CORA Cork Open Research Archive
Cartlann Taighde Oscailte Chorcaí

| Title | Genome-scale analyses of health-promoting bacteria: probiogenomics |
|---|---|
| Author(s) | Ventura, Marco; O'Flaherty, Sarah; Claesson, Marcus J.; Turroni, Francesca; Klaenhammer, Todd R.; van Sinderen, Douwe; O'Toole, Paul W. |
| Publication date | 2009-01 |
| 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 | http://dx.doi.org/10.1038/nrmicro2047 <br> Access to the full text of the published version may require a subscription. |
| Rights | © 2009 Macmillan Publishers Limited. All rights reserved <br> http://www.nature.com/authors/editorial_policies/license.html |
| Item downloaded from | http://hdl.handle.net/10468/118 |

UCC

University College Cork, Ireland
Coláiste na hOllscoile Corcaigh

1 **Genome-scale analyses of health-promoting bacteria: probiogenomics**

2

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

4

5 **Marco Ventura[*], Sarah O'Flaherty[‡], Marcus J. Claesson[§], Francesca Turroni[*], Todd R.**

6 **Klaenhammer[‡], Douwe van Sinderen[§], and Paul W. O'Toole[§]**

7

8

9

10 [*]Department of Genetics, Biology of Microorganisms, Anthropology and Evolution, University of

11 Parma, Italy, [‡]Department of Food, Bioprocessing and Nutrition Sciences, North Carolina State

12 University, Raleigh, North Carolina 27695, [§]Alimentary Pharmabiotic Centre and Department of

13 Microbiology, University College Cork, Western Road, Cork, Ireland

14

15

16

17

18

19

20

21

22

23

24

25 Correspondence to P.W.O

26 e.mail: pwotoole@ucc.ie

27  **Abstract**

28  The human body is colonized by an enormous population of bacteria (microbiota) that outnumbers

29  the human somatic and germ cells and provides the host with additional coding capacity and

30  metabolic activities. Among the human gut microbiota are health-promoting indigenous species,

31  also referred to as probiotic bacteria, which are commonly consumed as live dietary supplements.

32  Although there is a growing list of health benefits provided by the consumption of probiotics, their

33  precise mechanisms of action remain largely unknown. Recent genomics based studies

34  (probiogenomics) are starting to provide insights into the ways probiotic bacteria sense and adapt to

35  the gastrointestinal tract environment.  In this review, we will discuss the application of

36  probiogenomics in the elucidation of the molecular basis of probiosis using the well recognized

37  model probiotic bacteria *Bifidobacterium* and *Lactobacillus* as examples.

38

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

64  live dietary supplements[15]. The mechanisms by which probiotic micro-organisms beneficially affect

65  human health (reviewed in[16, 17]) are typically divided into a number of general categories, including

66  strengthening of the intestinal barrier, modulation of the immune response and antagonism of

67  pathogens either by the production of antimicrobial compounds or through competition for mucosal

68  binding sites[16, 18]. Although there is suggestive evidence for each of these functional claims, the

69  molecular mechanisms remain largely unknown.

70      Genomics offers the possibility of accelerating research into probiotic bacteria. In recent

71  years, genome sequencing of gut commensals and symbionts has come to the fore, currently

72  represented by the development of a novel scientific discipline, called probiogenomics[19], which

73  aims to provide insights into the diversity and evolution of commensal/probiotic bacteria and to

74  reveal the molecular basis for their health-promoting activities. The integration of probiogenomics

75  and functional genomic information with data on host gene expression in the human gut will expand

76  our understanding of the roles of (probiotic) microbiota, microbe-microbe and host-microbe

77  interactions. These "omics" approaches allow the simultaneous analysis of very large numbers of

78  genes or proteins[20]. Probiogenomics is thus one strand of gut systems microbiology. Significantly,

79  when studied in combination with host genome variation, probiogenomics offers a comprehensive

80  systems model, even at individual subject level.

81      Here we address current developments in analyzing the genome sequences of probiotic

82  bacteria and how these data can be integrated in a global view using omics approaches in order to

83  elucidate genome evolution and genetic adaptation of these bacteria to the human gut ecological

84  niche. We consider the well recognized model probiotic bacteria *Bifidobacterium* spp. and

85  *Lactobacillus* spp. which are phylogenetically distant (although well-characterized; Fig 1), have

86  distinguishing properties, and different depths of biological characterization.

87

88

89

**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-

142 oligosaccharides, human milk consumption provides large amounts of small peptides that are

143 derived from the digestion of milk proteins by the gastric protease pepsin[42]. *Bifidobacterium*

144 genomes encode a rich repertoire of enzymes involved in the breakdown and internalization of

145 peptides such as dipeptidyl aminopeptidases and oligopeptide uptake systems (Ventura et al.,

146 unpublished data).

147 *Molecular interaction with the host.*

148 Bacterium-host interactions that result in host benefit can be elucidated by identification and

149 detailed molecular analysis of the bacterial proteins or macromolecules involved. For example a

150 potential probiotic effector molecule, a eukaryotic-type serine protease inhibitor (serpin) was

151 identified in the genome of *B. longum* subsp. *longum*[25, 43]. Members of the serpin family regulate a

152 wide range of signalling pathways in eukaryotes and some are recognized for their ability to

153 suppress inflammatory responses by inhibiting elastase activity[44]. Recent findings showed that the

154 bifidobacterial serpin-like protein performs an immunomodulatory role in a murine colitis model,

155 by reducing intestinal inflammation [43].

156       Transcriptomic approaches facilitate studies of gene expression profiles and have been

157 successful in studying how individual organisms in bacterial communities affect each other's

158 transcriptome. Recent transcriptomic analyses were performed on bacteria from germ-free mice that

159 had been mono-associated with *B. thetaiotaomicron* —one of the dominant components of the

160 human gut microbiota — and subsequently challenged with *B. longum* subsp. *longum* . The

161 presence of *B. longum* subsp. *longum* provoked an expansion in the diversity of polysaccharides

162 targeted for breakdown by *B. thetaiotaomicron* such as mannose and xylose-containing glycans[45].

163 The changes in the transcriptional profiles of polysaccharide-utilization related genes by *B. longum*

164 subsp. *longum* and *B. thetaiotaomicron* may imply the existence of symbiosis between these

165 microbial species, where each species possesses a complement of GH activities, which when

166 combined allow both to participate in a synergic harvest of xylose and mannose-containing sugars.

167 This phenomenon has already been described in other microbial communities that degrade

168 cellulose[46]. Alternatively, the shifts in transcription patterns could represent response to competition

169 (see also below for lactobacilli).

170      The elucidation of the molecular impact generated by members of the human microbiota on

171 the human host was also analysed by studying the host epithelium response to co-colonization by *B.*

172 *longum* subsp. *longum* and *B. thetaiotaomicron*[45]. Remarkably, the host response to these two

173 bacterial species was different. In fact, the host response to *B. thetaiotaomicron* was more focused

174 on tumor necrosis factor α and LPS-responsive cytokine produced by natural killer and T

175 macrophages, whereas *B. longum* subsp. *longum* promoted the activation of T-cell-produced

176 cytokine interferon-γ and reduced production by the host of antibacterial proteins such as Reg3γ

177 (Regenerating islet-derived-3γ) and Pap (Pancreatitis-associated protein). Thus the host response to

178 enteric bifidobateria may not only promote their own survival in the human intestine but also affect

179 the composition of the overall human gut microbiota.

180

181 **Comparative genomics of bifidobacteria**

182 Comparisons at the nucleotide level of the fully sequenced bifidobacterial genomes revealed a high

183 degree of conservation and synteny across the entire genomes[19]. However, several breakpoint

184 regions were also reported, apparently representing inversions or DNA deletion/insertion points.

185 DNA regions uniquely present in one genome and absent in others were also identified. Most of

186 these correspond to genetic elements presumably acquired by horizontal gene transfer events

187 (HGT), including prophage-like elements, restriction modification systems, integrative plasmids,

188 and genes involved in the biosynthesis of extracellular structures such as exopolysaccharides (EPS)

189 (Fig. 3). Another set of genes disseminated via HGT in bifidobacteria is the CRISPR-related system

190 (CASS) implicated in defence against phages and plasmids[47], which have been identified in the

191 genome of *B. dentium* Bd1 as well as in the genome of *B. breve* UCC2003 (Ventura et al.,

192 unpublished data; Leahy and D. van Sinderen, unpublished data). Notably these *in silico* analyses

193 were also confirmed by comparative genome hybridization analyses[48].

194         There is relatively little phylogenetic diversity within the genus *Bifidobacterium* compared

195    to *Lactobacillus* (see below). This is underlined at whole genome level when one compares the oral

196    species (*B. dentium*), which is frequently identified as a component of the microbiota associated

197    with dental caries[49] with the probiotic species *B. adolescentis* (Fig. 3). Despite the large phenotypic

198    differences, there is a remarkable degree of overall synteny. This reductionist model of genome

199    evolution may be useful for identifying niche-specific genes and genes related to specialized

200    phenotypes.

201

202    **Genomics of *Lactobacillus***

203    The genus *Lactobacillus* has more than 100 species, and is noteworthy for its extreme phylogenetic,

204    phenotypic and ecological diversity[22]. The microbiological characterization of lactobacilli is

205    historically better developed than that of bifidobacteria, but the genomic analysis is similarly recent.

206    Of the 14 sequenced and published *Lactobacillus* genomes, eight (*L. acidophilus*, *L. casei*, *L.*

207    *fermentum*, *L. gasseri*, *L. johnsonii*, *L. reuteri*, *L. salivarius* and *L. plantarum*) are from

208    cultures/species considered probiotic (Table 1). Interestingly, 11% of the overall coding capacity of

209    the *L. salivarius* genome lies on the first megaplasmid described in lactic acid bacteria; pMP118[22].

210    This megaplasmid encodes biologically important features such as a locus for bacteriocin

211    production, a bile salt hydrolase, and two genes that complete the phosphoketolase pathway,

212    officially reclassifying this organism as a facultative heterofermenter[22]. In fact, plasmids account for

213    15% of the genome of *L. salivarius*, which is not the case with other sequenced probiotic

214    lactobacilli, even though members of this genus are considered relatively replete with plasmids[9].

215    *Adaptation to the human gut*.

216    The metabolic diversity revealed by the *Lactobacillus* genome sequences available to date is

217    illustrated in Fig 4. Taking the *L. plantarum* WCFS1 genome as reference, it is clear that there is

218    considerable variation in the COG assignments of the gene sets harboured by the respective

219    genomes. Intestinal lactobacilli compensate for their relative degree auxotrophy by being rich in

220    genes for transporters. Their genomes also contain genes that encode acid and bile resistance,

221    capacity for uptake of macromolecules, metabolism of complex carbohydrates, and cell surface

222    proteins that interact with the intestinal mucosa[60]. More strikingly than is evident for bifidobacteria,

223    this adaptation to life in the GIT is further evident when the genome sequences of intestinal isolates

224    are compared with food-adapted lactobacilli such as *L. bulgaricus* and *L. helveticus*. *L. bulgaricus*.

225    The latter, which is widely used as a starter culture in yogurt fermentations, has undergone genome

226    decay to adapt to the milk environment[53], and thus harbours numerous degraded or partial

227    carbohydrate pathways and harbours bile salt hydrolase pseudogenes[53, 60]. In addition, *L. bulgaricus*

228    shows a preference for growth in lactose, further emphasizing its niche adaptation to milk. The

229    genome sequence of *L. helveticus*, a widely used cheese starter culture, has been reported recently[52].

230    Compared to the closely related *L. acidophilus*, *L. helveticus* has additional genes for fatty acid

231    biosynthesis and specific amino acid metabolism, but notably fewer cell surface proteins and PEP-

232    PTS systems for sugar utilization[52]. Additionally, no functional mucus binding proteins or

233    transporters for complex carbohydrates such as raffinose and fructooligosaccharides are encoded by

234    the *L. helveticus* genome, reflecting the degree of adaptation of *L. helveticus* to a milk environment.

235    In contrast, *L. acidophilus* has adapted to the gut ecological niche by retaining the functional gene

236    sets lacking in *L. helveticus*, emphasizing their importance for probiotic functionality and niche

237    adaptation by autochthonous lactobacilli naturally residing in the GIT.

238    Several studies have examined commensal *Lactobacillus* gene expression in animal model systems.

239    Using a stringent lincomycin-resistance based selection, Walter and colleagues identified

240    surprisingly only three genes that were differentially expressed *in vivo* [69]. Bron *et al.* [70] used a

241    modified *in vivo* expression technology to identify 72 genes expressed by *L. plantarum* in the

242    mouse GIT, most of which were associated with carbon metabolism, amino acid metabolism, and

243    stress resistance[70], and many of which were functions previously identified as survival/adaptation

244    factors in pathogens. *L. casei* actively transcribes metabolic genes in the murine intestine, and

245    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

271 acids[63]. The metabolism of the administered probiotics, coupled with competition for substrates and

272 small molecules, are the likely reasons for the transcriptional and metabolite alterations described in

273 these studies.

274 Numerous studies have reported that administration of probiotics benefits a range of gastrointestinal

275 conditions and infections[64, 65], but mechanistic insights are generally lacking. Reduction in vaginal

276 *Lactobacillus* levels that leads to vaginosis has been linked to production of a bacteriocin-like

277 substance by commensal enterococci[66]. From the opposite perspective, the ability of *L. salivarius* to

278 eliminate *Listeria monocytogenes* in a mouse model was dependent on production of the broad

279 spectrum bacteriocin Abp118/salivaricin[67], and bacteriocin-producing lactobacilli become dominant

280 among strains in a cocktail that reduce *Salmonella* shedding in pigs[68]. Thus bacteriocin production

281 is likely an important general mechanism in the interaction of many lactobacilli and other

282 commensals.

283

284 **Comparative genomics of *Lactobacillus*.**

285 Sequencing of the genomes of twenty lactic acid bacteria (LAB) has demonstrated that loss and

286 decay of ancestral genes has played a key role in the evolution of *Lactobacillales*. *Lactobacillales*

287 diverged from their *Bacillus* ancestor with an estimated loss of 600-1200 genes from a total gene

288 repertoire of 2,100 to 2,200[50]. Many of these genes encoded biosynthetic enzymes or functioned in

289 the sporulation process[50]. However, in addition to major gene losses, gene gains also occurred

290 which appear to reflect the nutrient-rich niches occupied by the LAB, such as milk and the GIT. For

291 example, genes encoding for peptidases, amino acid transport proteins and genes involved in the

292 metabolism and transport of carbohydrates have been duplicated[50]. In addition, comparative

293 analysis between GIT-associated species *L. acidophilus*, *L. gasseri*, and *L. johnsonii* and the dairy

294 species *L. bulgaricus* and *L. helveticus* revealed selective pressure from niche-specific adaptation on

295 the genome evolution of these species[51-53].

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 afforded by its larger 3 Mb genome, and a low-GC-content region coding for sugar transport and metabolism genes which is likely to have been acquired by HGT[54]. 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[52, 55]. Moreover, production of reuterin (3-hydroxypropionaldehyde), a potent broad-spectrum antimicrobial compound[56], is encoded by a genomic island which is present in some *L. reuteri* strains[57-59], and absent in the sequenced genome of a mouse *L. reuteri* isolate[58] and the closely related *L. fermentum*[59].

With genomes of 12 of the 147 recognized species[74] now fully sequenced, *Lactobacillus* has been targeted for several comparative whole-genome analyses. Beginning with the report of extreme diversity between the first two available genomes[34], genome sequencing of *L. acidophilus*, *L. gasseri*, *L. delbrueckii* and *L. helveticus* allowed a more focused attention on the 'acidophilus complex'[25, 52, 75]. Large regions of synteny were observed between the species[25, 52]. Multi-locus sequence analysis of five housekeeping genes, comparative-genome hybridizations and DNA-typing showed consistent and stepwise-decreasing levels of similarity within the group, suggesting a strong role for vertical evolution[25]. Conversely, differences between trees from 16S rRNA genes and 401 core genes from *L. acidophilus*, *L. johnsonii* and *L. delbrueckii* indicated a much higher level (40%) of HGT[75].

In order 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 five lactobacilli underscored the substantial diversification of the genus, and suggested a subgeneric division into three groups [21]. Furthermore, two overlapping comparative studies, encompassing nine additional *Lactobacillales* genomes, saw the expansion of the gene core to 567 order-specific genes[50, 76]. Similarly, the majority of these encoded information-processing proteins. The finer granularity provided by

322 LaCOGs (*Lactobacillales*-specific COGs) allowed detection of two genes, whose gene-contexts

323 suggest housekeeping and protein-modification functions. Recently, we extracted 141 core genes

324 from 12 *Lactobacillus* genomes to investigate the case for a single congruent genus phylogeny[22].

325 Although this proved impossible at the time, four sub-generic groups were reliably distinguished.

326 These were operationally characterized by absent genes rather than gained/retained genes,

327 consistent with the findings of an earlier study[76].

328

329 **Common evolutionary trends in probiotic genomes**

330 Collective analyses of probiotic genome sequences so far available — the probiome — has revealed

331 some generally conserved genetic traits[22, 25, 52, 54, 55, 59, 76], which may reflect adaptation to the

332 intestinal niche [1]. However, since probiotic bacteria represent diverse and taxonomically

333 heterogeneous groups of microorganisms, the analysis of phyletic (phylogenetic) patterns, i.e.

334 patterns of gene presence/absence in a particular set of genomes, may be overwhelmingly

335 influenced by the evolutionary distance between these two distant phyla. Nevertheless, common

336 trends in the evolution of both *Bifidobacterium* and *Lactobacillus* genomes may be discerned. These

337 include gene loss (e.g. of genes encoding biosynthetic enzymes), gene duplication and HGT. The

338 adaptation of probiotic bacteria to successfully exist and compete in the human gut must have been

339 driven by the occurrence of DNA duplications and genetic acquisitions during their evolution.

340 Many genes involved in sugar metabolism and transport were duplicated or acquired early in the

341 evolution of probiotic bacteria, including those encoding enolase, β-galactosidase, and many other

342 GH[50]. In addition, expansion of peptidases and amino acid transporters has occurred in several

343 lineages of *Lactobacillales* and bifidobacteria. Furthermore, several expanded families include

344 proteins involved in antibiotic resistance in other bacteria, i.e. β-lactamases[77].

345 Horizontal gene transfer via bacteriophage-mediated or conjugative pathways has been extensively

346 documented in *Lactobacillales* and appears to be important for niche-specific adaptation in

347 probiotic bacteria. In probiotic lactobacilli, HGT played an important role in shaping the common

348  ancestor, in which 84 genes were inferred to be horizontally transferred from different sources[50]. In

349  some cases, the ancestor acquired an additional pseudoparalogous copy of a gene by HGT (e.g.

350  enolase in *Lactobacillales*) while on other occasions, xenologous displacement, i.e., acquisition of

351  genes by HGT followed by the loss of the ancestral orthologous gene[78] apparently took place.

352  A provocative future challenge will involve the identification of the hypothetical core

353  probiogenome, representing core genome functions of probiotic bacteria. However, only seven

354  genes present in the bifidobacteria but not in the genomes of the other members of the

355  *Actinobacteria* phylum are shared with *Lactobacillales*. Only one of these genes, which encodes a

356  functionally uncharacterized membrane protein, is present in all the *Lactobacillales* genomes so far

357  sequenced [50].

358

359  **Conclusions and future considerations**

360  Most of the probiotic bacteria marketed today were originally selected on the basis of technological

361  stability or by a variety of easily measurable phenotypes such as ability to tolerate bile salts or

362  survive GIT passage, but not necessarily for their ability to promote health benefits. It is crucial to

363  identify the precise mechanisms by which such probiotic microorganisms influence human health.

364  Such studies should be accelerated by omics approaches involving genomics and functional

365  analyses. Molecular interaction models are being currently developed, although more are required,

366  that monitor the activation of cellular and systemic responses *in vivo* in animal models and in

367  feeding trial participants through the measurement of previously validated biomarkers. The

368  combination of verified molecular models with functional and comparative genomics-based

369  approaches should enable selection of the most appropriate probiotic strain for a particular health

370  benefit or improvement of strain processing and administration regimes that optimize the

371  established health effect. Finally, this might allow the selection of specific probiotics for a particular

372  human genotype, in analogy to personalized genomic medicine efforts.

373  Several issues regarding the sequences of complete probiotic bacterial genomes remain unresolved

374  at present. So far, only a limited number of completed probiotic bacterial genome sequences are

375  available, which only partially represent the total biodiversity of probiotic bacteria residing in the

376  human gut. In this context, understanding of the human gut microbiome will be an important

377  challenge for the future[79]. Furthermore, sequencing the genomes of environmental organisms and

378  carrying out metagenomic surveys of diverse gut environments (human vs. animal GIT) will

379  provide not only an improved understanding of microbial biodiversity but also insights into the

380  evolution of bacterial factors  that may be crucial for the commensals (probiotics) establishment in

381  these different gut niches[80].

382  The first decade of bacterial genomics has afforded unprecedented insights into the evolution of

383  bacterial pathogens (bacterial pathogenomics)[81]. The next decade holds the promise of being even

384  more rewarding as the new discoveries about probiotic bacteria provided by probiogenomic efforts

385  are exploited.

386

387  **GLOSSARY**

388  Omics: The integration of genomics methodology and data with functional genomic analyses

389  involving transcriptomics, proteomics, metabolomics and interactomics.

390  Microbiota: The collective microbial community or population resident in a particular locale at a

391  given time-point.

392  Microbiome: The collective genome of the human microbial communities

393  Prebiotics: Growth substrates that are preferentially (or ideally, exclusively) metabolized by a single

394  genus or species, and that may thus be used as dietary supplements to promote growth of a targeted

395  microorganism.

396  Transcriptome: Subsets of genes transcribed in an organism. It represents dynamic links between

397  genomes, proteins and cellular phenotypes.

398  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.

412

**References**

414 1.    Backhed, F., Ley, R. E., Sonnenburg, J. L., Peterson, D. A. & Gordon, J. I. Host-bacterial
415       mutualism in the human intestine. Science 307, 1915-20 (2005).
416 **2.    Eckburg, P. B. et al. Diversity of the human intestinal microbial flora. Science 308,**
417       **1635-8 (2005).**
418 **This article describes the bacterial diversity occurring in the human gut using 16S rRNA gene**
419       **based libraries.**
420 3.    Seksik, P. et al. Alterations of the dominant faecal bacterial groups in patients with Crohn's
421       disease of the colon. Gut 52, 237-42 (2003).
422 4.    Turroni, F., Ribbera, A., Foroni, E., van Sinderen, D. & Ventura, M. Human gut microbiota
423       and bifidobacteria: from composition to functionality. Antonie Van Leeuwenhoek 94, 35-50
424       (2008).
425 **5.    Rajilic-Stojanovic, M., Smidt, H. & de Vos, W. M. Diversity of the human**
426       **gastrointestinal tract microbiota revisited. Environ. Microbiol. 9, 2125-36 (2007).**
427 **This review provides an integrated summary of data from culture independent studies of the**
428       **human gut microbiota**.
429 6.    Ley, R. E., Peterson, D. A. & Gordon, J. I. Ecological and evolutionary forces shaping
430       microbial diversity in the human intestine. Cell 124, 837-48 (2006).
431 7.    Guarner, F. & Malagelada, J. R. Gut flora in health and disease. Lancet 361, 512-9 (2003).
432 8.    Hooper, L. V. & Gordon, J. I. Commensal host-bacterial relationships in the gut. Science
433       292, 1115-8 (2001).
434 9.    Backhed, F. et al. The gut microbiota as an environmental factor that regulates fat storage.
435       Proc Natl Acad Sci U S A 101, 15718-23 (2004).
436 10.   Samuel, B. S. & Gordon, J. I. A humanized gnotobiotic mouse model of host-archaeal-
437       bacterial mutualism. Proc Natl Acad Sci U S A 103, 10011-6 (2006).
438 11.   Turnbaugh, P. J. et al. An obesity-associated gut microbiome with increased capacity for
439       energy harvest. Nature 444, 1027-31 (2006).
440 12.   Frank, D. N. et al. Molecular-phylogenetic characterization of microbial community
441       imbalances in human inflammatory bowel diseases. Proc Natl Acad Sci U S A 104, 13780-5
442       (2007).
443 13.   Kassinen, A. et al. The fecal microbiota of irritable bowel syndrome patients differs
444       significantly from that of healthy subjects. Gastroenterology 133, 24-33 (2007).
445 **14.  Manichanh, C. et al. Reduced diversity of faecal microbiota in Crohn's disease**
446       **revealed by a metagenomic approach. Gut 55, 205-11 (2006)**.
447 **The preceding two references provide evidence for significant microbiota alterations in**
448       **functional bowel disorders.**
449 15.   Salminen, S., Nurmi, J. & Gueimonde, M. The genomics of probiotic intestinal
450       microorganisms. Genome Biol 6, 225 (2005).
451 16.   Marco, M. L., Pavan, S. & Kleerebezem, M. Towards understanding molecular modes of
452       probiotic action. Curr Opin Biotechnol 17, 204-10 (2006).
453 17.   O'Hara, A. M. & Shanahan, F. Mechanisms of action of probiotics in intestinal diseases.
454       ScientificWorldJournal 7, 31-46 (2007).
455 18.   Saxelin, M., Tynkkynen, S., Mattila-Sandholm, T. & de Vos, W. M. Probiotic and other
456       functional microbes: from markets to mechanisms. Curr Opin Biotechnol 16, 204-11 (2005).
457 19.   Ventura, M. et al. Genomics of Actinobacteria: tracing the evolutionary history of an ancient
458       phylum. Microbiol Mol Biol Rev 71, 495-548 (2007).
459 20.   Joyce, A. R. & Palsson, B. O. The model organism as a system: integrating 'omics' data sets.
460       Nat Rev Mol Cell Biol 7, 198-210 (2006).

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

510 39. Katayama, T. et al. Molecular cloning and characterization of Bifidobacterium bifidum 1,2-
511     alpha-L-fucosidase (AfcA), a novel inverting glycosidase (glycoside hydrolase family 95). J
512     Bacteriol 186, 4885-93 (2004).
513 40. Ryan, S. M., Fitzgerald, G. F. & van Sinderen, D. Transcriptional regulation and
514     characterization of a novel beta-fructofuranosidase-encoding gene from Bifidobacterium
515     breve UCC2003. Appl Environ Microbiol 71, 3475-82 (2005).
516 41. Gonzalez, R., Klaassens, E. S., Malinen, E., de Vos, W. M. & Vaughan, E. E. Differential
517     transcriptional response of Bifidobacterium longum to human milk, formula milk and
518     galactooligosaccharide. Appl Environ Microbiol (2008).
519 42. Liepke, C. et al. Human milk provides peptides highly stimulating the growth of
520     bifidobacteria. Eur J Biochem 269, 712-8 (2002).
521 43. Ivanov, D. et al. A serpin from the gut bacterium Bifidobacterium longum inhibits
522     eukaryotic elastase-like serine proteases. J Biol Chem 281, 17246-52 (2006).
523 44. Potempa, J., Korzus, E. & Travis, J. The serpin superfamily of proteinase inhibitors:
524     structure, function, and regulation. J Biol Chem 269, 15957-60 (1994).
525 **45. Sonnenburg, J. L., Chen, C. T. & Gordon, J. I. Genomic and metabolic studies of the**
526     **impact of probiotics on a model gut symbiont and host. PLoS Biol 4, e413 (2006).**
527 **This paper describes the crosstalk existing between bifidobacteria and *Bacteroides* in the**
528     **murine intestine as well as between these bacteria and their host.**
529 46. Kato, S., Haruta, S., Cui, Z. J., Ishii, M. & Igarashi, Y. Stable coexistence of five bacterial
530     strains as a cellulose-degrading community. Appl Environ Microbiol 71, 7099-106 (2005).
531 47. Barrangou, R. et al. CRISPR provides acquired resistance against viruses in prokaryotes.
532     Science 315, 1709-12 (2007).
533 48. Klijn, A., Mercenier, A. & Arigoni, F. Lessons from the genomes of bifidobacteria. FEMS
534     Microbiol Rev 29, 491-509 (2005).
535 49. Aas, J. A. et al. Bacteria of dental caries in primary and permanent teeth in children and
536     young adults. J Clin Microbiol 46, 1407-17 (2008).
537 50. Makarova, K. S. & Koonin, E. V. Evolutionary genomics of lactic acid bacteria. J Bacteriol
538     189, 1199-208 (2007).
539 51. Makarova, K. et al. Comparative genomics of the lactic acid bacteria. Proc. Natl. Acad. Sci.
540     U S A 103, 15611-6 (2006).
541 **This landmark study provided a large tranche of genomic data to allow studies of genome**
542     **evolution in lactic acid bacteria.**
543 52. Altermann, E. et al. Complete genome sequence of the probiotic lactic acid bacterium
544     Lactobacillus acidophilus NCFM. Proc Natl Acad Sci U S A 102, 3906-12 (2005).
545 53. van de Guchte, M. et al. The complete genome sequence of Lactobacillus bulgaricus reveals
546     extensive and ongoing reductive evolution. Proc. Natl. Acad. Sci. U S A 103, 9274-9
547     (2006).
548 **54. Kleerebezem, M. et al. Complete genome sequence of Lactobacillus plantarum**
549     **WCFS1. Proc Natl Acad Sci U S A 100, 1990-5 (2003).**
550 **This is the first article describing the genome sequence of a member of the genus**
551     ***Lactobacillus*.**
552 **55. Pridmore, R. D. et al. The genome sequence of the probiotic intestinal bacterium**
553     **Lactobacillus johnsonii NCC 533. Proc Natl Acad Sci U S A 101, 2512-7 (2004).**
554 **This is paper describes the genome contents of a common used probiotic bacterium belonging**
555     **to the genus *Lactobacillus*.**
556 56. Talarico, T. L., Casas, I. A., Chung, T. C. & Dobrogosz, W. J. Production and isolation of
557     reuterin, a growth inhibitor produced by *Lactobacillus reuteri*. Antimicrob. Agents
558     Chemother. 32, 1854-8 (1988).
559 57. Santos, F. et al. The complete coenzyme B12 biosynthesis gene cluster of *Lactobacillus*
560     *reuteri* CRL1098. Microbiology 154, 81-93 (2008).

561 58. Sriramulu, D. D. et al. Lactobacillus reuteri DSM 20016 produces cobalamin-dependent diol
562     dehydratase in metabolosomes and metabolizes 1,2-propanediol by disproportionation. J.
563     Bacteriol. 190, 4559-67 (2008).
564 59. Morita, H. et al. Comparative Genome Analysis of Lactobacillus reuteri and Lactobacillus
565     fermentum Reveal a Genomic Island for Reuterin and Cobalamin Production. DNA Res 15,
566     151-61 (2008).
567 60. Pfeiler, E. A. & Klaenhammer, T. R. The genomics of lactic acid bacteria. Trends Microbiol
568     15, 546-53 (2007).
569 61. Whitman, W. B., Coleman, D. C. & Wiebe, W. J. Prokaryotes: the unseen majority. Proc.
570     Natl. Acad. Sci. U S A 95, 6578-83 (1998).
571 62. Tannock, G. W. et al. Analysis of the fecal microflora of human subjects consuming a
572     probiotic product containing *Lactobacillus rhamnosus* DR20. Appl. Environ. Microbiol. 66,
573     2578-88 (2000).
574 63. Martin, F. P. et al. Probiotic modulation of symbiotic gut microbial-host metabolic
575     interactions in a humanized microbiome mouse model. Mol. Syst. Biol. 4, 157 (2008).
576 64. Hickson, M. et al. Use of probiotic *Lactobacillus* preparation to prevent diarrhoea associated
577     with antibiotics: randomised double blind placebo controlled trial. Brit. Med. J. 335, 80
578     (2007).
579 65. Sullivan, A. & Nord, C. E. Probiotics and gastrointestinal diseases. J. Intern. Med. 257, 78-
580     92 (2005).
581 66. Kelly, M. C., Mequio, M. J. & Pybus, V. Inhibition of vaginal lactobacilli by a bacteriocin-
582     like inhibitor produced by Enterococcus faecium 62-6: potential significance for bacterial
583     vaginosis. Infect. Dis. Obstet. Gynecol. 11, 147-56 (2003).
584 **67. Corr, S. C. et al. Bacteriocin production as a mechanism for the antiinfective activity**
585     **of *Lactobacillus salivarius* UCC118. Proc. Natl. Acad. Sci. U S A 104, 7617-7621 (2007).**
586 **This study identified the first molecular mechanism whereby probiotic bacteria modulate the**
587     **microbiota *in vivo*.**
588 68. Casey, P. G. et al. A five-strain probiotic combination reduces pathogen shedding and
589     alleviates disease signs in pigs challenged with *Salmonella enterica* Serovar Typhimurium.
590     Appl. Environ. Microbiol. 73, 1858-63 (2007).
591 **69. Walter, J. et al. Identification of *Lactobacillus reuteri* genes specifically induced in the**
592     **mouse gastrointestinal tract. Appl. Environ. Microbiol. 69, 2044-51 (2003).**
593 **70. Bron, P. A., Grangette, C., Mercenier, A., de Vos, W. M. & Kleerebezem, M.**
594     **Identification of Lactobacillus plantarum genes that are induced in the gastrointestinal**
595     **tract of mice. J Bacteriol 186, 5721-9 (2004).**
596 **This manuscript provides an insight into the molecular interactions between a commensal**
597     **microorganism and a murine-model host.**
598 71. Oozeer, R. et al. Differential activities of four *Lactobacillus casei* promoters during bacterial
599     transit through the gastrointestinal tracts of human-microbiota-associated mice. Appl.
600     Environ. Microbiol. 71, 1356-63 (2005).
601 72. Denou, E. et al. Gene expression of commensal *Lactobacillus johnsonii* strain NCC533
602     during in vitro growth and in the murine gut. J. Bacteriol. 189, 8109-19 (2007).
603 73. Denou, E. et al. Identification of genes associated with the long-gut-persistence phenotype
604     of the probiotic *Lactobacillus johnsonii* strain NCC533 using a combination of genomics
605     and transcriptome analysis. J. Bacteriol. 190, 3161-8 (2008).
606 74. Euzèby, J. P. (2007). List of Bacterial names with Standing in Nomenclature. Int J Syst
607     Bacteriol. 47, 590-592 (1997).
608 75. Nicolas, P., Bessieres, P., Ehrlich, S. D., Maguin, E. & van de Guchte, M. Extensive
609     horizontal transfer of core genome genes between two *Lactobacillus* species found in the
610     gastrointestinal tract. BMC Evol. Biol. 7, 141 (2007).

611 76. Makarova, K. et al. Comparative genomics of the lactic acid bacteria. Proc Natl Acad Sci U
612 S A 103, 15611-6 (2006).
613 77. Teuber, M., Meile, L. & Schwarz, F. Acquired antibiotic resistance in lactic acid bacteria
614 from food. Antonie Van Leeuwenhoek 76, 115-37 (1999).
615 78. Koonin, E. V., Makarova, K. S. & Aravind, L. Horizontal gene transfer in prokaryotes:
616 quantification and classification. Annu Rev Microbiol 55, 709-42 (2001).
617 79. Turnbaugh, P. J. et al. The human microbiome project. Nature 449, 804-10 (2007).
618 80. **Ley, R. E. et al. Evolution of mammals and their gut microbes. Science 320, 1647-51**
619 **(2008).**
620 **This paper describes the bacterial diversity present in the gut of numerous mammals.**
621 81. Pallen M.J., & Wren, B.M. Bacterial pathogenomics. Nature 449, 835-842 (2007).
622
623 82. Lee, J. H. et al. Comparative genomic analysis of the gut bacterium Bifidobacterium longum
624 reveals loci susceptible to deletion during pure culture growth. BMC Genomics 9, 247
625 (2008).
626 83. Leahy, S. C., Higgins, D. G., Fitzgerald, G. F. & van Sinderen, D. Getting better with
627 bifidobacteria. J Appl Microbiol 98, 1303-15 (2005).
628
629
630
631
632
633
634
635
636
637

**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. *lactis* 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.
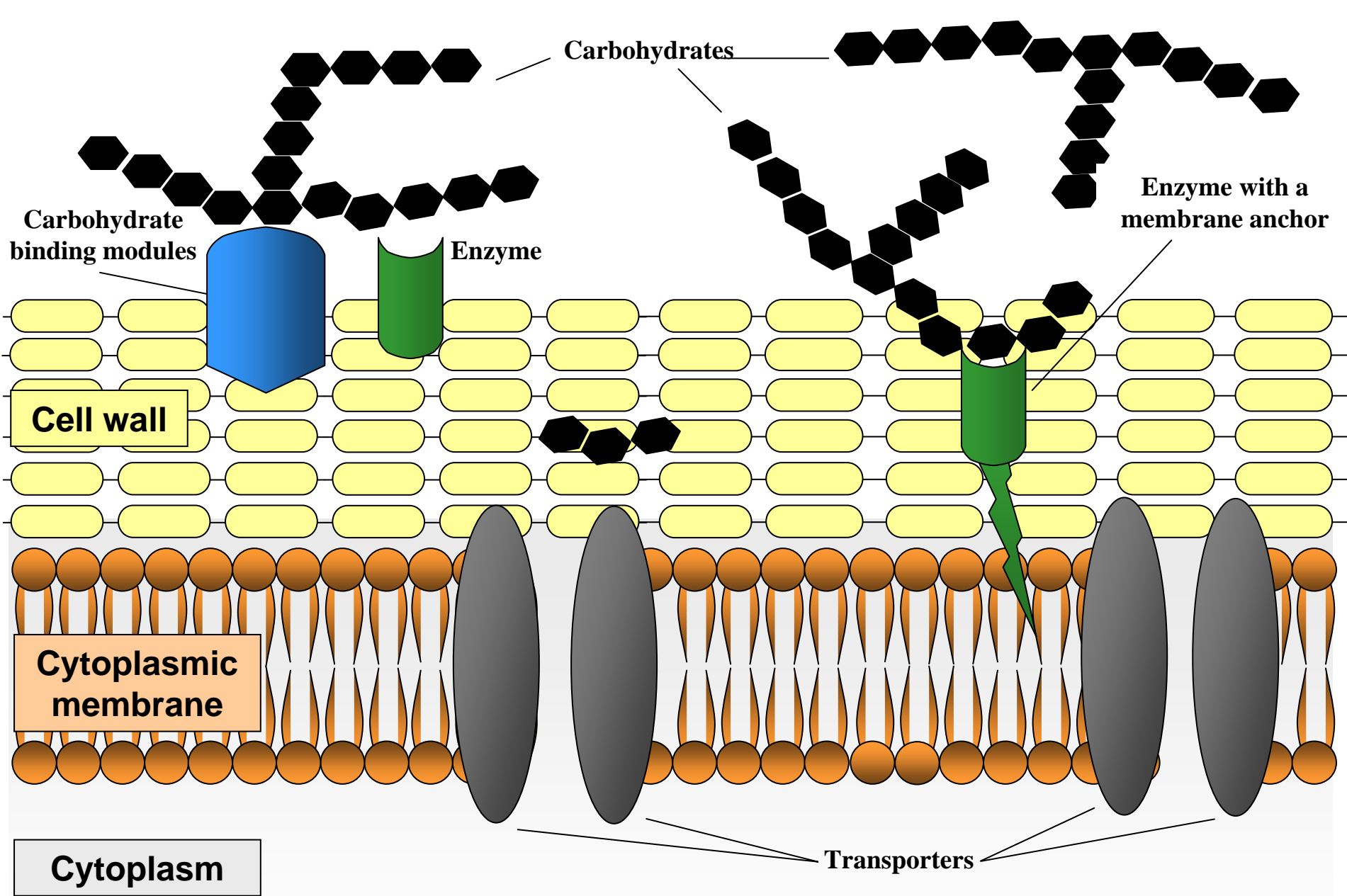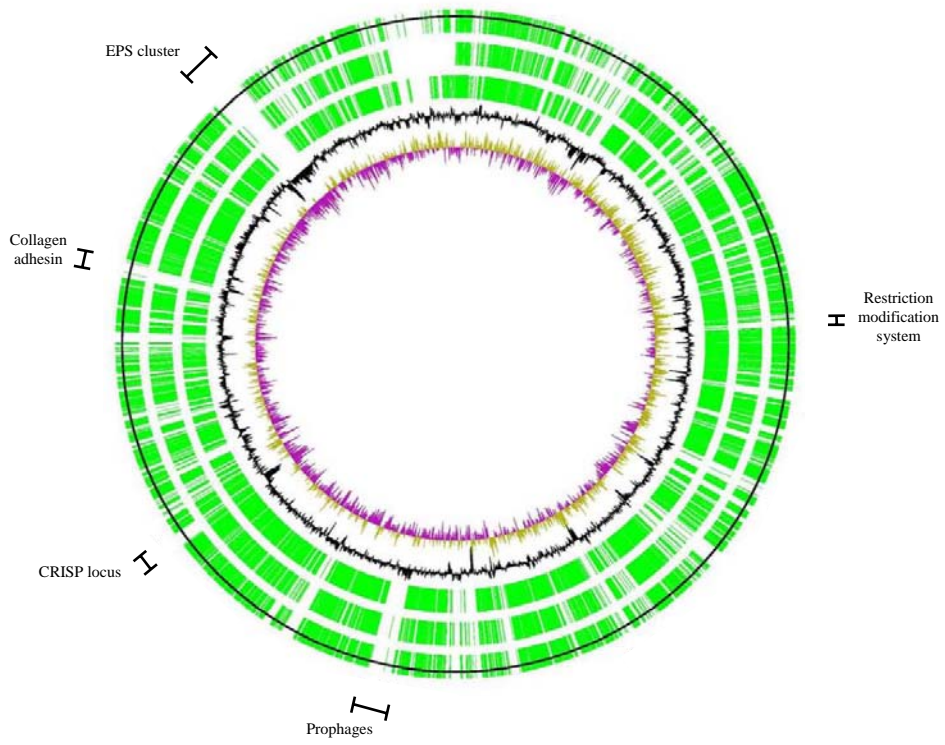
**a)**

Host → Bacteria → Pathogenicity ⟷ Mutualism

Commensalism | Symbiosis

**b)**

**c)**

**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

Restriction
modification
system

CRISP locus

Prophages

b)



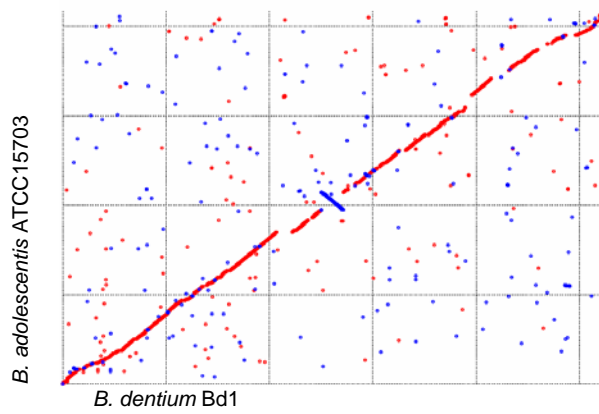*B. adolescentis* ATCC15703

*B. dentium* Bd1

**Figure 3**

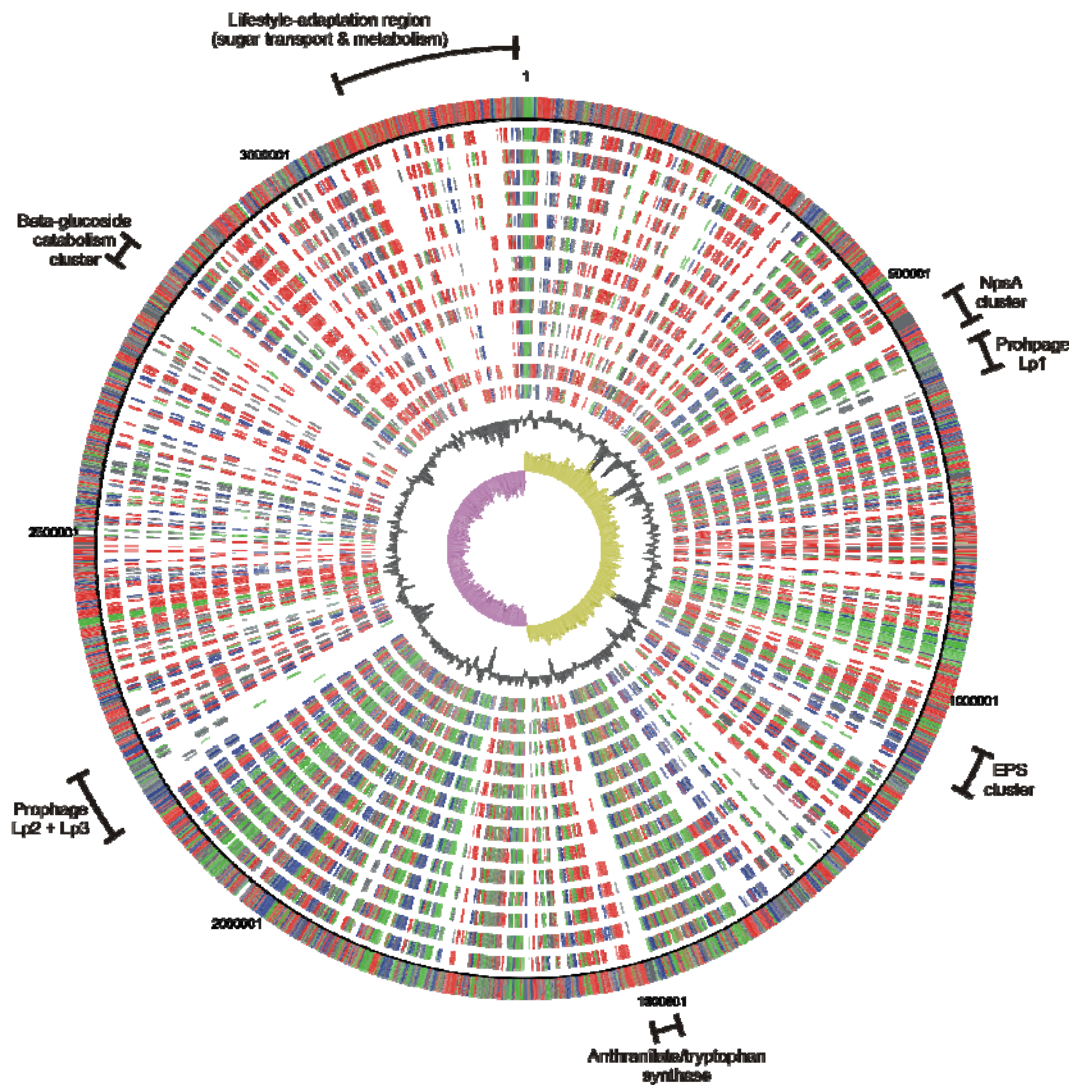Figure 4

**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. *lactis* 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] |