carbohydrate utilization system
130 identified in the genome of *L. plantarum*[34], and a system used by *Bacteriodes thetaiotaomicron* for
131 starch utilization[35]. Enteric bifidobacteria are also able to utilize sialic acid-containing complex
132 carbohydrates in mucin, glycosphingolipids and human milk[36, 37]. Thus, these bifidobacteria have
133 acquired adaptations to allow them to exploit a rich repertoire of otherwise indigestible components
134 of the human or animal diet.
135 Characterization of the metabolism of prebiotic compounds by bifidobacteria has identified specific
136 transporters and hydrolases for oligosaccharides[30, 38, 39]. These studies indicated that bifidobacteria
137 ferment different types of fructo-oligosaccharides (FOS); accordingly, the respective FOS
138 metabolism operons possess different genetic architectures[40], suggesting that these genes were
139 acquired following evolutionary divergence of the species. Prebiotic oligosaccharides are also
140 contained in human milk (e.g., galacto-oligosaccharides), which are hydrolyzed by bifidobacteria
141 through the action of extracellular enzymes encoded by the *gal*A gene[30, 41]. In addition to galacto-

oligosaccharides, human milk consumption provides large amounts of small peptides that are derived from the digestion of milk proteins by the gastric protease pepsin[42]. *Bifidobacterium* genomes encode a rich repertoire of enzymes involved in the breakdown and internalization of peptides such as dipeptidyl aminopeptidases and oligopeptide uptake systems (Ventura et al., unpublished data).

Molecular interaction with the host.

Bacterium-host interactions that result in host benefit can be elucidated by identification and detailed molecular analysis of the bacterial proteins or macromolecules involved. For example a potential probiotic effector molecule, a eukaryotic-type serine protease inhibitor (serpin) was identified in the genome of *B. longum* subsp. *longum*[25, 43]. Members of the serpin family regulate a wide range of signalling pathways in eukaryotes and some are recognized for their ability to suppress inflammatory responses by inhibiting elastase activity[44]. Recent findings showed that the bifidobacterial serpin-like protein performs an immunomodulatory role in a murine colitis model, by reducing intestinal inflammation [43].

Transcriptomic approaches facilitate studies of gene expression profiles and have been successful in studying how individual organisms in bacterial communities affect each other's transcriptome. Recent transcriptomic analyses were performed on bacteria from germ-free mice that had been mono-associated with *B. thetaiotaomicron* —one of the dominant components of the human gut microbiota — and subsequently challenged with *B. longum* subsp. *longum* . The presence of *B. longum* subsp. *longum* provoked an expansion in the diversity of polysaccharides targeted for breakdown by *B. thetaiotaomicron* such as mannose and xylose-containing glycans[45]. The changes in the transcriptional profiles of polysaccharide-utilization related genes by *B. longum* subsp. *longum* and *B. thetaiotaomicron* may imply the existence of symbiosis between these microbial species, where each species possesses a complement of GH activities, which when combined allow both to participate in a synergic harvest of xylose and mannose-containing sugars. This phenomenon has already been described in other microbial communities that degrade

cellulose[46]. Alternatively, the shifts in transcription patterns could represent response to competition (see also below for lactobacilli).

The elucidation of the molecular impact generated by members of the human microbiota on the human host was also analysed by studying the host epithelium response to co-colonization by *B. longum* subsp. *longum* and *B. thetaiotaomicron*[45]. Remarkably, the host response to these two bacterial species was different. In fact, the host response to *B. thetaiotaomicron* was more focused on tumor necrosis factor α and LPS-responsive cytokine produced by natural killer and T macrophages, whereas *B. longum* subsp. *longum* promoted the activation of T-cell-produced cytokine interferon-γ and reduced production by the host of antibacterial proteins such as Reg3γ (Regenerating islet-derived-3γ) and Pap (Pancreatitis-associated protein). Thus the host response to enteric bifidobateria may not only promote their own survival in the human intestine but also affect the composition of the overall human gut microbiota.

Comparative genomics of bifidobacteria

Comparisons at the nucleotide level of the fully sequenced bifidobacterial genomes revealed a high degree of conservation and synteny across the entire genomes[19]. However, several breakpoint regions were also reported, apparently representing inversions or DNA deletion/insertion points. DNA regions uniquely present in one genome and absent in others were also identified. Most of these correspond to genetic elements presumably acquired by horizontal gene transfer events (HGT), including prophage-like elements, restriction modification systems, integrative plasmids, and genes involved in the biosynthesis of extracellular structures such as exopolysaccharides (EPS) (Fig. 3). Another set of genes disseminated via HGT in bifidobacteria is the CRISPR-related system (CASS) implicated in defence against phages and plasmids[47], which have been identified in the genome of *B. dentium* Bd1 as well as in the genome of *B. breve* UCC2003 (Ventura et al., unpublished data; Leahy and D. van Sinderen, unpublished data). Notably these *in silico* analyses were also confirmed by comparative genome hybridization analyses[48].

There is relatively little phylogenetic diversity within the genus *Bifidobacterium* compared to *Lactobacillus* (see below). This is underlined at whole genome level when one compares the oral species (*B. dentium*), which is frequently identified as a component of the microbiota associated with dental caries[49] with the probiotic species *B. adolescentis* (Fig. 3). Despite the large phenotypic differences, there is a remarkable degree of overall synteny. This reductionist model of genome evolution may be useful for identifying niche-specific genes and genes related to specialized phenotypes.

Genomics of *Lactobacillus*

The genus *Lactobacillus* has more than 100 species, and is noteworthy for its extreme phylogenetic, phenotypic and ecological diversity[22]. The microbiological characterization of lactobacilli is historically better developed than that of bifidobacteria, but the genomic analysis is similarly recent. Of the 14 sequenced and published *Lactobacillus* genomes, eight (*L. acidophilus*, *L. casei*, *L. fermentum*, *L. gasseri*, *L. johnsonii*, *L. reuteri*, *L. salivarius* and *L. plantarum*) are from cultures/species considered probiotic (Table 1). Interestingly, 11% of the overall coding capacity of the *L. salivarius* genome lies on the first megaplasmid described in lactic acid bacteria; pMP118[22]. This megaplasmid encodes biologically important features such as a locus for bacteriocin production, a bile salt hydrolase, and two genes that complete the phosphoketolase pathway, officially reclassifying this organism as a facultative heterofermenter[22]. In fact, plasmids account for 15% of the genome of *L. salivarius*, which is not the case with other sequenced probiotic lactobacilli, even though members of this genus are considered relatively replete with plasmids[9].

*Adaptation to the human gut***.**

The metabolic diversity revealed by the *Lactobacillus* genome sequences available to date is illustrated in Fig 4. Taking the *L. plantarum* WCFS1 genome as reference, it is clear that there is considerable variation in the COG assignments of the gene sets harboured by the respective genomes. Intestinal lactobacilli compensate for their relative degree auxotrophy by being rich in

genes for transporters. Their genomes also contain genes that encode acid and bile resistance, capacity for uptake of macromolecules, metabolism of complex carbohydrates, and cell surface proteins that interact with the intestinal mucosa[60]. More strikingly than is evident for bifidobacteria, this adaptation to life in the GIT is further evident when the genome sequences of intestinal isolates are compared with food-adapted lactobacilli such as *L. bulgaricus* and *L. helveticus*. *L. bulgaricus*. The latter, which is widely used as a starter culture in yogurt fermentations, has undergone genome decay to adapt to the milk environment[53], and thus harbours numerous degraded or partial carbohydrate pathways and harbours bile salt hydrolase pseudogenes[53, 60]. In addition, *L. bulgaricus* shows a preference for growth in lactose, further emphasizing its niche adaptation to milk. The genome sequence of *L. helveticus*, a widely used cheese starter culture, has been reported recently[52]. Compared to the closely related *L. acidophilus*, *L. helveticus* has additional genes for fatty acid biosynthesis and specific amino acid metabolism, but notably fewer cell surface proteins and PEP-PTS systems for sugar utilization[52]. Additionally, no functional mucus binding proteins or transporters for complex carbohydrates such as raffinose and fructooligosaccharides are encoded by the *L. helveticus* genome, reflecting the degree of adaptation of *L. helveticus* to a milk environment. In contrast, *L. acidophilus* has adapted to the gut ecological niche by retaining the functional gene sets lacking in *L. helveticus*, emphasizing their importance for probiotic functionality and niche adaptation by autochthonous lactobacilli naturally residing in the GIT.

Several studies have examined commensal *Lactobacillus* gene expression in animal model systems. Using a stringent lincomycin-resistance based selection, Walter and colleagues identified surprisingly only three genes that were differentially expressed *in vivo* [69]. Bron *et al.* [70] used a modified *in vivo* expression technology to identify 72 genes expressed by *L. plantarum* in the mouse GIT, most of which were associated with carbon metabolism, amino acid metabolism, and stress resistance[70], and many of which were functions previously identified as survival/adaptation factors in pathogens. *L. casei* actively transcribes metabolic genes in the murine intestine, and initiates *de novo* protein synthesis[71]. *L. johnsonii* NCC533 expresses different sets of genes

246 depending on its location in the GIT[72], and surprisingly, 44% of the genome remains untranscribed
247 either *in vitro* or *in vivo* [72]. Interestingly, the prolonged murine gut persistence of NCC533 but not
248 of *L. johnsonii* was recently shown to induce expression of exopolysaccharide synthesis genes,
249 mannose uptake genes and a gene for a putative protease in this strain[73]. In summary, while there
250 are tantalizing glimpses of commensal *Lactobacillus* gene expression *in vivo*, these are as yet
251 limited to animal models; data from human volunteer studies is keenly awaited.

252 *Molecular basis of the interaction with other commensal bacteria.*

253 Although the biology of commensal bacteria can be investigated in isolation, it must ultimately be
254 understood in the context of the extremely complex intestinal ecosystem[61]. *Lactobacillaceae*
255 account for approximately 36 phylotypes among the >1000 phylotypes in the human
256 gastrointestinal microbiota[5]. In the short term, intervention studies in animal models and human
257 subjects provide the key insights into our current understanding of interaction with other
258 commensals.
259 Some lactobacilli may have quite subtle effects on the microbiota. Consumption of *L. rhamnosus*
260 DR20 transiently altered the levels of lactobacilli, bifidobacteria, enterococci, and *Bacteroidetes*,
261 but the variations were generally small[62] and mechanisms were not investigated. The development
262 of genomic tools facilitated a study[45], in germ-free mice that were mono-associated with *B.*
263 *thetaiotaomicron*, *B. longum*, *L. casei*, or combinations of these organisms[45]. Presence of *L. casei*
264 resulted in an expanded capacity of *B. thetaiotaomicron* to metabolize polysaccharides, and
265 increased expression of genes for inorganic ion transport and metabolism[45]. The *L. casei*-induced
266 changes in the *Bacteroides* transcriptome were functionally similar to those caused by *B. longum*,
267 but distinct from those induced by administration of *B. animalis* to the mice. Administration of *L.*
268 *paracasei* or *L. rhamnosus* to germ-free mice colonized with human infant microbiota caused
269 modest changes in levels of a limited number of species monitored by culture techniques, but major
270 changes to levels of diverse metabolites including amino acids, methylamines and short-chain fatty

acids[63]. The metabolism of the administered probiotics, coupled with competition for substrates and small molecules, are the likely reasons for the transcriptional and metabolite alterations described in these studies.

Numerous studies have reported that administration of probiotics benefits a range of gastrointestinal conditions and infections[64, 65], but mechanistic insights are generally lacking. Reduction in vaginal *Lactobacillus* levels that leads to vaginosis has been linked to production of a bacteriocin-like substance by commensal enterococci[66]. From the opposite perspective, the ability of *L. salivarius* to eliminate *Listeria monocytogenes* in a mouse model was dependent on production of the broad spectrum bacteriocin Abp118/salivaricin[67], and bacteriocin-producing lactobacilli become dominant among strains in a cocktail that reduce *Salmonella* shedding in pigs[68]. Thus bacteriocin production is likely an important general mechanism in the interaction of many lactobacilli and other commensals.

Comparative genomics of *Lactobacillus*.

Sequencing of the genomes of twenty lactic acid bacteria (LAB) has demonstrated that loss and decay of ancestral genes has played a key role in the evolution of *Lactobacillales*. *Lactobacillales* diverged from their *Bacillus* ancestor with an estimated loss of 600-1200 genes from a total gene repertoire of 2,100 to 2,200[50]. Many of these genes encoded biosynthetic enzymes or functioned in the sporulation process[50]. However, in addition to major gene losses, gene gains also occurred which appear to reflect the nutrient-rich niches occupied by the LAB, such as milk and the GIT. For example, genes encoding for peptidases, amino acid transport proteins and genes involved in the metabolism and transport of carbohydrates have been duplicated[50]. In addition, comparative analysis between GIT-associated species *L. acidophilus*, *L. gasseri*, and *L. johnsonii* and the dairy species *L. bulgaricus* and *L. helveticus* revealed selective pressure from niche-specific adaptation on the genome evolution of these species[51-53].

296 In addition to gene duplication, HGT is also evident in probiotic lactobacilli. For example, the
297 metabolic diversity of *L. plantarum* is underpinned by the expanded coding capacity afforded by its
298 larger 3 Mb genome, and a low-GC-content region coding for sugar transport and metabolism genes
299 which is likely to have been acquired by HGT[54]. Genes encoding cell surface factors in *L. johnsonii*
300 and the exopolysaccaride cluster in the *L. acidophilus* complex are further examples of HGT in
301 probiotic lactobacilli[52, 55]. Moreover, production of reuterin (3-hydroxypropionaldehyde), a potent
302 broad-spectrum antimicrobial compound[56], is encoded by a genomic island which is present in some
303 *L. reuteri* strains[57-59], and absent in the sequenced genome of a mouse *L. reuteri* isolate[58] and the
304 closely related *L. fermentum*[59].
305 With genomes of 12 of the 147 recognized species[74] now fully sequenced, *Lactobacillus* has been
306 targeted for several comparative whole-genome analyses. Beginning with the report of extreme
307 diversity between the first two available genomes[34], genome sequencing of *L. acidophilus*, *L.*
308 *gasseri*, *L. delbrueckii* and *L. helveticus* allowed a more focused attention on the 'acidophilus
309 complex'[25, 52, 75]. Large regions of synteny were observed between the species[25, 52]. Multi-locus
310 sequence analysis of five housekeeping genes, comparative-genome hybridizations and DNA-
311 typing showed consistent and stepwise-decreasing levels of similarity within the group, suggesting
312 a strong role for vertical evolution[25]. Conversely, differences between trees from 16S rRNA genes
313 and 401 core genes from *L. acidophilus*, *L. johnsonii* and *L. delbrueckii* indicated a much higher
314 level (40%) of HGT[75].
315 In order to infer robust phylogenetic relationships with minimal incongruence, or to elucidate
316 functional differences between species, a set of carefully selected single-copy ubiquitously-present
317 genes is necessary. A comparison of 354 core genes from five lactobacilli underscored the
318 substantial diversification of the genus, and suggested a subgeneric division into three groups [21].
319 Furthermore, two overlapping comparative studies, encompassing nine additional *Lactobacillales*
320 genomes, saw the expansion of the gene core to 567 order-specific genes[50, 76]. Similarly, the
321 majority of these encoded information-processing proteins. The finer granularity provided by

LaCOGs (*Lactobacillales*-specific COGs) allowed detection of two genes, whose gene-contexts suggest housekeeping and protein-modification functions. Recently, we extracted 141 core genes from 12 *Lactobacillus* genomes to investigate the case for a single congruent genus phylogeny[22]. Although this proved impossible at the time, four sub-generic groups were reliably distinguished. These were operationally characterized by absent genes rather than gained/retained genes, consistent with the findings of an earlier study[76].

Common evolutionary trends in probiotic genomes

Collective analyses of probiotic genome sequences so far available — the probiome — has revealed some generally conserved genetic traits[22, 25, 52, 54, 55, 59, 76], which may reflect adaptation to the intestinal niche [1]. However, since probiotic bacteria represent diverse and taxonomically heterogeneous groups of microorganisms, the analysis of phyletic (phylogenetic) patterns, i.e. patterns of gene presence/absence in a particular set of genomes, may be overwhelmingly influenced by the evolutionary distance between these two distant phyla. Nevertheless, common trends in the evolution of both *Bifidobacterium* and *Lactobacillus* genomes may be discerned. These include gene loss (e.g. of genes encoding biosynthetic enzymes), gene duplication and HGT. The adaptation of probiotic bacteria to successfully exist and compete in the human gut must have been driven by the occurrence of DNA duplications and genetic acquisitions during their evolution. Many genes involved in sugar metabolism and transport were duplicated or acquired early in the evolution of probiotic bacteria, including those encoding enolase, β-galactosidase, and many other GH[50]. In addition, expansion of peptidases and amino acid transporters has occurred in several lineages of *Lactobacillales* and bifidobacteria. Furthermore, several expanded families include proteins involved in antibiotic resistance in other bacteria, i.e. β-lactamases[77].

Horizontal gene transfer via bacteriophage-mediated or conjugative pathways has been extensively documented in *Lactobacillales* and appears to be important for niche-specific adaptation in probiotic bacteria. In probiotic lactobacilli, HGT played an important role in shaping the common

ancestor, in which 84 genes were inferred to be horizontally transferred from different sources[50]. In some cases, the ancestor acquired an additional pseudoparalogous copy of a gene by HGT (e.g. enolase in *Lactobacillales*) while on other occasions, xenologous displacement, i.e., acquisition of genes by HGT followed by the loss of the ancestral orthologous gene[78] apparently took place.

A provocative future challenge will involve the identification of the hypothetical core probiogenome, representing core genome functions of probiotic bacteria. However, only seven genes present in the bifidobacteria but not in the genomes of the other members of the *Actinobacteria* phylum are shared with *Lactobacillales*. Only one of these genes, which encodes a functionally uncharacterized membrane protein, is present in all the *Lactobacillales* genomes so far sequenced [50].

Conclusions and future considerations

Most of the probiotic bacteria marketed today were originally selected on the basis of technological stability or by a variety of easily measurable phenotypes such as ability to tolerate bile salts or survive GIT passage, but not necessarily for their ability to promote health benefits. It is crucial to identify the precise mechanisms by which such probiotic microorganisms influence human health. Such studies should be accelerated by omics approaches involving genomics and functional analyses. Molecular interaction models are being currently developed, although more are required, that monitor the activation of cellular and systemic responses *in vivo* in animal models and in feeding trial participants through the measurement of previously validated biomarkers. The combination of verified molecular models with functional and comparative genomics-based approaches should enable selection of the most appropriate probiotic strain for a particular health benefit or improvement of strain processing and administration regimes that optimize the established health effect. Finally, this might allow the selection of specific probiotics for a particular human genotype, in analogy to personalized genomic medicine efforts.

Several issues regarding the sequences of complete probiotic bacterial genomes remain unresolved at present. So far, only a limited number of completed probiotic bacterial genome sequences are available, which only partially represent the total biodiversity of probiotic bacteria residing in the human gut. In this context, understanding of the human gut microbiome will be an important challenge for the future[79]. Furthermore, sequencing the genomes of environmental organisms and carrying out metagenomic surveys of diverse gut environments (human vs. animal GIT) will provide not only an improved understanding of microbial biodiversity but also insights into the evolution of bacterial factors that may be crucial for the commensals (probiotics) establishment in these different gut niches[80].

The first decade of bacterial genomics has afforded unprecedented insights into the evolution of bacterial pathogens (bacterial pathogenomics)[81]. The next decade holds the promise of being even more rewarding as the new discoveries about probiotic bacteria provided by probiogenomic efforts are exploited.

GLOSSARY

Omics: The integration of genomics methodology and data with functional genomic analyses involving transcriptomics, proteomics, metabolomics and interactomics.

Microbiota: The collective microbial community or population resident in a particular locale at a given time-point.

Microbiome: The collective genome of the human microbial communities

Prebiotics: Growth substrates that are preferentially (or ideally, exclusively) metabolized by a single genus or species, and that may thus be used as dietary supplements to promote growth of a targeted microorganism.

Transcriptome: Subsets of genes transcribed in an organism. It represents dynamic links between genomes, proteins and cellular phenotypes.

Synteny: Genetic linkage or conservation of gene order.

399 COGs: Clusters of Orthologous Groups are delineated by comparing protein sequences encoded in
400 complete genomes, representing major phylogenetic lineages. Each COG consists of individual
401 proteins or groups of paralogs from at least 3 lineages and thus corresponds to an ancient conserved
402 domain.
403 Neighbour-joining tree: Tree that reconstruct the evolutionary development of organisms based on
404 distances between each pair of taxa.

405 **Acknowledgements**

406 Work in the laboratories of DvS and PWOT is supported by an SFI CSET award to the Alimentary
407 Pharmabiotic Centre, and a DAF/HRB FHRI award to the ELDERMET project. MV was supported
408 by an Italian Award for Outstanding Young Researcher scheme “Incentivazione alla mobilità di
409 studiosi stranieri e italiani residente all’estero” 2005-2009, a Marie Curie Reintegration Grant
410 (MERG-CT-2005-03080) and Parmalat spa, Italy. We thank also S. Leahy, Univ. College Cork and
411 D. John, Trinity College Dublin for the scanning electron microscope images.

References

1. Backhed, F., Ley, R. E., Sonnenburg, J. L., Peterson, D. A. & Gordon, J. I. Host-bacterial mutualism in the human intestine. Science 307, 1915-20 (2005).
2. **Eckburg, P. B. et al. Diversity of the human intestinal microbial flora. Science 308, 1635-8 (2005).**

This article describes the bacterial diversity occurring in the human gut using 16S rRNA gene based libraries.

3. Seksik, P. et al. Alterations of the dominant faecal bacterial groups in patients with Crohn's disease of the colon. Gut 52, 237-42 (2003).
4. Turroni, F., Ribbera, A., Foroni, E., van Sinderen, D. & Ventura, M. Human gut microbiota and bifidobacteria: from composition to functionality. Antonie Van Leeuwenhoek 94, 35-50 (2008).
5. **Rajilic-Stojanovic, M., Smidt, H. & de Vos, W. M. Diversity of the human gastrointestinal tract microbiota revisited. Environ. Microbiol. 9, 2125-36 (2007).**

This review provides an integrated summary of data from culture independent studies of the human gut microbiota.

6. Ley, R. E., Peterson, D. A. & Gordon, J. I. Ecological and evolutionary forces shaping microbial diversity in the human intestine. Cell 124, 837-48 (2006).
7. Guarner, F. & Malagelada, J. R. Gut flora in health and disease. Lancet 361, 512-9 (2003).
8. Hooper, L. V. & Gordon, J. I. Commensal host-bacterial relationships in the gut. Science 292, 1115-8 (2001).
9. Backhed, F. et al. The gut microbiota as an environmental factor that regulates fat storage. Proc Natl Acad Sci U S A 101, 15718-23 (2004).
10. Samuel, B. S. & Gordon, J. I. A humanized gnotobiotic mouse model of host-archaeal-bacterial mutualism. Proc Natl Acad Sci U S A 103, 10011-6 (2006).
11. Turnbaugh, P. J. et al. An obesity-associated gut microbiome with increased capacity for energy harvest. Nature 444, 1027-31 (2006).
12. Frank, D. N. et al. Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases. Proc Natl Acad Sci U S A 104, 13780-5 (2007).
13. Kassinen, A. et al. The fecal microbiota of irritable bowel syndrome patients differs significantly from that of healthy subjects. Gastroenterology 133, 24-33 (2007).
14. **Manichanh, C. et al. Reduced diversity of faecal microbiota in Crohn's disease revealed by a metagenomic approach. Gut 55, 205-11 (2006)**.

The preceding two references provide evidence for significant microbiota alterations in functional bowel disorders.

15. Salminen, S., Nurmi, J. & Gueimonde, M. The genomics of probiotic intestinal microorganisms. Genome Biol 6, 225 (2005).
16. Marco, M. L., Pavan, S. & Kleerebezem, M. Towards understanding molecular modes of probiotic action. Curr Opin Biotechnol 17, 204-10 (2006).
17. O'Hara, A. M. & Shanahan, F. Mechanisms of action of probiotics in intestinal diseases. ScientificWorldJournal 7, 31-46 (2007).
18. Saxelin, M., Tynkkynen, S., Mattila-Sandholm, T. & de Vos, W. M. Probiotic and other functional microbes: from markets to mechanisms. Curr Opin Biotechnol 16, 204-11 (2005).
19. Ventura, M. et al. Genomics of Actinobacteria: tracing the evolutionary history of an ancient phylum. Microbiol Mol Biol Rev 71, 495-548 (2007).
20. Joyce, A. R. & Palsson, B. O. The model organism as a system: integrating 'omics' data sets. Nat Rev Mol Cell Biol 7, 198-210 (2006).

21. Ventura, M. et al. Analysis of bifidobacterial evolution using a multilocus approach. Int J Syst Evol Microbiol 56, 2783-92 (2006).
22. Claesson, M. J. et al. Multireplicon genome architecture of Lactobacillus salivarius. Proc Natl Acad Sci U S A 103, 6718-23 (2006).
23. Tissier, H. Traitement des infections intestinales par la methode de la flore bacterienne de l'intestin. Crit. Rev. Soc. Biol. 60, 359-361 (1906).
24. Ventura, M., Canchaya, C., Fitzgerald, G. F., Gupta, R. S. & van Sinderen, D. Genomics as a means to understand bacterial phylogeny and ecological adaptation: the case of bifidobacteria. Antonie Van Leeuwenhoek 91, 351-72 (2007).
25. Schell, M. A. et al. The genome sequence of Bifidobacterium longum reflects its adaptation to the human gastrointestinal tract. Proc Natl Acad Sci U S A 99, 14422-7 (2002).
26. Gibson, G. R. & Roberfroid, M. B. Dietary modulation of the human colonic microbiota: introducing the concept of prebiotics. J Nutr 125, 1401-12 (1995).
27. Flint, H. J., Bayer, E. A., Rincon, M. T., Lamed, R. & White, B. A. Polysaccharide utilization by gut bacteria: potential for new insights from genomic analysis. Nat Rev Microbiol 6, 121-31 (2008).
28. Sonnenburg, J. L. et al. Glycan foraging in vivo by an intestine-adapted bacterial symbiont. Science 307, 1955-9 (2005).
29. Hooper, L. V., Xu, J., Falk, P. G., Midtvedt, T. & Gordon, J. I. 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, 9833-8 (1999).
30. Hinz, S. W., Verhoef, R., Schols, H. A., Vincken, J. P. & Voragen, A. G. Type I arabinogalactan contains beta-D-Galp-(1-->3)-beta-D-Galp structural elements. Carbohydr Res 340, 2135-43 (2005).
31. Ryan, S. M., Fitzgerald, G. F. & van Sinderen, D. Screening for and identification of starch-, amylopectin-, and pullulan-degrading activities in bifidobacterial strains. Appl Environ Microbiol 72, 5289-96 (2006).
32. Maze, A., O'Connell-Motherway, M., Fitzgerald, G. F., Deutscher, J. & van Sinderen, D. Identification and characterization of a fructose phosphotransferase system in Bifidobacterium breve UCC2003. Appl Environ Microbiol 73, 545-53 (2007).
33. van den Broek, L. A., Hinz, S. W., Beldman, G., Vincken, J. P. & Voragen, A. G. Bifidobacterium carbohydrases-their role in breakdown and synthesis of (potential) prebiotics. Mol Nutr Food Res 52, 146-63 (2008).
This paper is providing the state of the art in the area of enzymes encoded by bifidobacteria involved in the hydrolysis of carbohydrates
34. Siezen, R. et al. Lactobacillus plantarum gene clusters encoding putative cell-surface protein complexes for carbohydrate utilization are conserved in specific gram-positive bacteria. BMC Genomics 7, 126 (2006).
35. Hooper, L. V., Midtvedt, T. & Gordon, J. I. How host-microbial interactions shape the nutrient environment of the mammalian intestine. Annu Rev Nutr 22, 283-307 (2002).
36. Hoskins, L. C. et al. Mucin degradation in human colon ecosystems. Isolation and properties of fecal strains that degrade ABH blood group antigens and oligosaccharides from mucin glycoproteins. J Clin Invest 75, 944-53 (1985).
37. Ruas-Madiedo, P., Gueimonde, M., Fernandez-Garcia, M., de los Reyes-Gavilan, C. G. & Margolles, A. Mucin degradation by Bifidobacterium strains isolated from the human intestinal microbiota. Appl Environ Microbiol 74, 1936-40 (2008).
38. Ehrmann, M. A., Korakli, M. & Vogel, R. F. Identification of the gene for beta-fructofuranosidase of Bifidobacterium lactis DSM10140(T) and characterization of the enzyme expressed in Escherichia coli. Curr Microbiol 46, 391-7 (2003).

39. Katayama, T. et al. Molecular cloning and characterization of Bifidobacterium bifidum 1,2-alpha-L-fucosidase (AfcA), a novel inverting glycosidase (glycoside hydrolase family 95). J Bacteriol 186, 4885-93 (2004).
40. Ryan, S. M., Fitzgerald, G. F. & van Sinderen, D. Transcriptional regulation and characterization of a novel beta-fructofuranosidase-encoding gene from Bifidobacterium breve UCC2003. Appl Environ Microbiol 71, 3475-82 (2005).
41. Gonzalez, R., Klaassens, E. S., Malinen, E., de Vos, W. M. & Vaughan, E. E. Differential transcriptional response of Bifidobacterium longum to human milk, formula milk and galactooligosaccharide. Appl Environ Microbiol (2008).
42. Liepke, C. et al. Human milk provides peptides highly stimulating the growth of bifidobacteria. Eur J Biochem 269, 712-8 (2002).
43. Ivanov, D. et al. A serpin from the gut bacterium Bifidobacterium longum inhibits eukaryotic elastase-like serine proteases. J Biol Chem 281, 17246-52 (2006).
44. Potempa, J., Korzus, E. & Travis, J. The serpin superfamily of proteinase inhibitors: structure, function, and regulation. J Biol Chem 269, 15957-60 (1994).
45. Sonnenburg, J. L., Chen, C. T. & Gordon, J. I. Genomic and metabolic studies of the impact of probiotics on a model gut symbiont and host. PLoS Biol 4, e413 (2006).
This paper describes the crosstalk existing between bifidobacteria and *Bacteroides* in the murine intestine as well as between these bacteria and their host.
46. Kato, S., Haruta, S., Cui, Z. J., Ishii, M. & Igarashi, Y. Stable coexistence of five bacterial strains as a cellulose-degrading community. Appl Environ Microbiol 71, 7099-106 (2005).
47. Barrangou, R. et al. CRISPR provides acquired resistance against viruses in prokaryotes. Science 315, 1709-12 (2007).
48. Klijn, A., Mercenier, A. & Arigoni, F. Lessons from the genomes of bifidobacteria. FEMS Microbiol Rev 29, 491-509 (2005).
49. Aas, J. A. et al. Bacteria of dental caries in primary and permanent teeth in children and young adults. J Clin Microbiol 46, 1407-17 (2008).
50. Makarova, K. S. & Koonin, E. V. Evolutionary genomics of lactic acid bacteria. J Bacteriol 189, 1199-208 (2007).
51. Makarova, K. et al. Comparative genomics of the lactic acid bacteria. Proc. Natl. Acad. Sci. U S A 103, 15611-6 (2006).
This landmark study provided a large tranche of genomic data to allow studies of genome evolution in lactic acid bacteria.
52. Altermann, E. et al. Complete genome sequence of the probiotic lactic acid bacterium Lactobacillus acidophilus NCFM. Proc Natl Acad Sci U S A 102, 3906-12 (2005).
53. van de Guchte, M. et al. The complete genome sequence of Lactobacillus bulgaricus reveals extensive and ongoing reductive evolution. Proc. Natl. Acad. Sci. U S A 103, 9274-9 (2006).
54. Kleerebezem, M. et al. Complete genome sequence of Lactobacillus plantarum WCFS1. Proc Natl Acad Sci U S A 100, 1990-5 (2003).
This is the first article describing the genome sequence of a member of the genus *Lactobacillus*.
55. Pridmore, R. D. et al. The genome sequence of the probiotic intestinal bacterium Lactobacillus johnsonii NCC 533. Proc Natl Acad Sci U S A 101, 2512-7 (2004).
This is paper describes the genome contents of a common used probiotic bacterium belonging to the genus *Lactobacillus*.
56. Talarico, T. L., Casas, I. A., Chung, T. C. & Dobrogosz, W. J. Production and isolation of reuterin, a growth inhibitor produced by *Lactobacillus reuteri*. Antimicrob. Agents Chemother. 32, 1854-8 (1988).
57. Santos, F. et al. The complete coenzyme B12 biosynthesis gene cluster of *Lactobacillus reuteri* CRL1098. Microbiology 154, 81-93 (2008).

58. Sriramulu, D. D. et al. Lactobacillus reuteri DSM 20016 produces cobalamin-dependent diol dehydratase in metabolosomes and metabolizes 1,2-propanediol by disproportionation. J. Bacteriol. 190, 4559-67 (2008).
59. Morita, H. et al. Comparative Genome Analysis of Lactobacillus reuteri and Lactobacillus fermentum Reveal a Genomic Island for Reuterin and Cobalamin Production. DNA Res 15, 151-61 (2008).
60. Pfeiler, E. A. & Klaenhammer, T. R. The genomics of lactic acid bacteria. Trends Microbiol 15, 546-53 (2007).
61. Whitman, W. B., Coleman, D. C. & Wiebe, W. J. Prokaryotes: the unseen majority. Proc. Natl. Acad. Sci. U S A 95, 6578-83 (1998).
62. Tannock, G. W. et al. Analysis of the fecal microflora of human subjects consuming a probiotic product containing *Lactobacillus rhamnosus* DR20. Appl. Environ. Microbiol. 66, 2578-88 (2000).
63. Martin, F. P. et al. Probiotic modulation of symbiotic gut microbial-host metabolic interactions in a humanized microbiome mouse model. Mol. Syst. Biol. 4, 157 (2008).
64. Hickson, M. et al. Use of probiotic *Lactobacillus* preparation to prevent diarrhoea associated with antibiotics: randomised double blind placebo controlled trial. Brit. Med. J. 335, 80 (2007).
65. Sullivan, A. & Nord, C. E. Probiotics and gastrointestinal diseases. J. Intern. Med. 257, 78-92 (2005).
66. Kelly, M. C., Mequio, M. J. & Pybus, V. Inhibition of vaginal lactobacilli by a bacteriocin-like inhibitor produced by Enterococcus faecium 62-6: potential significance for bacterial vaginosis. Infect. Dis. Obstet. Gynecol. 11, 147-56 (2003).
67. Corr, S. C. et al. Bacteriocin production as a mechanism for the antiinfective activity of *Lactobacillus salivarius* UCC118. Proc. Natl. Acad. Sci. U S A 104, 7617-7621 (2007).
This study identified the first molecular mechanism whereby probiotic bacteria modulate the microbiota *in vivo*.
68. Casey, P. G. et al. A five-strain probiotic combination reduces pathogen shedding and alleviates disease signs in pigs challenged with *Salmonella enterica* Serovar Typhimurium. Appl. Environ. Microbiol. 73, 1858-63 (2007).
69. Walter, J. et al. Identification of *Lactobacillus reuteri* genes specifically induced in the mouse gastrointestinal tract. Appl. Environ. Microbiol. 69, 2044-51 (2003).
70. Bron, P. A., Grangette, C., Mercenier, A., de Vos, W. M. & Kleerebezem, M. Identification of Lactobacillus plantarum genes that are induced in the gastrointestinal tract of mice. J Bacteriol 186, 5721-9 (2004).
This manuscript provides an insight into the molecular interactions between a commensal microorganism and a murine-model host.
71. Oozeer, R. et al. Differential activities of four *Lactobacillus casei* promoters during bacterial transit through the gastrointestinal tracts of human-microbiota-associated mice. Appl. Environ. Microbiol. 71, 1356-63 (2005).
72. Denou, E. et al. Gene expression of commensal *Lactobacillus johnsonii* strain NCC533 during in vitro growth and in the murine gut. J. Bacteriol. 189, 8109-19 (2007).
73. Denou, E. et al. Identification of genes associated with the long-gut-persistence phenotype of the probiotic *Lactobacillus johnsonii* strain NCC533 using a combination of genomics and transcriptome analysis. J. Bacteriol. 190, 3161-8 (2008).
74. Euzèby, J. P. (2007). List of Bacterial names with Standing in Nomenclature. Int J Syst Bacteriol. 47, 590-592 (1997).
75. Nicolas, P., Bessieres, P., Ehrlich, S. D., Maguin, E. & van de Guchte, M. Extensive horizontal transfer of core genome genes between two *Lactobacillus* species found in the gastrointestinal tract. BMC Evol. Biol. 7, 141 (2007).

76. Makarova, K. et al. Comparative genomics of the lactic acid bacteria. Proc Natl Acad Sci U S A 103, 15611-6 (2006).
77. Teuber, M., Meile, L. & Schwarz, F. Acquired antibiotic resistance in lactic acid bacteria from food. Antonie Van Leeuwenhoek 76, 115-37 (1999).
78. Koonin, E. V., Makarova, K. S. & Aravind, L. Horizontal gene transfer in prokaryotes: quantification and classification. Annu Rev Microbiol 55, 709-42 (2001).
79. Turnbaugh, P. J. et al. The human microbiome project. Nature 449, 804-10 (2007).
80. **Ley, R. E. et al. Evolution of mammals and their gut microbes. Science 320, 1647-51 (2008).**
This paper describes the bacterial diversity present in the gut of numerous mammals.
81. Pallen M.J., & Wren, B.M. Bacterial pathogenomics. Nature 449, 835-842 (2007).
82. Lee, J. H. et al. Comparative genomic analysis of the gut bacterium Bifidobacterium longum reveals loci susceptible to deletion during pure culture growth. BMC Genomics 9, 247 (2008).
83. Leahy, S. C., Higgins, D. G., Fitzgerald, G. F. & van Sinderen, D. Getting better with bifidobacteria. J Appl Microbiol 98, 1303-15 (2005).

Table 1: General features of sequenced *Bifidobacterium* and *Lactobacillus* genomes.

Species	Genome size (bp)	%GC	Gene numbers	Proteins	Source	Accession number	Reference
B. longum subsp. *longum* NCC2705	2,256,640	60%	1798	1727	Human GIT	NC_004307	[25]
B. longum subsp. *longum* DJ010A	2,375,286	59%	1908	1908	Human GIT	NC_010816	[82]
B. breve UCC2003	2422668	59%	1868		Infant feces	Project ID: 13487	[83]
B. adolescentis ATCC15703	2,089,645	59%	1701	1631	Human GIT	NC_008618	-
B. adolescentis L2-32	2,385,710	59%	2499	2428	Infant feces	NZ_AAXD00000000	-
B. animalis subsp. l*actis* HN019	1,915,892	60%	1632	1578	-	NZ_ABOT00000000	-
Lactobacillus acidophilus NCFM	1,993,560	34%	1936	1862	Human GIT	NC_006814	[52]
Lactobacillus casei ATCC334	2,895,264	46%	2909	2751	Emmental cheese	NC_008526	[76]
Lactobacillus gasseri ATCC33323	1,894,360	35%	1898	1755	Human GIT	NC_008530	[50]
Lactobacillus jonsonii NCC533	1,992,676	34%	1918	1821	Human GIT	NC_005362	[55]
Lactobacillus plantarum WCFS1	3,308,274	44%	3135	3007	Human saliva	NC_004567	[54]
Lactobacillus reuteri F275	1,999,618	38%	2027	1900	Human GIT	NC_009513	[60]
Lactobacillus fermentum IFO 3956	2,098,685	51%	1912	1843	-	NC_010610	[60]
Lactobacillus salivarius susp. *salivarius* UCC118	1,827,111	32%	1864	1717	Human GIT	NC_007929	[22]

LEGENDS

Figure 1: Ecological, evolutionary and morphological overview of bifidobacteria and lactobacillae. [A| Schematic representation of the biological relationships between bacteria and the human body. Commensalisms or symbiosis is a consequence of the co-evolution of host-bacterial relationships. B| Evolutionary relationships between the main GIT commensal bacterial groups (bifidobacteria on the left and lactobacillae on the right) based on neighbour-joining tree of 16S rRNA genes sequences. Bar indicates scale for computed distances. Bacterial taxa for which the whole genome sequences is available are shaded in blue, whereas for those that is still on progress are shaded in grey. C| electron micrographs illustrating the cell morphology of bifidobacteria (e.g., *B. breve* UCC2003) (right panel) and lactobacillae (e.g., *L. salivarius* UCC118) (Left panel). Both scanning electron microscope images were prepared by. S. Leahy, Univ. College Cork and D. John, Trinity College Dublin. Magnification ca. 20,000 fold; scale bar is 2 micrometres.

Figure 2: Putative strategy adopted by bifidobacteria to secure sugar nutrients for their own benefit. Bifidobacteria use a kind of docking station to capture complex sugars (e.g., xylan and arabino based molecules) and bind these to the bacterial cell surface, without loosing them to nearby competitors. In the latter case the docking station is a complex of modular glycanases, which are anchored at the cell surface by a transmembrane domain. The enzymatic activities degrade the arabinoxylan molecules to oligosaccharides that are subsequently transported across the bacterial membrane by a transporter protein; the presence of the bacterial cell wall may prohibit diffusion of these nutrients away from the transporter.

Figure 3. Comparative analysis of *Bifidobacterium* genomes. A| A comparison of the *B. dentium* Bd1 and *B. adolescentis* ATCC15703 genomes. B| Comparison of gene order conservation between two genome pairs, illustrating different forms of bifidobacterial genome evolution. X and Y axes represent the linearised chromosomes of *B. dentium* Bd1 and *B. adolescentis* ATCC15703, respectively.

Figure 4. Comparative analysis of *Lactobacillus* genomes. Circular genome atlas of *L. plantarum* WCFS1 with mapped orthologs (defined as reciprocal best FastA hits with more than 30% identity over at least 80% of both protein lengths) in 13 publicly available *Lactobacillus* genomes. The outer circle shows *L. plantarum* followed, inwards, by *L. salivarius*, *L. brevis*, *L. reuteri* F275, *L. reuteri* F275 (Japanese), *L. fermentum*, *L. acidophilus*, *L. helveticus*, *L. johnsonii*, *L. gasseri*, *L. bulgaricus* ATCC 11842, *L. bulgaricus* ATCC BAA-365, *L. casei*, *L. sakei*, G+C percentage, and GC skew (window-sizes 10,000 bp). Red colour represents COG categories in Metabolism, green - Information Storage and Processing, blue - Cellular Processes and Signalling, and grey - poorly or not categorised.

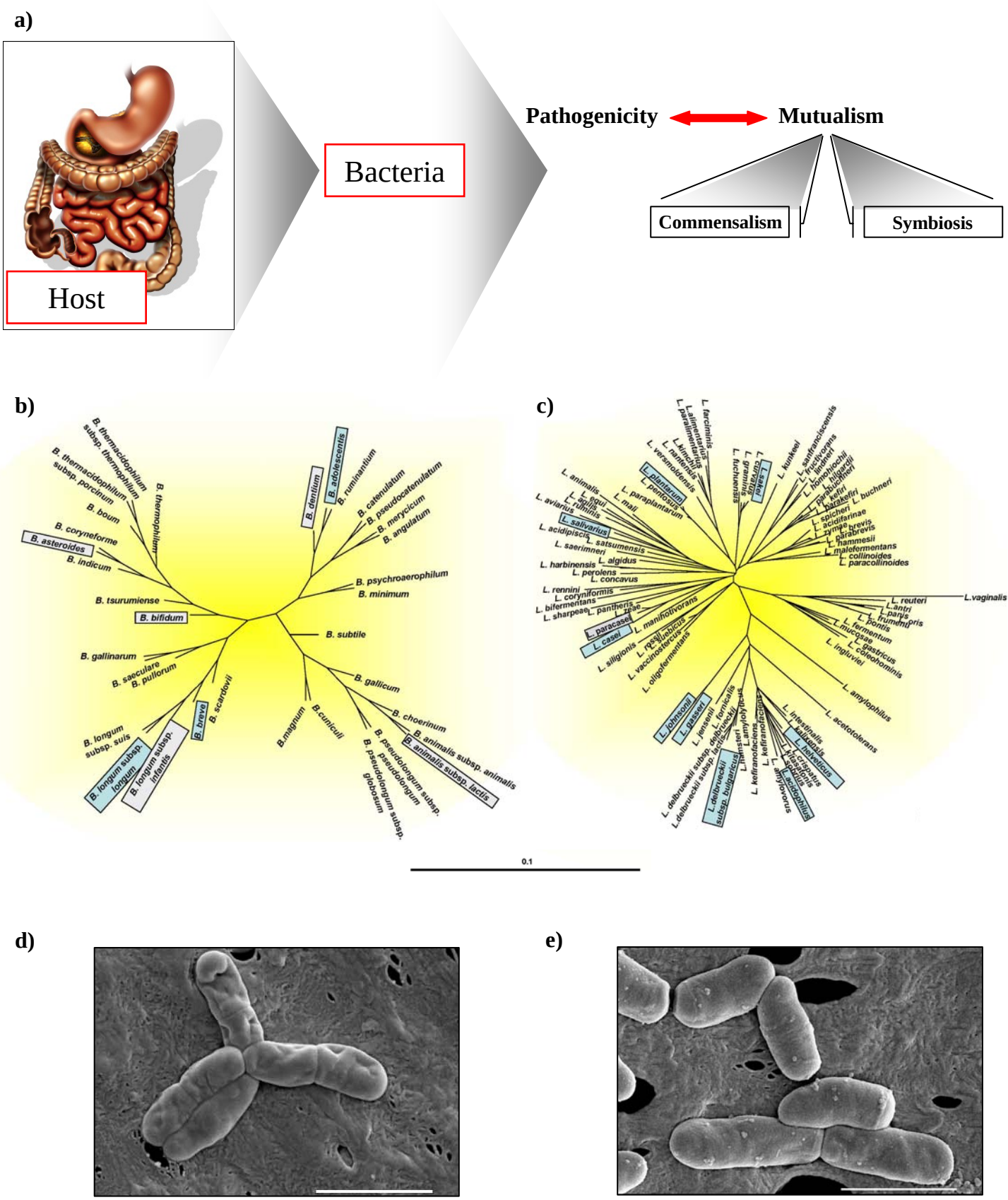

a)
Host
Bacteria
Pathogenicity
Mutualism
Commensalism
Symbiosis
b)
B. thermacidophilum subsp. thermophilum
B. thermacidophilum subsp. porcinum
B. thermophilum
B. boum
B. coryneforme
B. asteroides
B. indicum
B. tsurumiense
B. bifidum
B. gallinarum
B. saeculare
B. pullorum
B. longum subsp. suis
B. longum subsp. longum
B. longum subsp. infantis
B. breve
B. scardovii
B. dentium
B. adolescentis
B. ruminantium
B. catenulatum
B. pseudocatenulatum
B. merycicum
B. angulatum
B. psychroaerophilum
B. minimum
B. subtile
B. gallicum
B. choerinum
B. animalis subsp. animalis
B. animalis subsp. lactis
B. pseudolongum subsp. pseudolongum
B. pseudolongum subsp. globosum
B. magnum
B. cuniculi
0.1
c)
L. farciminis
L. alimentarius
L. paralimentarius
L. kimchii
L. nantensis
L. versmoldensis
L. plantarum
L. pentosus
L. paraplantarum
L. mali
L. animalis
L. equi
L. agilis
L. ruminis
L. aviarius
L. salivarius
L. acidipiscis
L. satsumensis
L. saerimneri
L. harbinensis
L. algidus
L. perolens
L. concavus
L. rennini
L. coryniformis
L. bifermentans
L. sharpeae
L. pantheris
L. paracasei
L. casei
L. manihotivorans
L. siliginis
L. rossii
L. vaccinostercus
L. oligofermentans
L. johnsonii
L. gasseri
L. jensenii
L. fornicalis
L. delbrueckii subsp. delbrueckii
L. delbrueckii subsp. lactis
L. delbrueckii subsp. bulgaricus
L. amylolyticus
L. hamsteri
L. kefiranofaciens
L. kefiranofaciens
L. amylovorus
L. acidophilus
L. crispatus
L. kitasatonis
L. helveticus
L. intestinalis
L. kalixensis
L. acetotolerans
L. amylophilus
L. ingluviei
L. coleohominis
L. gastricus
L. mucosae
L. fermentum
L. pontis
L. frumenti
L. panis
L. antri
L. reuteri
L. vaginalis
L. paracollinoides
L. collinoides
L. malefermentans
L. hammesii
L. parabrevis
L. zymae
L. acidifarinae
L. spicheri
L. kefiri
L. parakefiri
L. buchneri
L. hilgardii
L. homohiochii
L. lindneri
L. fructivorans
L. sanfranciscensis
L. kunkeei
L. curvatus
L. graminis
L. fuchuensis
L. sakei
0.1
d)
e)

Figure 1

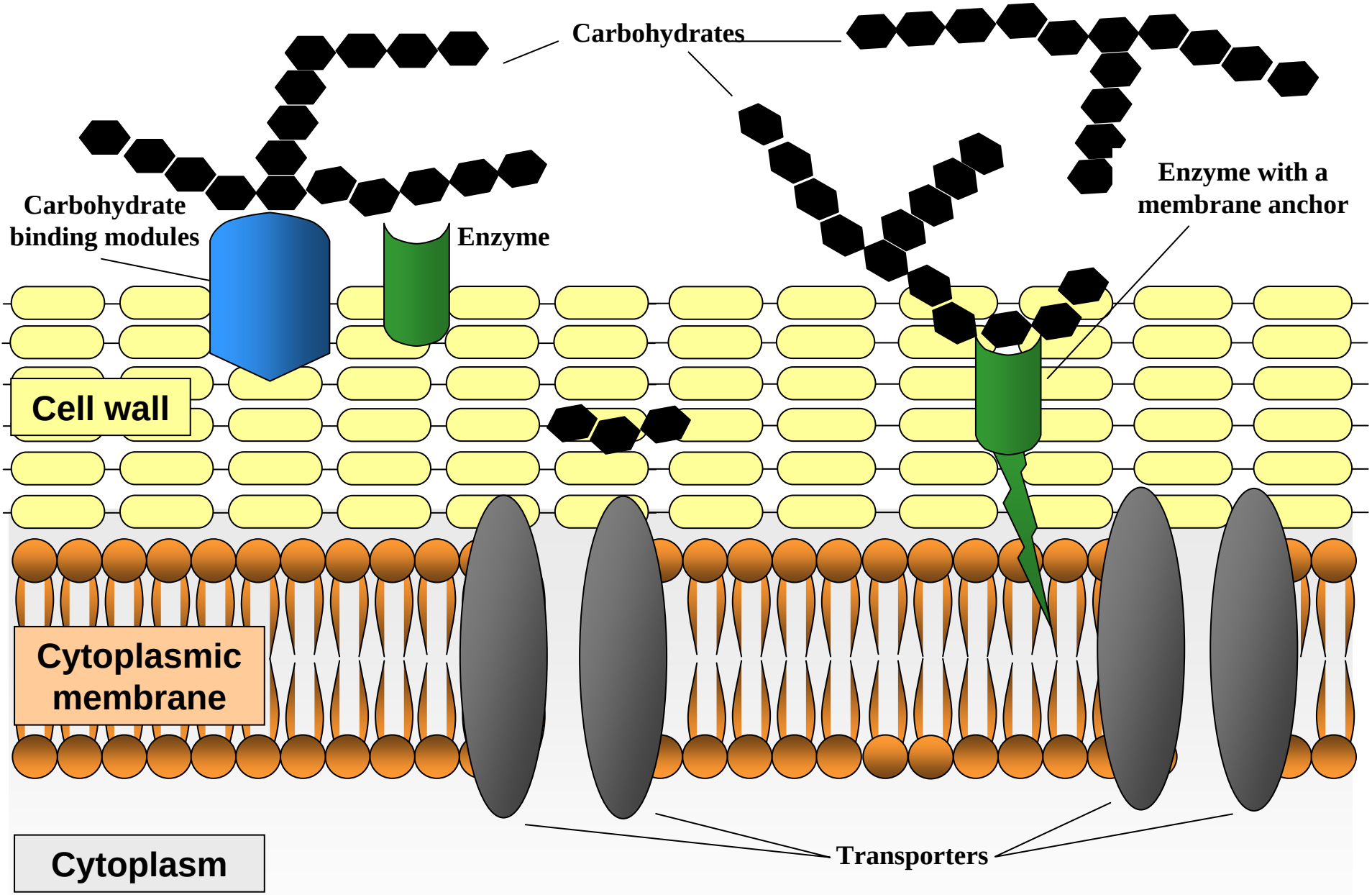
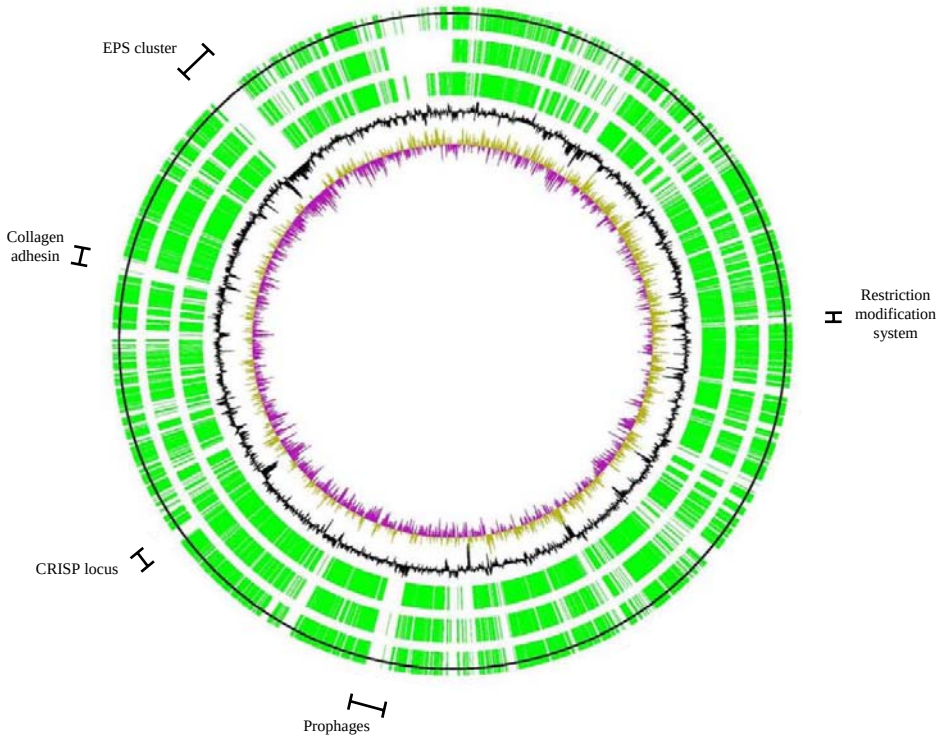

Carbohydrates
Carbohydrate binding modules
Enzyme
Enzyme with a membrane anchor
Cell wall
Cytoplasmic membrane
Cytoplasm
Transporters

Figure 2

a)


EPS cluster
Collagen adhesin
CRISP locus
Prophages
Restriction modification system

b)

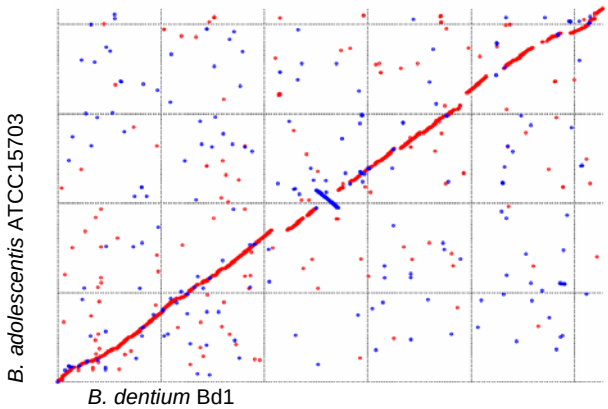

B. adolescentis ATCC15703
B. dentium Bd1

Figure 3

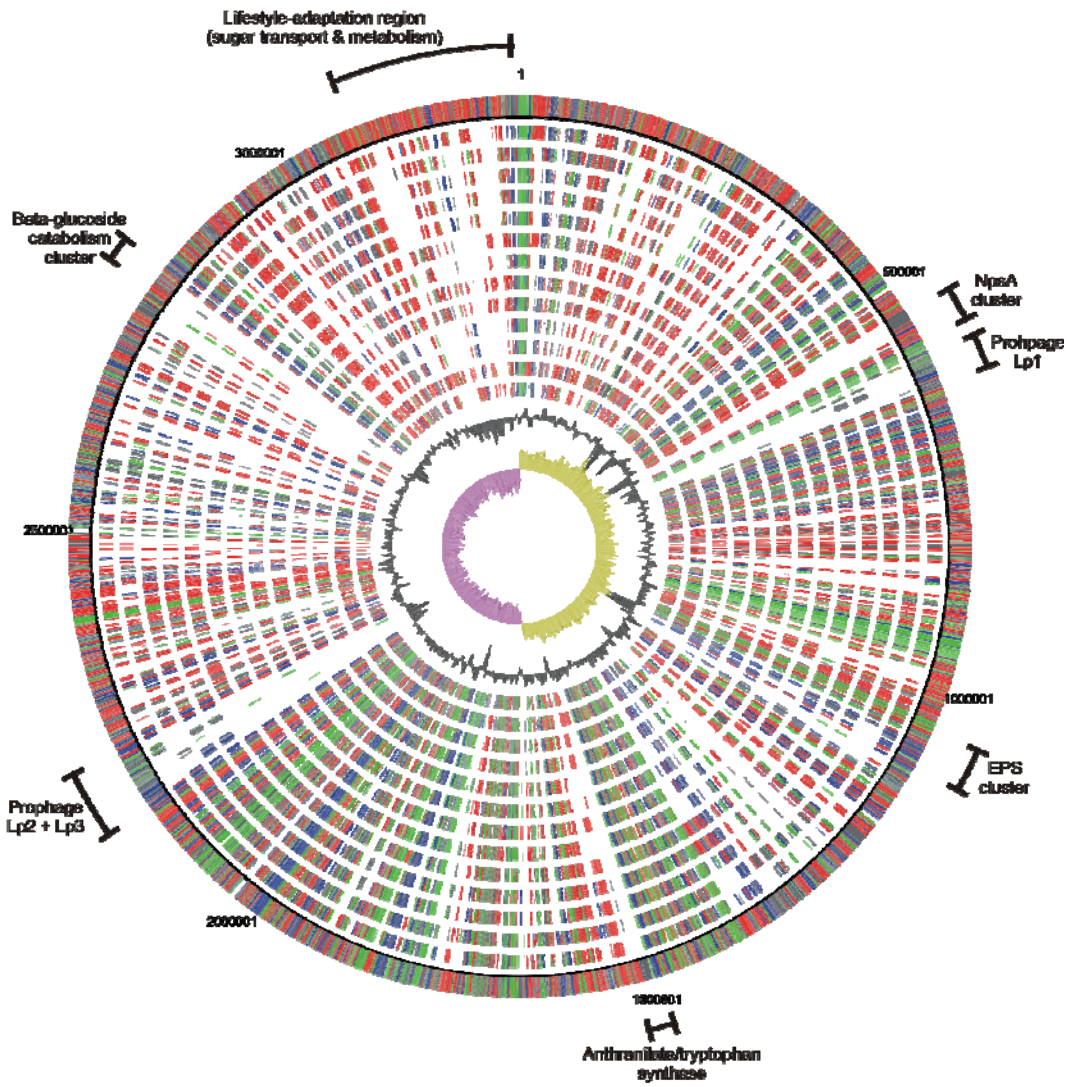

Lifestyle-adaptation region
(sugar transport & metabolism)
1
500001
1000001
1500001
2000001
2500001
3000001
Beta-glucoside
catabolism
cluster
NpsA
cluster
Prohpage
Lp1
EPS
cluster
Anthranilate/tryptophan
synthase
Prophage
Lp2 + Lp3

Figure 4

Table 1: General features of sequenced *Bifidobacterium* and *Lactobacillus* genomes.

Species	Genome size (bp)	%GC	Gene numbers	Proteins	Source	Accession number	Reference
B. longum subsp. *longum* NCC2705	2,256,640	60%	1798	1727	Human GIT	NC_004307	[25]
B. longum subsp. *longum* DJ010A	2,375,286	59%	1908	1908	Human GIT	NC_010816	[82]
B. breve UCC2003	2422668	59%	1868		Infant feces	Project ID: 13487	[83]
B. adolescentis ATCC15703	2,089,645	59%	1701	1631	Human GIT	NC_008618	-
B. adolescentis L2-32	2,385,710	59%	2499	2428	Infant feces	NZ_AAXD00000000	-
B. animalis subsp. l*actis* HN019	1,915,892	60%	1632	1578	-	NZ_ABOT00000000	-
Lactobacillus acidophilus NCFM	1,993,560	34%	1936	1862	Human GIT	NC_006814	[52]
Lactobacillus casei ATCC334	2,895,264	46%	2909	2751	Emmental cheese	NC_008526	[76]
Lactobacillus gasseri ATCC33323	1,894,360	35%	1898	1755	Human GIT	NC_008530	[50]
Lactobacillus jonsonii NCC533	1,992,676	34%	1918	1821	Human GIT	NC_005362	[55]
Lactobacillus plantarum WCFS1	3,308,274	44%	3135	3007	Human saliva	NC_004567	[54]
Lactobacillus reuteri F275	1,999,618	38%	2027	1900	Human GIT	NC_009513	[60]
Lactobacillus fermentum IFO 3956	2,098,685	51%	1912	1843	-	NC_010610	[60]
Lactobacillus salivarius susp. *salivarius* UCC118	1,827,111	32%	1864	1717	Human GIT	NC_007929	[22]