Genome-scale analyses of health-promoting bacteria: probiogenomics

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

Abstract | The human body is colonized by an enormous population of bacteria (microbiota) that provides the host with coding capacity and metabolic activities. Among the human gut microbiota are health-promoting indigenous species (probiotic bacteria) that are commonly consumed as live dietary supplements. Recent genomics-based studies (probiogenomics) are starting to provide insights into how probiotic bacteria sense and adapt to the gastrointestinal tract environment. In this Review, we discuss the application of probiogenomics in the elucidation of the molecular basis of probiosis using the well-recognized model probiotic bacteria genera *Bifidobacterium* and *Lactobacillus* as examples.

Microbiota
The collective microbial community or population that resides in a particular locale at a given time.

Phylotypes
Groups of bacteria that are defined by percentage identity in their 16S rRNA gene sequences.

**Department of Genetics, Biology of Microorganisms, Anthropology and Evolution, University of Parma 43100, Italy.*
‡Department of Food, Bioprocessing and Nutrition Sciences, North Carolina State University, Raleigh, North Carolina 27695, USA.
§Alimentary Pharmabiotic Centre and Department of Microbiology, University College Cork, Western Road, Cork, Ireland.
Correspondence to P.W.O.
e-mail: pwotoole@ucc.ie
doi:10.1038/nrmicro2047
Published online 24 November 2008.

The availability of the human genome sequence has enabled us to better understand the genetic basis of many aspects of human health and disease. However, to fully understand the human genotype and its relationship with susceptibility to disease we need better information on how environmental and developmental factors interact with the genome to influence health. Human beings are colonized by, or transiently harbour, a diverse, complex and dynamic collection of bacteria that outnumber the human somatic and germ cells and that collectively represent significantly more genetic variety than the genomes of their hosts[1]. However, the components of the human microbiota remain poorly characterized. Recent culture-independent studies of the microbiota of the human gastrointestinal tract (GIT) have identified more than 1,000 phylotypes, which represent more than 7,000 strains and belong to 8 major phyla[1–4] (reviewed in REF. 5).

It has been suggested that the composition of the gut microbiota is the result of selective pressures that are imposed by the host, and is further modulated by competition between constituent bacterial members[6]. The interactions between bacteria and the human host can be categorized as a continuum that ranges from symbiosis and commensalism (mutualism) to pathogenesis. In the human gut, adaptive co-evolution of humans and bacteria has resulted in the establishment of commensal relationships in which neither partner is disadvantaged and in symbiotic relationships in which both partners benefit, be it from unique metabolic activities or from other benefits. The intestinal microbiota contributes to host nutrition[1,7,8] and impacts on intestinal cell proliferation and differentiation, pH, the development of the immune system and innate and acquired responses to pathogens[1,9,10].

Alterations in the composition of the intestinal microbiota have recently been linked to various conditions, including inflammatory bowel disease, allergy and obesity[6,11–14]. Among the variable constituents of the microbiota are health-promoting indigenous species (or mucosa-adherent microbiota). According to the Food and Agriculture Organization (FAO)/WHO criteria, probiotics are defined as "live microorganisms which when administered in adequate amounts confer a health benefit on the host"[15].

The mechanisms by which probiotic microorganisms benefit human health (reviewed in REFS 16,17) are typically divided into several 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 some evidence for each of these functional claims, the molecular mechanisms by which these activities are achieved remain largely unknown.

Genomics could accelerate 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 new discipline called probiogenomics[19], which aims to provide insights into the diversity and evolution of commensal and probiotic bacteria and to reveal the molecular basis for their health-promoting activities. The integration of

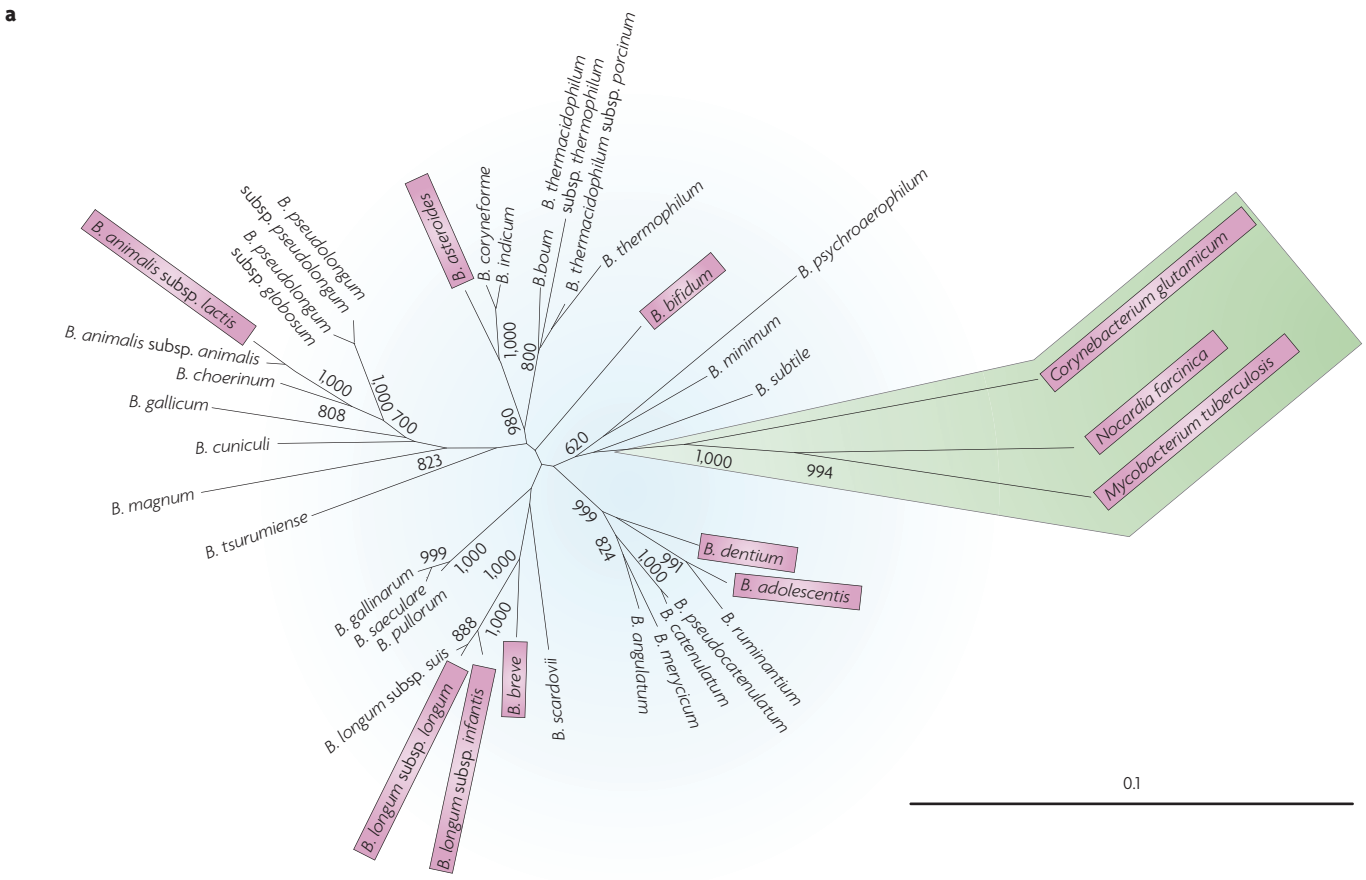

Table 1 | **General features of sequenced *Bifidobacterium* and *Lactobacillus* genomes**

Species	Genome size (basepairs)	% GC	Genes	Proteins	Source	Accession number	References
Bifidobacterium longum subsp. *longum* NCC2705	2,256,640	60%	1,798	1,727	Human GIT	NC_004307	24
Bifidobacterium longum subsp. *longum* DJ010A	2,375,286	59%	1,908	1,908	Human GIT	NC_010816	91
Bifidobacterium breve UCC2003	2,422,668	59%	1,868	1,590	Infant faeces	Unpublished	92
Bifidobacterium adolescentis ATCC15703	2,089,645	59%	1,701	1,631	Human GIT	NC_008618	Unpublished
Bifidobacterium adolescentis L2-32	2,385,710	59%	2,499	2,428	Infant faeces	NZ_AAXD00000000	Unpublished
Bifidobacterium animalis subsp. *lactis* HN019	1,915,892	60%	1,632	1,578	Unknown	NZ_ABOT00000000	Unpublished
Lactobacillus acidophilus NCFM	1,993,560	34%	1,936	1,862	Human GIT	NC_006814	55
Lactobacillus casei ATCC334	2,895,264	46%	2,909	2,751	Emmental cheese	NC_008526	82
Lactobacillus gasseri ATCC33323	1,894,360	35%	1,898	1,755	Human GIT	NC_008530	50
Lactobacillus johnsonii NCC533	1,992,676	34%	1,918	1,821	Human GIT	NC_005362	71
Lactobacillus plantarum WCFS1	3,308,274	44%	3,135	3,007	Human saliva	NC_004567	70
Lactobacillus reuteri F275	1,999,618	38%	2,027	1,900	Human GIT	NC_009513	60
Lactobacillus fermentum IFO 3956	2,098,685	51%	1,912	1,843	Fermented plant material	NC_010610	60
Lactobacillus salivarius subsp. *salivarius* UCC118	1,827,111	32%	1,864	1,717	Human GIT	NC_007929	51

GIT, gastrointestinal tract.

Neighbour-joining tree
A tree that reconstructs the evolutionary development of organisms on the basis of distances between pairs of taxa.

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

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 huge numbers of genes and proteins[20]. Probiogenomics is thus just one strand of gut systems microbiology. Significantly, when studied in combination with host genome variation, probiogenomics offers a comprehensive systems model, even at the individual subject level.

Here we address current developments in analysing the genome sequences of probiotic bacteria and how these data can be integrated into a global view using omics approaches to elucidate genome evolution and genetic adaptation of these bacteria to the human gut niche. We have focused on the model probiotic bacteria *Bifidobacterium* spp. and *Lactobacillus* spp., which are phylogenetically distant relatives (FIG. 1) that have different features from one another.

Genomics of the genus *Bifidobacterium*

The genus *Bifidobacterium* is small, with 30 characterized species and a low level of phylogenetic and genomic diversity[21] (FIG. 1a). Bifidobacteria were originally isolated from a breast-fed infant[22] and 30 species have since been isolated from the GIT contents of mammals, birds and insects[19]. Those bifidobacterial species that have been isolated from the human intestine have attracted the interest of genomic researchers owing to their probiotic properties. However, of the bifidobacterial taxa described so far, genomes of only three species, which belong to the *Bifidobacterium longum* and *Bifidobacterium adolescentis* groups, have been sequenced to completion (TABLE 1). The availability of six genome sequences provides genetic evidence that bifidobacteria are prototrophic and therefore well adapted to growth in an environment such as the human colon, which contains low concentrations of some growth substrates (for example, vitamins, amino acids and nucleotides)[23]. These bifidobacterial genome sequences harbour genes for the synthesis of at least 19 amino acids and they encode all of the enzymes that are needed for the biosynthesis of pyrimidine and purine nucleotides, as well as those that are required for the synthesis of the B vitamins, folic acid, thiamine and nicotinate[24] (S. Leahy and D.v.S., unpublished observations). Annotation and pathway prediction revealed that bifidobacterial species possess the genetic information that is required to shunt many monosaccharides or disaccharides into the fructose-6-phosphate pathway[23].

Adaptation to the human gut. The amount and types of 'non-digestible' saccharides in the diet (some of which are referred to as prebiotics) have major influences on the numbers and metabolic activities of different groups of bacteria in the enteric microbiota[25]. The range of polysaccharide substrates that arrive in the intestine is extremely broad[26]. This diversity of carbon substrates potentially generates a vast array

◀ Figure 1 | **Evolutionary relationships between the main gastrointestinal tract commensal bacterial groups.** Bifidobacteria are shown in panel **a** and lactobacilli are shown in panel **b**. Both panels are based on a neighbour-joining tree of 16S rRNA gene sequences. Bacterial taxa for which the whole-genome sequences are available are shaded in pink. Bootstrap values above 600 are indicated. The outgroups are shaded in green. Scale bars indicate 0.1 nucleotide substitutions per site.

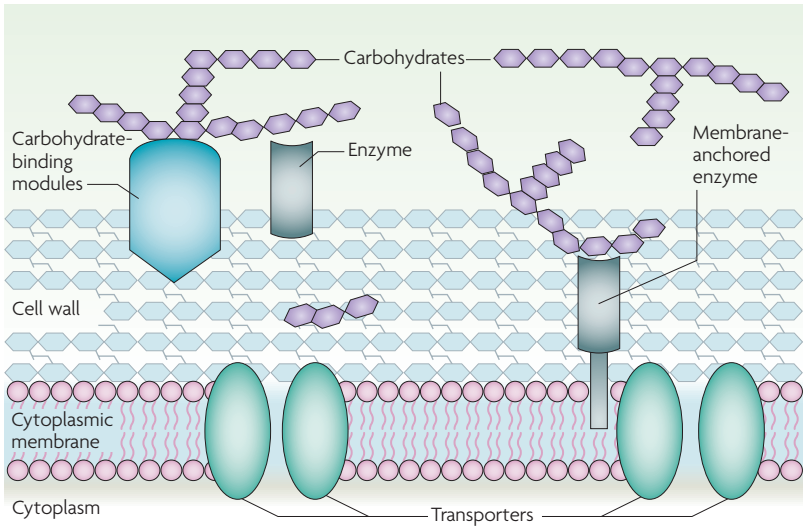

Figure 2 | **Acquisition of sugars by bifidobacteria.** The figure shows a strategy that might be adopted by bifidobacteria to acquire sugar nutrients. Bifidobacteria use a 'docking station' to capture complex sugars, such as xylan- and arabino-based molecules, and bind these to the bacterial cell surface to prevent loss of the sugars to competitors. 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 arabino- or xylan-based molecules to oligosaccharides that are subsequently transported across the bacterial membrane by a transporter protein; the presence of the bacterial cell-wall material might prohibit diffusion of nutrients away from the transporters.

of ecological niches that can be exploited by gut bacteria. Although some members of the gut microbiota can switch rapidly between using different substrates (for example, derived from diet or from host origin), others (for example, those bacteria associated with insoluble substrates) are far more specialized[27]. In this context, bifidobacteria have been presumed to have an ecological advantage owing to their capacity to metabolize complex sugars that are derived from the diet as well as from the host[28]. Genome annotation has confirmed that genes that are required for the breakdown of complex sugars are abundant in sequenced bifidobacterial genomes[19]. More than 8% of annotated bifidobacterial genes encode enzymes that are involved in carbohydrate metabolism. This is 30% higher than GIT-resident bacteria such as *Escherichia coli* or *Enterococcus faecium* and than non-GIT residents such as *Lactococcus lactis*[19]. However, the level of sugar-fermentative coding capacity in bifidobacteria is similar to that of one other intestinal commensal genus, *Bacteroides*[19]. Bifidobacterial enzymes that are involved in sugar metabolism include various glycosyl hydrolases (GH), which are used on diverse, but in most cases unidentified, plant-derived dietary fibres or complex carbohydrate structures.

Most of the bifidobacterial GHs are predicted to be intracellular, including those that are predicted to hydrolyse arabinogalactans and arabinoxylans, starch and related polysaccharides[24,29,30]. The genes for these GHs are associated with genetic loci for the uptake of structurally diverse sugar substrates. Altogether, about 5% of the total bifidobacterial gene content is dedicated to sugar internalization, through ATP-binding cassette (ABC) transporters, permeases and proton symporters rather than through phosphoenolpyruvate phosphotransferase systems[24,31,32]. Bifidobacteria use a 'docking station' to sequester and capture high-molecular-weight carbohydrate molecules such as xylose- and arabinose-containing polysaccharides (FIG. 2) and bind these to their cell surface[29,32], presumably to avoid losing them to nearby competitors. This is reminiscent of a putative carbohydrate utilization system that was identified in the genome of *Lactobacillus plantarum*[33] and in a system used by *Bacteroides thetaiotaomicron* for starch utilization[34]. Enteric bifidobacteria can also use sialic-acid-containing complex carbohydrates in mucin, glycosphingolipids and human milk[35,36]. Thus, these bifidobacteria have acquired adaptations to allow them to exploit a rich repertoire of otherwise indigestible components of the human or animal diet.

Whole bacterial genome sequencing efforts have also provided general indications about the genetic adaptation of some organisms to specific ecological niches. In the case of bifidobacteria, although genomic information is still currently limited to a few genomes, it was possible to identify an operon that encodes for enzymes that are involved in the breakdown of complex sugars such as starch, amylopectin and pullulan, which is present only in the genomes of *Bifidobacterium breve*[31]. As *B. breve* is one of the dominant bacteria in the infant microbiota[37], this enzyme might be important during weaning when non-milk foods are supplemented in the diet and when infants are, for the first time, exposed to complex carbohydrates that are different from those present in mother's milk.

Characterization of the metabolism of prebiotic compounds by bifidobacteria has led to the identification of specific transporters and hydrolases for oligosaccharides[29,38,39]. These studies indicated that bifidobacteria ferment different types of fructo-oligosaccharides; accordingly, the respective fructo-oligosaccharide metabolism operons have different genetic architectures[40], suggesting that these genes were acquired following evolutionary divergence of the species. Prebiotic oligosaccharides (such as galacto-oligosaccharides) are also contained in human milk and these are hydrolysed by bifidobacteria through the action of extracellular enzymes that are encoded by the *gal*A gene[29,41]. In addition to galacto-oligosaccharides, human milk provides large amounts of small peptides, which are derived from the digestion of milk proteins by the gastric protease pepsin[42]. *Bifidobacterium* genomes encode several enzymes, such as dipeptidyl aminopeptidases and oligopeptide uptake systems, that are involved in the breakdown and internalization of peptides (M.V. and D.v.S. unpublished observations).

Interaction with the host. Bacterium–host interactions that benefit the host can be elucidated by identification and molecular analysis of the bacterial proteins

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 health-promoting microorganism.

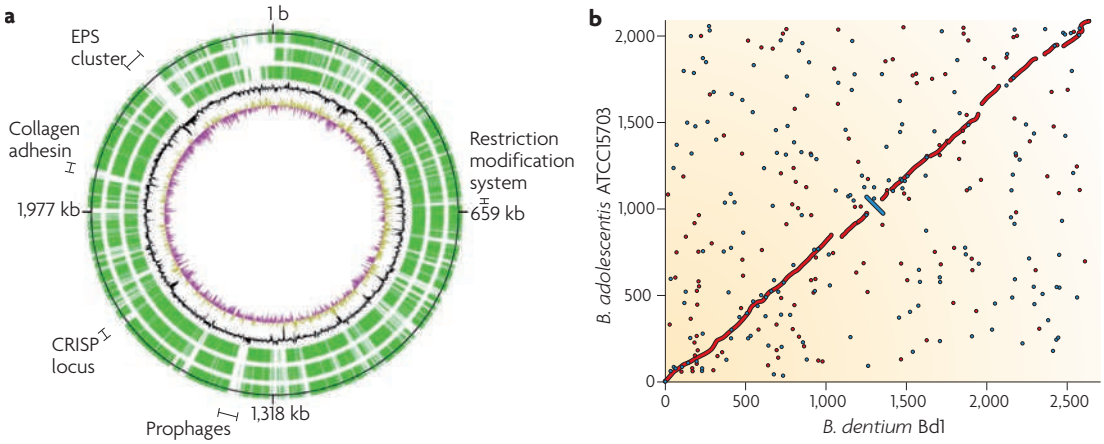

Figure 3 | **Comparative analysis of *Bifidobacterium* genomes. a** | Circular plot of genome diversity in bifidobacteria. The white and green colouring in the three outer rings indicates genome regions present and absent, respectively, in the bifidobacterial genomes, relative to the *Bifidobacterium dentium* Bd1 genome map. From outside to the outside: ring 1 shows a comparison with the genome sequence of *Bifidobacterium longum* subsp. *longum* NCC2705; ring 2 shows a comparison with the genome sequence of *B. longum* subsp. *longum* DJO10A; ring 3 shows a comparison with the genome sequence of *Bifidobacterium adolescentis* ATCC15703; ring 4 shows the GC content; ring 5 shows the GC deviation. Deviations from the average GC content are shown in either green (high GC spike) or violet (low GC spike). **b** | Comparison of gene-order conservation between two genome pairs, illustrating different forms of bifidobacterial genome evolution. The x and y axes represent the linearized chromosomes of *B. dentium* Bd1 and *B. adolescentis* ATCC15703, respectively. Blue dots indicate pairs of homologous genes that are in the same orientation in both genomes, whereas red dots indicate pairs that are in an inverted orientation in one relative to the other.

or macromolecules involved. For example, a potential probiotic effector molecule that is a homologue of the eukaryotic-type serine protease inhibitor (serpin) was identified in the genome of *B. longum* subsp. *longum*[24,43]. Members of the serpin family regulate various 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 model of colitis by reducing intestinal inflammation[43].

Transcriptomic approaches have been useful for studying how individual organisms in bacterial communities affect one another's transcriptomes. 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 that are 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* might 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. Complementation of phenotypes among community members has already been described in other microbial communities that degrade cellulose[46]. Alternatively, shifts in transcription patterns could represent responses to competition (see below).

The elucidation of the molecular impact of the human microbiota on the human host was 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. The host response to *B. thetaiotaomicron* was focused on tumour necrosis factor-α and lipopolysaccharide-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 host production of antibacterial proteins such as regenerating islet-derived-3γ (Reg3γ) and pancreatitis-associated protein (Pap). Thus, the host response to enteric bifidobacteria may not only promote bifidobacterial survival in the human intestine, but may 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

Transcriptome
The subset of genes that are transcribed in an organism. It represents dynamic links between a genome, proteins and cellular phenotypes.

Synteny
Genetic linkage or conservation of gene order.

Bacteriocins
Proteinaceous substances that are produced by one bacterium to kill another bacterium, usually by inducing leakage or lysis. Bacteriocins are composed of one or two short peptides that can be post-translationally modified.

COGs
Clusters of orthologous groups are delineated by comparing protein sequences that are encoded in complete genomes, representing major phylogenetic lineages. Each COG consists of individual proteins or groups of paralogues from at least 3 lineages and thus corresponds to an ancient conserved domain.

Autochthonous
Members of the microbiota that are growing where they are found, as distinct from transient species that are only passing through the environment.

identified. Most of these, including prophage-like elements, restriction modification systems, integrative plasmids and genes that are involved in the biosynthesis of extracellular structures such as exopolysaccharides, correspond to genetic elements that were presumably acquired by horizontal gene transfer (HGT) events (FIG. 3). Another set of genes that disseminated via HGT in bifidobacteria is the CRISPR-related system (CASS), which is implicated in defence against phages and plasmids[47] and which has been identified in the genome of *Bifidobacterium dentium* Bd1 as well as in the genome of *B. breve* UCC2003 (M.V. and D.v.S., unpublished observations; S. Leahy and D.v.S., unpublished observations). Notably, these *in silico* analyses were also confirmed by comparative genome hybridization analyses[48].

There is little phylogenetic diversity in the genus *Bifidobacterium* compared with *Lactobacillus* (see below). This is underlined at the whole-genome level when one compares the oral species (*B. dentium*), which is frequently identified as a component of the microbiota that is 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 that are related to specialized phenotypes.

Genomics of the genus *Lactobacillus*

The genus *Lactobacillus* has more than 100 cultured species (and probably more that are poorly culturable or non-culturable) and is noteworthy for its extreme phylogenetic, phenotypic and ecological diversity[50] (FIG. 1b). However, the real extent of *Lactobacillus* diversity is not fully known and culture-independent 16S rRNA gene surveys of complex ecosystems (for example, the human gut microbiota) are expected to uncover novel phylotypes that belong to the genus *Lactobacillus*. The microbiological characterization of lactobacilli is historically better developed than that of bifidobacteria, but the genomic analysis is recent. Of the 14 sequenced and published *Lactobacillus* genomes, 8 (*Lactobacillus acidophilus*, *Lactobacillus casei*, *Lactobacillus fermentum*, *Lactobacillus gasseri*, *Lactobacillus johnsonii*, *Lactobacillus reuteri*, *Lactobacillus salivarius* and *L. plantarum*) are from cultures or species that are considered to be probiotic (TABLE 1). Interestingly, 11% of the overall coding capacity of the *L. salivarius* genome is present on pMP118, the first megaplasmid described in lactic acid bacteria[51]. 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[51]. 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 to be replete with plasmids[9].

Adaptation to the human gut. The metabolic diversity of the *Lactobacillus* genome sequences that are available so far is illustrated in FIG. 4. Taking the *L. plantarum* WCFS1 genome as a reference, it is clear that there is considerable variation in the COG assignments of the gene sets that are harboured by the respective genomes. Intestinal lactobacilli compensate for their auxotrophy by encoding multiple 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[52]. More strikingly than is evident for bifidobacteria, the adaptation to life in the GIT becomes evident when the genome sequences of intestinal isolates are compared with food-adapted lactobacilli such as *Lactobacillus bulgaricus* and *Lactobacillus helveticus*. *L. bulgaricus* is widely used as a starter culture in yoghurt fermentations and has undergone genome decay to adapt to the milk environment[53]. Thus, it harbours numerous degraded or partial carbohydrate pathways and bile salt hydrolase pseudogenes[52,53]. In addition, *L. bulgaricus* has a preference for growth on lactose, further emphasizing its niche adaptation to milk. The genome sequence of *L. helveticus*, a widely used cheese starter culture, has been reported recently[54]. 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 phosphoenolpyruvate phosphotransferase systems for sugar utilization[54,55]. Additionally, no functional mucus-binding proteins or transporters for complex carbohydrates, such as raffinose and fructo-oligosaccharides, are encoded by the *L. helveticus* genome, reflecting the degree of adaptation of *L. helveticus* to a milk environment.

By contrast, *L. acidophilus* has adapted to the gut ecological niche by retaining the functional gene sets that are absent from *L. helveticus*, emphasizing the importance of these gene sets for probiotic functionality and niche adaptation by autochthonous lactobacilli that naturally reside 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 just three genes that were differentially expressed *in vivo*[56]. Bron *et al.*[57] used a modified *in vivo* expression technology to identify 72 genes that are expressed by *L. plantarum* in the mouse GIT, most of which were associated with carbon metabolism, amino-acid metabolism and stress resistance[57]. Notably, many of these functions in pathogens were associated with survival or adaptation. *L. casei* actively transcribes metabolic genes in the murine intestine and initiates *de novo* protein synthesis[58]. *L. johnsonii* NCC533 expresses different sets of genes depending on its location in the GIT[59], and surprisingly, 44% of the genome remains untranscribed both *in vitro* and *in vivo*[59]. Interestingly, the prolonged murine gut persistence of NCC533, but not

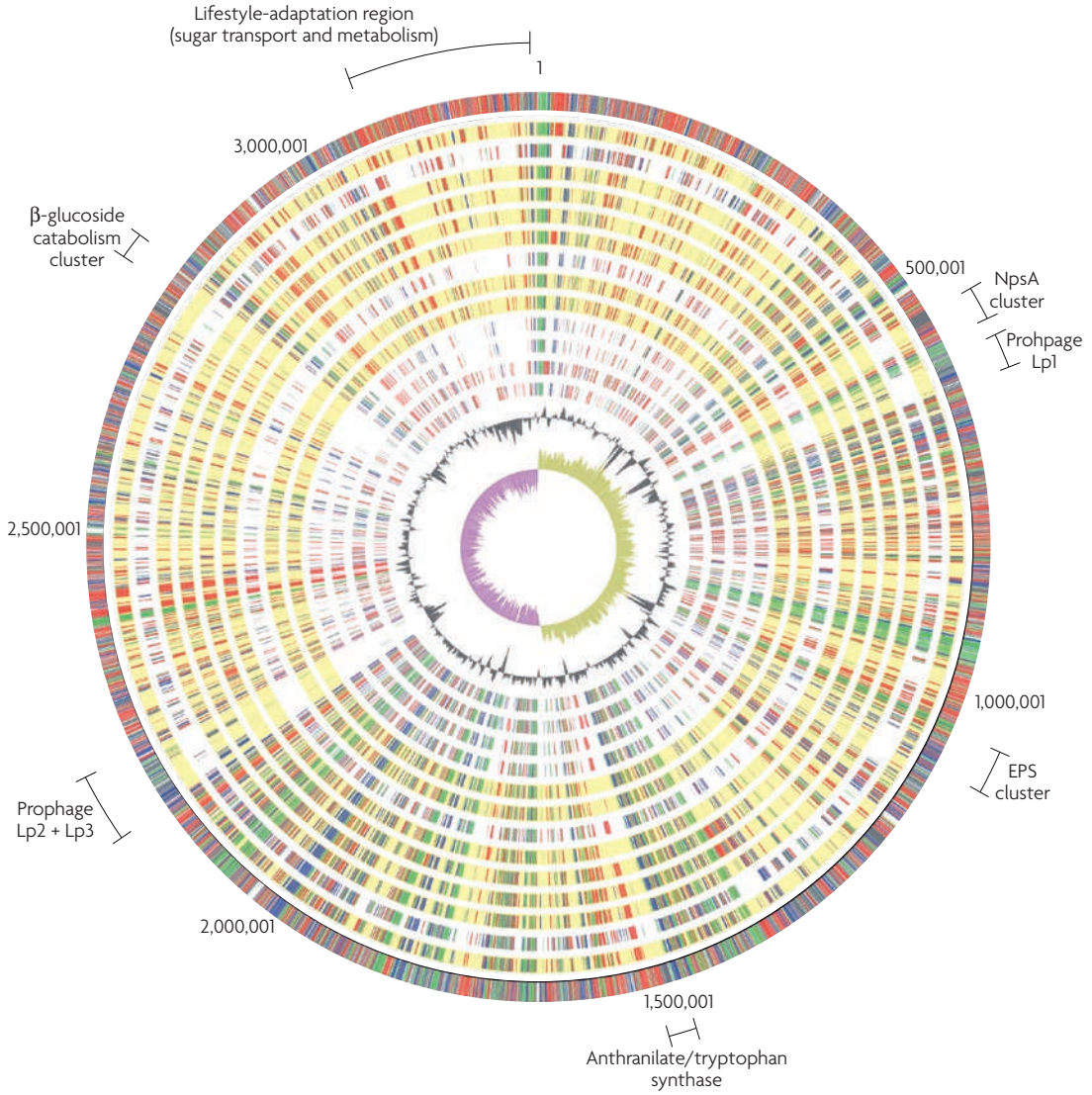

Figure 4 | **Comparative analysis of *Lactobacillus* genomes.** Circular genome atlas of *Lactobacillus plantarum* WCFS1 with mapped orthologues (defined as reciprocal best FastA hits with more than 30% identity over at least 80% of both protein lengths) from 13 publicly available *Lactobacillus* genomes. The outer circle shows *L. plantarum* WCFS1 followed, inwards, by *Lactobacillus salivarius*, *Lactobacillus brevis*, *Lactobacillus reuteri* F275, *L. reuteri* F275 (Japanese), *Lactobacillus fermentum*, *Lactobacillus acidophilus*, *Lactobacillus helveticus*, *Lactobacillus johnsonii*, *Lactobacillus gasseri*, *Lactobacillus bulgaricus* ATCC 11842, *L. bulgaricus* ATCC BAA-365, *Lactobacillus casei*, *Lactobacillus sakei*, GC percentage, and GC skew (green shows high GC spikes whereas violet shows low GC spikes; window-sizes 10,000 basepairs). COG categories in metabolism are shown in red, information storage and processing are shown in green, cellular processes and signalling are shown in blue, and poorly or not categorized COGs are shown in grey. Rings on yellow backgrounds indicate genomes from species that are considered to be resident in the gastrointestinal tract. EPS, exopolysaccharides; NpsA, non-ribosomal peptide synthetase.

of *L. johnsonii*, was recently shown to induce expression of exopolysaccharide synthesis genes, mannose-uptake genes and a gene for a putative protease in this strain[60]. In summary, although there are tantalizing glimpses of commensal *Lactobacillus* gene expression *in vivo*, these are as yet limited to animal models; data from human volunteer studies is keenly awaited.

Interaction with other commensal bacteria. Although the biology of commensal bacteria can be investigated in isolation, it must ultimately be understood in the context of the extremely complex intestinal ecosystem[61]. Lactobacillaceae account for approximately 36 phylotypes out of the >1,000 phylotypes in the human GIT microbiota[5]. In the short term, intervention studies in animal models and human subjects should provide key insights into our current understanding of interaction with other intestinal commensals.

Some lactobacilli have subtle effects on the microbiota. Consumption of *Lactobacillus rhamnosus* DR20 transiently alters the proportions of lactobacilli, bifidobacteria, enterococci and Bacteroidetes, but the

variations were generally small[62] and mechanisms were not investigated. The development of genomic tools facilitated a study that examined the molecular basis of interactions between the different components of the gut microbiota[45]. Such analyses were performed by the colonization of germ-free mice with *B. thetaiotaomicron* and *B. longum* as well as with *L. casei*, or combinations of these organisms[45]. Presence of *L. casei* resulted in an expanded capacity of *B. thetaiotaomicron* to metabolize polysaccharides and increased expression of genes for inorganic ion transport and metabolism[45]. The *L. casei*-induced changes in the *B. thetaiotaomicron* transcriptome were functionally similar to those caused by *B. longum*, but distinct from those induced by administration of *Bifidobacterium animalis* to the mice. Administration of *Lactobacillus paracasei* or *L. rhamnosus* to germ-free mice colonized with human infant microbiota caused modest changes in levels of a limited number of species monitored by culture techniques, but major 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 metabolic alterations that are described in these studies.

Numerous studies have reported that consumption of probiotics provides benefits for a range of GIT conditions and infections[64,65,66,67], but mechanistic insights are generally lacking. A reduction in the levels of vaginal *Lactobacillus spp.,* which results in vaginosis, has been linked to the production of a bacteriocin-like substance by commensal enterococci[66]. Also, the ability of *L. salivarius* to eliminate *Listeria monocytogenes* from a mouse model was dependent on the production of the broad spectrum bacteriocin Abp118 (also known as salivaricin)[67], and bacteriocin-producing lactobacilli become dominant among strains in a cocktail that reduces *Salmonella* shedding in pigs[68]. Thus, bacteriocin production is probably an important mechanism in the interaction of many lactobacilli with other commensals.

Comparative genomics of *Lactobacillus*

Sequencing of the genomes of 20 lactic acid bacteria 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 600–1,200 genes from a total gene repertoire of 2,100–2,200 (REF. 50). Many of these genes encoded biosynthetic enzymes or functioned in sporulation[50]. However, in addition to major gene losses, gene gains also occurred that seem to reflect the nutrient-rich niches, such as milk and the GIT, that are occupied by lactic acid bacteria. For example, genes encoding peptidases and amino-acid transport proteins as well as 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 that selective pressure from niche-specific adaptation has impacted on the genome evolution of these species[53,54,69].

In addition to gene duplication, HGT is also evident in probiotic lactobacilli. For example, the metabolic diversity of *L. plantarum* is underpinned by the expanded coding capacity that is afforded by its larger 3 Mb genome and by a low-GC-content region coding for sugar transport and metabolism genes that is likely to have been acquired by HGT[70]. Genes encoding cell-surface factors in *L. johnsonii* and the exopolysaccharide cluster in the *L. acidophilus* complex are further examples of HGT in probiotic lactobacilli[55,71]. Moreover, production of reuterin (3-hydroxypropionaldehyde), a potent broad-spectrum antimicrobial compound[72], is encoded by a genomic island that is present in some *L. reuteri* strains[73–75] and that is absent from the sequenced genome of a mouse *L. reuteri* isolate[74] and the closely related *L. fermentum*[75].

With genomes of 12 of the 147 recognized species[76] now fully sequenced, *Lactobacillus* spp. have been targeted for several comparative whole-genome analyses. Starting with the report of extreme diversity between the first two available genomes[77], genome sequencing of *L. acidophilus*, *L. gasseri*, *Lactobacillus delbrueckii* and *L. helveticus* allowed attention to be focused on the 'acidophilus complex'[54,55,78–80]. Large regions of synteny were observed between these species[55,78]. Multilocus sequence analysis of five housekeeping genes, comparative-genome hybridizations and DNA-typing revealed consistent and stepwise-decreasing levels of similarity in the group, indicating a strong role for vertical evolution[78]. Conversely, differences between trees from 16S rRNA genes and 401 core genes from *L. acidophilus*, *L. johnsonii* and *L. delbrueckii* indicated a high level (40%) of HGT[79].

To infer robust phylogenetic relationships with minimal incongruence, or to elucidate functional differences between species, a set of carefully selected single-copy ubiquitously-present genes is necessary. A comparison of 354 core genes from 5 lactobacilli underscored the substantial diversification of the genus and suggested that these lactobacilli could be subdivided into 3 groups[81]. Furthermore, 2 overlapping comparative studies, which included 9 additional Lactobacillales genomes, expanded the core genome to 567 order-specific genes[50,82]. The finer granularity provided by LaCOGs (Lactobacillales-specific COGs) allowed detection of two genes, the gene-contexts of which suggest housekeeping and protein-modification functions. Recently, we extracted 141 core genes from 12 *Lactobacillus* spp. genomes to investigate the case for a single congruent genus phylogeny[51,83]. These were operationally characterized by absent genes rather than by gained or retained genes, consistent with the findings of an earlier study[82].

Evolutionary trends in probiotic genomes

Collective analyses of probiotic genome sequences have revealed some conserved genetic traits[24,51,55,70,71,75,82], which might reflect adaptation to the intestinal niche[1].

However, as probiotic bacteria are diverse and taxonomically heterogeneous groups of microorganisms, the analysis of phyletic (phylogenetic) patterns, that is, patterns of gene presence/absence in a particular set of genomes, may be overwhelmingly influenced by the evolutionary distance between distant phyla. Nevertheless, common trends in the evolution of the genomes of both *Bifidobacterium* and *Lactobacillus* species can be discerned. These include gene loss (for example, 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. Many genes that are involved in sugar metabolism and transport were duplicated or acquired early in the evolution of probiotic bacteria, including those that encode enolase, β-galactosidase and many other GHs[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, such as β-lactamases, that are involved in antibiotic resistance in other bacteria[84].

Extensive evidence of HGT by bacteriophages or conjugation has been documented in Lactobacillales and seems 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 acquired by horizontal transfer from different sources[50]. In some cases the ancestor acquired an additional pseudoparalogous copy of a gene by HGT (for example, enolase in Lactobacillales), whereas in other cases xenologous displacement, that is, acquisition of genes by HGT followed by the loss of the ancestral orthologous gene[85], seems to have occurred.

With the imminent availability of an even greater number of whole-genome sequences from probiotic bacteria, a future challenge is the identification of the core probiogenome, which would comprise the core genome functions of probiotic bacteria. However, only seven genes present in bifidobacteria, but absent from 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 of the Lactobacillales genomes that have been sequenced so far[50].

Notably, many current claims of health-promoting properties in commercially available products that include probiotic agents are based on strain-specific properties. Thus, another intriguing goal of probiogenomics is to provide the molecular basis for such strain-specific genes and gene products. Large-scale parallel sequencing of multiple strains of single species will resolve issues such as conserved and variable gene families at inter- and intra-specific levels. The power of this approach has been demonstrated by a recent pathogenomic study that narrowed 10-fold the focus of a follow-up investigative phase of effector molecules[86]. In the case of *L. plantarum*, biodiversity-based screening was used to correlate comparative genomic hybridization patterns with a particular phenotype (mannose-sensitive adhesin) to successfully identify this gene from the genomic background[87]. Thus, comparative genomic analysis of probiotic strains with well-defined phenotypic characteristics can be a fruitful approach to identify strain-specific effector molecules/mechanisms that can then be functionally validated. However, other effector mechanisms that are probably involved in probiosis, such as the modulation of cytokine production by the composition of lipoteichoic acid[88], were not identified by a comparative genomics approach at all, so conserved components must not be overlooked.

Conclusions

Most of the probiotic bacteria marketed today were originally selected on the basis of technological stability or by various easily measurable phenotypes such as ability to tolerate bile salts or survive GIT passage, but not necessarily for their ability to confer health benefits. It is crucial to identify the precise mechanisms by which such probiotic microorganisms affect human health. Such studies should be accelerated by omics approaches, including genomics and functional analyses. Molecular interaction models are currently being developed, although more are required, to 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 validated molecular models with functional and comparative genomics-based approaches should enable selection of the most appropriate probiotic strain for a particular health benefit or should enable improvement of strain processing and administration regimes that optimize established health effects. This might allow the selection of specific probiotics for a particular human genotype, by analogy with personalized genomic medicine efforts.

Several issues regarding the sequences of complete probiotic bacterial genomes remain unresolved. So far, only a limited number of completed probiotic bacterial genome sequences are available, and these only partially represent the total biodiversity of probiotic bacteria residing in the human gut. In this context, understanding the human gut microbiome will be an important challenge for the future[89]. Furthermore, sequencing the genomes of environmental organisms and carrying out metagenomic surveys of diverse gut environments (human versus animal GITs, for example) 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 establishment of commensals (probiotics) in these different gut niches[90].

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 can be exploited.

Pseudoparalogous
An extra copy of a gene that is already present in a genome that was acquired by lateral gene transfer rather than by gene duplication.

Microbiome
The collective genome of microbial communities.

1. Backhed, F., Ley, R. E., Sonnenburg, J. L., Peterson, D. A. & Gordon, J. I. Host-bacterial mutualism in the human intestine. *Science* **307**, 1915–1920 (2005).
2. Eckburg, P. B. *et al.* Diversity of the human intestinal microbial flora. *Science* **308**, 1635–1638 (2005).
This article describes the bacterial diversity that occurs in the human gut, assessed 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**, 237–242 (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**, 2125–2136 (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**, 837–848 (2006).
7. Guarner, F. & Malagelada, J. R. Gut flora in health and disease. *Lancet* **361**, 512–519 (2003).
8. Hooper, L. V. & Gordon, J. I. Commensal host-bacterial relationships in the gut. *Science* **292**, 1115–1118 (2001).
9. Backhed, F. *et al.* The gut microbiota as an environmental factor that regulates fat storage. *Proc. Natl Acad. Sci. USA* **101**, 15718–15723 (2004).
10. Samuel, B. S. & Gordon, J. I. A humanized gnotobiotic mouse model of host–archaeal–bacterial mutualism. *Proc. Natl Acad. Sci. USA* **103**, 10011–10016 (2006).
11. Turnbaugh, P. J. *et al.* An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* **444**, 1027–1031 (2006).
12. Frank, D. N. *et al.* Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases. *Proc. Natl Acad. Sci. USA* **104**, 13780–13785 (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**, 205–211 (2006).
References 13 and 14 provide evidence for significant microbiota alterations in functional bowel disorders.
15. Food and Agriculture Organization of the United Nations and World Health Organization. Health and nutritional properties of probiotics in food including powder milk with live lactic acid bacteria. (FAO/WHO, Cordoba, Argentina, 2001).
16. Marco, M. L., Pavan, S. & Kleerebezem, M. Towards understanding molecular modes of probiotic action. *Curr. Opin. Biotechnol.* **17**, 204–210 (2006).
17. O'Hara, A. M. & Shanahan, F. Mechanisms of action of probiotics in intestinal diseases. *Scientific World J.* **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**, 204–211 (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. *Nature 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**, 2783–2792 (2006).
22. Tissier, M. H. *Recherche Sur La Flore Intestinale Des Nourissons (Etat Normal Et Pathologique)*. Thesis, Univ. Paris, France (1906).
23. 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**, 351–372 (2007).
24. Schell, M. A. *et al.* The genome sequence of *Bifidobacterium longum* reflects its adaptation to the human gastrointestinal tract. *Proc. Natl Acad. Sci. USA* **99**, 14422–14427 (2002).
25. Gibson, G. R. & Roberfroid, M. B. Dietary modulation of the human colonic microbiota: introducing the concept of prebiotics. *J. Nutr.* **125**, 1401–1412 (1995).
26. 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. *Nature Rev. Microbiol.* **6**, 121–131 (2008).
27. Sonnenburg, J. L. *et al.* Glycan foraging *in vivo* by an intestine-adapted bacterial symbiont. *Science* **307**, 1955–1959 (2005).
28. 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. USA* **96**, 9833–9838 (1999).
29. Hinz, S. W., Verhoef, R., Schols, H. A., Vincken, J. P. & Voragen, A. G. Type I arabinogalactan contains β-D-Galp-(1→3)-β-D-Galp structural elements. *Carbohydr. Res.* **340**, 2135–2143 (2005).
30. 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**, 5289–5296 (2006).
31. 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**, 545–553 (2007).
32. 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**, 146–163 (2008).
This paper provides the most up-to-date description of the enzymes encoded by bifidobacteria that are involved in the hydrolysis of carbohydrates.
33. 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).
34. 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).
35. 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**, 944–953 (1985).
36. 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**, 1936–1940 (2008).
37. Ventura, M., van Sinderen, D., Fitzgerald, G. F. & Zink, R. Insights into the taxonomy, genetics and physiology of bifidobacteria. *Antonie Van Leeuwenhoek* **86**, 205–223 (2004).
38. Ehrmann, M. A., Korakli, M. & Vogel, R. F. Identification of the gene for β-fructofuranosidase of *Bifidobacterium lactis* DSM10140(T) and characterization of the enzyme expressed in *Escherichia coli*. *Curr. Microbiol.* **46**, 391–397 (2003).
39. Katayama, T. *et al.* Molecular cloning and characterization of *Bifidobacterium bifidum* 1,2-α-L-fucosidase (AfcA), a novel inverting glycosidase (glycoside hydrolase family 95). *J. Bacteriol.* **186**, 4885–4893 (2004).
40. Ryan, S. M., Fitzgerald, G. F. & van Sinderen, D. Transcriptional regulation and characterization of a novel β-fructofuranosidase-encoding gene from *Bifidobacterium breve* UCC2003. *Appl. Environ. Microbiol.* **71**, 3475–3482 (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.* **74**, 4686–4694 (2008). .
42. Liepke, C. *et al.* Human milk provides peptides highly stimulating the growth of bifidobacteria. *Eur. J. Biochem.* **269**, 712–718 (2002).
43. Ivanov, D. *et al.* A serpin from the gut bacterium *Bifidobacterium longum* inhibits eukaryotic elastase-like serine proteases. *J. Biol. Chem.* **281**, 17246–17252 (2006).
44. Potempa, J., Korzus, E. & Travis, J. The serpin superfamily of proteinase inhibitors: structure, function, and regulation. *J. Biol. Chem.* **269**, 15957–15960 (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 that exists between bifidobacteria and *Bacteroides* in the murine intestine as well as between these bacteria and their hosts.
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**, 7099–7106 (2005).
47. Barrangou, R. *et al.* CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315**, 1709–1712 (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**, 1407–1417 (2008).
50. Makarova, K. S. & Koonin, E. V. Evolutionary genomics of lactic acid bacteria. *J. Bacteriol.* **189**, 1199–1208 (2007).
51. Claesson, M. J. *et al.* Multireplicon genome architecture of *Lactobacillus salivarius*. *Proc. Natl Acad. Sci. USA* **103**, 6718–6723 (2006).
52. Pfeiler, E. A. & Klaenhammer, T. R. The genomics of lactic acid bacteria. *Trends Microbiol.* **15**, 546–553 (2007).
53. van de Guchte, M. *et al.* The complete genome sequence of *Lactobacillus bulgaricus* reveals extensive and ongoing reductive evolution. *Proc. Natl Acad. Sci. USA* **103**, 9274–9279 (2006).
54. Callanan, M. *et al.* Genome sequence of *Lactobacillus helveticus*, an organism distinguished by selective gene loss and insertion sequence element expansion. *J. Bacteriol.* **190**, 727–735 (2008).
55. Altermann, E. *et al.* Complete genome sequence of the probiotic lactic acid bacterium *Lactobacillus acidophilus* NCFM. *Proc. Natl Acad. Sci. USA* **102**, 3906–3912 (2005).
56. Walter, J. *et al.* Identification of *Lactobacillus reuteri* genes specifically induced in the mouse gastrointestinal tract. *Appl. Environ. Microbiol.* **69**, 2044–2051 (2003).
57. 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**, 5721–5729 (2004).
This manuscript provides insight into the interactions between a commensal bacterium and its murine host.
58. 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**, 1356–1363 (2005).
59. Denou, E. *et al.* Gene expression of commensal *Lactobacillus johnsonii* strain NCC533 during *in vitro* growth and in the murine gut. *J. Bacteriol.* **189**, 8109–8119 (2007).
60. 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**, 3161–3168 (2008).
61. Whitman, W. B., Coleman, D. C. & Wiebe, W. J. Prokaryotes: the unseen majority. *Proc. Natl Acad. Sci. USA* **95**, 6578–6583 (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**, 2578–2588 (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* 62–66: potential significance for bacterial vaginosis. *Infect. Dis. Obstet. Gynecol.* **11**, 147–156 (2003).

67. Corr, S. C. *et al.* Bacteriocin production as a mechanism for the antiinfective activity of *Lactobacillus salivarius* UCC118. *Proc. Natl Acad. Sci. USA* **104**, 7617–7621 (2007).
This study identified the first molecular mechanism by which 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**, 1858–1863 (2007).
69. Makarova, K. *et al.* Comparative genomics of the lactic acid bacteria. *Proc. Natl Acad. Sci. USA* **103**, 15611–15616 (2006).
This landmark study provided a large tranche of genomic data to allow studies of genome evolution in lactic acid bacteria.
70. Kleerebezem, M. *et al.* Complete genome sequence of *Lactobacillus plantarum* WCFS1. *Proc. Natl Acad. Sci. USA* **100**, 1990–1995 (2003).
This is the first article describing the genome sequence of a member of the genus *Lactobacillus*.
71. Pridmore, R. D. *et al.* The genome sequence of the probiotic intestinal bacterium *Lactobacillus johnsonii* NCC 533. *Proc. Natl Acad. Sci. USA* **101**, 2512–2517 (2004).
This paper describes the genome of a commonly used probiotic bacterium belonging to the genus *Lactobacillus*.
72. 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**, 1854–1858 (1988).
73. Santos, F. *et al.* The complete coenzyme B12 biosynthesis gene cluster of *Lactobacillus reuteri* CRL1098. *Microbiology* **154**, 81–93 (2008).
74. 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**, 4559–4567 (2008).
75. 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**, 151–161 (2008).
76. Euzeby, J. P. List of bacterial names with standing in nomenclature: a folder available on the internet. *Int. J. Syst. Bacteriol.* **47**, 590–592 (1997).
77. Boekhorst, J. *et al.* The complete genomes of *Lactobacillus plantarum* and *Lactobacillus johnsonii* reveal extensive differences in chromosome organization and gene content. *Microbiology* **150**, 3601–3611 (2004).
78. Berger, B. *et al.* Similarity and differences in the *Lactobacillus acidophilus* group identified by polyphasic analysis and comparative genomics. *J. Bacteriol.* **189**, 1311–1321 (2007).
79. 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).
80. Klaenhammer, T. R., Barrangou, R., Buck, B. L., Azcarate-Peril, M. A. & Altermann, E. Genomic features of lactic acid bacteria effecting bioprocessing and health. *FEMS Microbiol. Rev.* **29**, 393–409 (2005).
81. Canchaya, C., Claesson, M. J., Fitzgerald, G. F., van Sinderen, D. & O'Toole, P. W. Diversity of the genus *Lactobacillus* revealed by comparative genomics of five species. *Microbiology* **152**, 3185–3196 (2006).
82. Makarova, K. *et al.* Comparative genomics of the lactic acid bacteria. *Proc. Natl Acad. Sci. USA* **103**, 15611–15616 (2006).
83. Claesson M. J., von Sinderen, D. & O'Toole, P. W. *Lactobacillus* phylogenomics — towards a reclassification of the genus. *Int. J. Sys. Evo. Microbiol.* (in press).
84. Teuber, M., Meile, L. & Schwarz, F. Acquired antibiotic resistance in lactic acid bacteria from food. *Antonie Van Leeuwenhoek* **76**, 115–137 (1999).
85. Koonin, E. V., Makarova, K. S. & Aravind, L. Horizontal gene transfer in prokaryotes: quantification and classification. *Annu. Rev. Microbiol.* **55**, 709–742 (2001).
86. Lloyd, A. L., Rasko, D. A. & Mobley, H. L. Defining genomic islands and uropathogen-specific genes in uropathogenic *Escherichia coli*. *J. Bacteriol.* **189**, 3532–3546 (2007).
87. Pretzer, G. *et al.* Biodiversity-based identification and functional characterization of the mannose-specific adhesin of *Lactobacillus plantarum*. *J. Bacteriol.* **187**, 6128–6136 (2005).
88. Grangette, C. *et al.* Enhanced antiinflammatory capacity of a *Lactobacillus plantarum* mutant synthesizing modified teichoic acids. *Proc. Natl Acad. Sci. USA* **102**, 10321–10326 (2005).
89. Turnbaugh, P. J. *et al.* The human microbiome project. *Nature* **449**, 804–810 (2007).
90. Ley, R. E. *et al.* Evolution of mammals and their gut microbes. *Science* **320**, 1647–1651 (2008).
This paper describes the bacterial diversity that exists in the gut of numerous mammals.
91. 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).
92. Leahy, S. C., Higgins, D. G., Fitzgerald, G. F. & van Sinderen, D. Getting better with bifidobacteria. *J. Appl. Microbiol.* **98**, 1303–1315 (2005).

Acknowledgements
Work in the laboratories of D.v.S. and P.W.O.T. is supported by a Science Foundation Ireland Centres for Science, Engineering & Technology (SFI CSET) award to the Alimentary Pharmabiotic Centre and a Department of Agriculture and Food (DAF)/Health Research Board, Food-Health Research Initiative (HRB FHRI) FHRI award to the ELDERMET project. M.V. was supported by an Italian Award for Outstanding Young Researcher scheme "Incentivazione alla mobilità di studiosi stranieri e italiani residente all'estero" 2005–2009, a Marie Curie Reintegration Grant (MERG-CT-2005-03,080) and Parmalat spa, Italy. We also thank C. Canchaya for helpful discussions. Work on genomics of lactobacilli at North Carolina State University, USA, is supported by the NC Dairy Foundation, Danisco USA Inc. and Dairy Management Inc.

DATABASES
Entrez Genome Project: http://www.ncbi.nlm.nih.gov/sites/entrez?Db=genomeprj
Bacteroides thetaiotaomicron | *Bifidobacterium adolescentis* | *Bifidobacterium breve* | *Bifidobacterium dentium* | *Bifidobacterium longum* | *Enterococcus faecium* | *Escherichia coli* | *Lactobacillus acidophilus* | *Lactobacillus delbrueckii* | *Lactobacillus fermentum* | *Lactobacillus gasseri* | *Lactobacillus helveticus* | *Lactobacillus johnsonii* | *Lactobacillus plantarum* | *Lactobacillus reuteri* | *Lactobacillus rhamnosus* | *Lactobacillus salivarius* | *Lactococcus lactis*

FURTHER INFORMATION
Alimentary Pharmabiotic Centre: http://www.ucc.ie/research/apc/content
ELDERMET: http://eldermet.ucc.ie
Univeristy of Parma: http://www.unipr.it

ALL LINKS ARE ACTIVE IN THE ONLINE PDF

Biographies

Marco Ventura is a non-tenure track Lecturer at the Department of Genetics, Anthropology and Evolution, University of Parma, Italy. He leads a probiogenomics laboratory and coordinates selected genome sequencing projects. He has worked as a postdoctoral research scientist at the Department of Microbiology, National University of Ireland, Cork, and as a Ph.D. student at the Nestle Research Center, Lausanne, Switzerland. Ventura's main research interest is human gut systems microbiology with specific emphasis on the exploration of the genome functionality of probiotic bacteria. His laboratory also studies the bacterial diversity present in the human gut, genetics of stress response in bifidobacteria, bioinformatics of probiotics, the molecular biology of bacteriophage that infect lactic acid bacteria and bacterial phylogeny, particularly of high GC-content Gram-positive bacteria.

Sarah O'Flaherty obtained her B.A. (Mod) degree in Microbiology from Trinity College Dublin, Ireland, and Ph.D. degree in Microbiology from University College Cork, Ireland. Her Ph.D. research focused on the role of bacteriophages as potential antimicrobial agents for *Staphylococcus aureus*. In 2006, she joined the laboratory of Todd Klaenhammer in the Food, Bioprocessing and Nutrition Sciences Department at North Carolina State University, USA, as a postdoctoral research associate. Her current research focuses on functional genomics of probiotics and bacteria–host interactions.

Marcus Claesson completed a B.Sc. in Chemical and Physics Engineering and an M.Sc. in Bioinformatics at Chalmers University of Technology Gothenburg, Sweden. He worked for a year at the start-up biotechnology company AngioGenetics before enrolling for a Ph.D. in bacterial genomics with Paul O'Toole and Douwe van Sinderen at University College Cork, Ireland. After graduation in 2006, Claesson worked for 9 months at the campus-based biotechnology company Alimentary Health. Since January 2008, he has led the bioinformatics analysis platform in the ELDERMET metagenomics project, which aims to use metagenomic approaches to determine the composition and activities of the gut microbiome in several hundred elderly Irish subjects.

Francesca Turroni is a Ph.D. student in the laboratory of probiogenomics at the Department of Genetics, Anthropology and Evolution, University of Parma, Italy. In 2005, she received a B.Sc. in Food Sciences at the University of Parma and since then she has focused on research in the fields of food microbiology and microbial ecology. In 2006 she began her Ph.D., which focuses on the molecular and genomic analysis of bifidobacterial communities that reside in humans, with a particular emphasis on those bacteria that exert health-promoting effects on the host (probiotic bifidobacteria).

Todd R. Klaenhammer obtained degrees in Microbiology (B.S.), and Food Science (M.S. and Ph.D.) from the University of Minnesota, USA. In 1978 he joined the North Carolina State University, USA, and currently holds faculty appointments in the Departments of Food Science, Microbiology, and Genetics, as a distinguished university professor. For 30 years he has directed a research programme on the genetics of lactic acid bacteria that are used as probiotics or as starter cultures for food bioprocessing and biotechnology applications. Todd is Fellow in the American Academy of Microbiology, the Institute of Food Technologists and the American Dairy Science Association. In 2001 he was elected into the National Academy of Sciences.

Douwe van Sinderen is an associate professor in the Department of Microbiology at University College Cork (UCC), Ireland, and a Principal Investigator in the Alimentary Pharmabiotic Centre (APC), Ireland, at UCC. He received his B.Sc. (in Biochemistry) and Ph.D. (in Molecular Genetics) degrees from the University of Groningen, The Netherlands, before working as a research scientist and long-term EMBO postdoctoral fellow in the National Food Biotechnology Centre at UCC. He is (co)author of more than 125 research papers, on topics that relate to the biotechnology, molecular biology and genomics of lactic acid bacteria, bifidobacteria and their bacteriophages. In the APC, his research activities are focused on the molecular genetics of stress response, environmental sensing and carbohydrate metabolism of bifidobacteria as relevant to probiotic functioning, efficacy and survival.

Paul O'Toole is Senior Lecturer in the Department of Microbiology and Principal Investigator in the Alimentary Pharmabiotic Centre at University College Cork, Ireland. Following education in Trinity College Dublin, he underwent postdoctoral training in Sweden and Canada, and was a lecturer for 7 years at Massey University, New Zealand. He has authored more than 60 research papers on the molecular biology and genomics of gut bacteria. The Alimentary Pharmabiotic Centre, Ireland, is a multi-disciplinary national centre of excellence for research into gut bacteria. He is also coordinator of ELDERMET, a multi-partner metagenomics project that is funded by the Irish Department of Agriculture & Food and the Irish Health Research Board. His research focuses on the genomic basis for survival and adaptation of pathogens and commensals in the human and animal intestine, and on how variations in the gut microbiota affect health and disease.

ToC

The genomics of probiotic bacteria, or probiogenomics, could shed light on how beneficial gut bacteria adapt to the gut environment and promote better gut health.

Online summary

- The human gastrointestinal tract (GIT) is a complex ecosystem, the bacterial components (microbiota) of which are thought to have a significant role in normal gut function and in maintaining host health. The human gut microbiota include health-promoting indigenous species such as *Bifidobacterium* and *Lactobacillus*, also referred to as probiotic bacteria. Probiotic bacteria are commonly consumed live as dietary supplements.
- The molecular mechanisms by which probiotic bacteria exert their health-promoting effects remain largely unclear. However, the advent of a novel scientific discipline, called probiogenomics, has recently provided new insights into the diversity and evolution of probiotic bacteria and has revealed the molecular basis of probiosis.
- Probiogenomic efforts have shown how the genome content of bifidobacteria and lactobacilli reflect adaptations to the human intestinal niche. Genomic evidence for adaptations to the GIT includes metabolic features, such as the capacity for uptake of macromolecules and breakdown of undigested complex carbohydrates, and the ability to interact with the host through the production of cell-surface proteins that interact with the intestinal mucosa.
- The interaction of probiotic bacteria with the host, as well as with other components of the human gut microbiota, is considered a key feature of probiosis. Bifidobacteria induce an expansion in the diversity of polysaccharides that are targeted for degradation

by common intestinal bacteria (*Bacteroides*) and also induce the expression of host genes that play a part in innate immunity.

- Comparisons among completely sequenced bifidobacterial and lactobacilli genomes revealed that the main force that drives evolution in these genomes is horizontal gene transfer.
- Probiotic bacteria are diverse and taxonomically heterogeneous groups of microorganisms, so the analysis of phyletic patterns — that is, patterns of gene presence or absence in a particular set of genomes — might be influenced by the evolutionary distance between these distant phyla. Nevertheless, comparative analyses of genomes from probiotic bacteria revealed a core genome (probiogenome), which encodes key functions of this group of microorganisms.