



UNIVERSITÀ DI PARMA

UNIVERSITÀ DEGLI STUDI DI PARMA

Ph.D. in Biotechnology and Life Sciences

XXXI cycle

Application of metagenomic approaches in discovering the gut microbiota functionality

Ph.D. Coordinator: Prof. Simone Ottonello

Tutor: Prof. Marco Ventura

Industrial Tutor: Prof. Francesca Turrone

Ph.D. student: **Marta Mangifesta**

2016-2018

Table of Contents

Summary	5
Chapter 1: General Introduction	7
A. The gut Microbiota: a complex ecosystem	8
Development of the gut microbiota.....	9
B. Implication of microbiota on host's health.....	11
Human health and microbiota.	11
The microbiota of animals.....	12
C. Correlation between microbiota and disease	14
Gut microbiota in Inflammatory Bowel disease (IBD).....	15
Colorectal Cancer.....	17
Modulation of the microbiota: prebiotics.....	18
D. Gut microbiota and microbial biomarkers	20
Bifidobacteria as a marker for health	20
<i>Lactobacillus</i> genus as a microbial biomarker	22
E. Technical approaches for microbiota determination.....	23
Methodologies for the investigation of the microbiota composition	25
Approaches to complete a high-definition image of the gut microbiota composition	26
Culturomics approaches	27
Chapter 2: Outline of the thesis	29
Chapter 3: Unveiling bifidobacterial biogeography across the mammalian branch of the tree of life	33
Chapter 4: <i>Bifidobacterium vansinderenii</i> sp. nov., isolated from faeces of emperor tamarin (<i>Saguinus imperator</i>)	55
Chapter 5: Phylogenetic classification of six novel species belonging to the genus <i>Bifidobacterium</i> comprising <i>Bifidobacterium anseris</i> sp. nov., <i>Bifidobacterium criceti</i> sp. nov., <i>Bifidobacterium imperatoris</i> sp. nov., <i>Bifidobacterium italicum</i> sp. nov., <i>Bifidobacterium margollesii</i> sp. nov. and <i>Bifidobacterium parmae</i> sp. nov.	73
Chapter 6: The Sortase-Dependent Fimbriome of the Genus <i>Bifidobacterium</i>: Extracellular Structures with Potential To Modulate Microbe-Host Dialogue	97
Chapter 7: Mucosal microbiota of intestinal polyps reveals putative biomarkers of colorectal cancer	123
Chapter 8: Compositional assessment of bacterial communities in probiotic supplements by means of metagenomic techniques.....	141
Chapter 9: Phylotype-Level Profiling of Lactobacilli in Highly Complex Environments by Means of an Internal Transcribed Spacer-Based Metagenomic Approach	161

Chapter 10: General Conclusion	183
Advances in Next generation sequencing analyses: unveiling the biodiversity of complex microbial communities	184
Identification of novel microbial biomarkers	186
References	187
Publications	195

Summary

In the last years, the advent of the Next-Generation Sequencing (NGS) technologies allowed to unveil the composition and functionality of the microbial populations inhabiting humans and other animals, i.e. microbiota. In the human context, it has been estimated that 10^{14} microorganisms reside in various compartments of the body such as the gastrointestinal, genitourinary and respiratory tracts and the skin surfaces. Among these, the human intestine retains a complex microbial ecosystem, that has co-evolved through the decades to form a mutually beneficial relationship with the host. In this regard, the range of studies focusing on this topic suggested that the gut microbiota exploits crucial functions largely responsible for the healthy status of the host. The gut microbiota can promote, through various mechanisms, intestinal metabolism and immune system modulation. Conversely, several factors, such as the environment and the lifestyle, may cause an upset in the composition of the microbiota, and thus causing dysbiosis.

Gut dysbiosis is typically related to a negative impact on host health, and it is associated with a large array of human disorders and disease, such as Inflammatory Bowel Disease (IBD), metabolic syndrome, neurological disorders and colorectal cancer (CRC).

The aim of this Ph.D. thesis is to explore the role of intestinal bacteria in human and animal health by means of several metagenomics approaches. In particular, it aims to investigate bifidobacterial compositional differences in various mammalian species (including humans), through a metagenomic analysis, based on the bioinformatics sequencing and analysis of the Internal Transcribed Spacer (ITS) sequences for bifidobacterial profiling (Bifidobacterial-ITS profiling). In addition, it discusses the role of bifidobacteria in the gastrointestinal tract of mammals and how these microorganisms can persist and colonize the human gut.

Furthermore, it provides an in depth overview of the mucosa microbiota composition associated with gastrointestinal polyps. This PhD thesis also evaluates the correlation between adenomatous polyps and publicly available metagenomics datasets of CRC, in order to identify possible microbial biomarkers associated with an early detection of CRC.

Despite the extensive commercial exploitation of probiotic formulations, there are still major knowledge gaps regarding the precise molecular mechanism of action and corresponding genetic/genomic properties of probiotic bacteria. In this context, this PhD thesis aims to unveil microbiota composition of different probiotic products through the development of an industrial pipeline encompassing 16S rRNA-microbial analysis, combined with a shotgun metagenomics approach. Thus, in order to confirm the obtained *in silico* metagenomic results, a culturomic approach

was used to perform a census of the bacterial populations of all assayed probiotics products. Finally, the purpose of this PhD thesis is also to provide a new metagenomic tool for species-level profiling of complex lactobacilli communities, which complements phylogenetic group assignments that can be obtained from 16S rRNA gene profiling data.

Chapter 1

General Introduction

A. The gut microbiota: a complex ecosystem

The human and animal body encompasses a complex and dynamic population of microorganisms, collectively referred to as the microbiota. In the human context, the majority of these microorganisms reside within the gastrointestinal tract, especially in the distal intestine, and it has been estimated to exceed ~10 times more the total numbers of human somatic and germ cells (Gill et al., 2006).

The gut microbiota can be described as an additional organ of mammalian's body in which the enormous amount of microorganisms possesses the capability to communicate with one to another and with the host. Some of these microorganisms are natural residents of the gastrointestinal tract (i.e., autochthonous microorganisms), whereas others reach the intestine from food ingestion and environmental contaminations (i.e., transient or allochthonous bacteria) (Ventura et al., 2009).

The gut microbiota has co-evolved through the decades to form a mutually beneficial relationship with the host (Thursby and Juge, 2017). It contributes to host health through biosynthesis of essential amino acids and vitamins. In addition, it exerts important physiological and metabolic activities for the host such as protection against pathogens, degradation of indigestible polysaccharides, production of energy for intestinal epithelial cells and strengthen the mucosal barrier [i.e., Short chain fatty acid (SCFA) byproducts like propionate, butyrate and acetate] (Topping and Clifton, 2001; Backhed et al., 2005). The gut microbiota can promote, through various mechanisms, epithelial cell maturation and angiogenesis, and it is fundamental in driving adaptive immunity in recognizing and responding to specific microorganisms or in programming many aspects of T cell differentiation. In this regard, many studies performed in germ-free mice (born and raised in the absence of any microorganism), provide important insights into how intestinal epithelial cells exhibit reduced expression of molecules involved in pathogen sensing and antigen presentation. (Lee and Mazmanian, 2010; Belkaid and Hand, 2014).

Development of the gut microbiota

The gastrointestinal tract (GIT) contains the highest cell densities compared to any other human body district (Sender et al., 2016) (Fig.1). In fact, the gut microbiota spans the tree of life and thus including bacteria, archaea, eukaryotes and also viruses (Lloyd-Price et al., 2016). Nonetheless, predominant bacterial groups in the microbiota are Gram positive Firmicutes (~51%) and Gram negative Bacteroidetes (~48%) (Eckburg et al., 2005; Singh et al., 2017), whereas Actinobacteria, Proteobacteria, and Verrucomicrobia are frequent but generally minor components. It seems that human intestine begins to be colonized by microorganisms after delivery, but it is worth mentioning that this dogma is changing thanks to the growing numbers of studies suggesting bacterial translocation from mother by placenta or umbilical cord (Jimenez et al., 2005; Aagaard et al., 2014). The newborn, after delivery is exposed to a wide range of microbes, many of which are provided by the mother if naturally delivered. The gut microbiota goes into dramatic changes until it reaches maturation after weaning and only a subset of the microbes to which the infant is initially exposed will permanently colonize the GIT and contribute to the distinctive microbiotas harbored by the gut of adults. Nevertheless, feeding methods, delivery mode and antibiotics have large effects on the infant microbiota (Turrone et al., 2012). In this context, diet plays a pivotal role in generating diversity and compositional changes in the microbiota in the early stage of host's life, which is characterized by the dominance of members of the *Bifidobacterium* genus. This colonization is stimulated by the presence of human milk oligosaccharides (HMOs), which is one the most abundant carbohydrate component of the breast milk. Notably, HMOs are not digested by the mammalian's enzyme and are typically hydrolyzed by few *Bifidobacterium* spp.. On the other hand, formula-fed infants possess higher proportions of *Bacteroides* and members of the *Clostridium* spp. and *Lactobacillus* spp. (Fallani et al., 2010; Bezirtzoglou et al., 2011). As mention before, infants born vaginally display a gut microbiota closer to that of their mother's, which is predicted to be maternally inherited through a vertical transmission by the birth canal. The gut microbiota of vaginally delivered babies is dominated by bifidobacteria (as above mentioned) and members of *Bacteroides* genera (Mitsou et al., 2008; Biasucci et al., 2010). On the contrary, Cesarean section (C-section) delivery upsets normal colonization of the infant gut with horizontally transmitted microbes through mother's skin and hospital environment. C-section infants possess a microbiota mostly dominated by *Staphylococcus*, *Corynebacteria*, and *Propionibacterium* spp. (Gronlund et al., 1999; Penders et al., 2006; Mitsou et al., 2008; Biasucci et al., 2010; Azad et al., 2013). These differences may have important implication on host's physiology and immunology since *Lactobacillus* and *Bifidobacterium* genera are considered to exploit health promoting activities, whereas some *Staphylococcus* spp. and *Clostridium* spp.

displayed possible pathogenic behaviors (Yang et al., 2016). Compositional differences driven by these factors in childhood, may contribute to disease predisposition, including asthma and atopic diseases into adulthood (Dominguez-Bello et al., 2010; Lozupone et al., 2012).

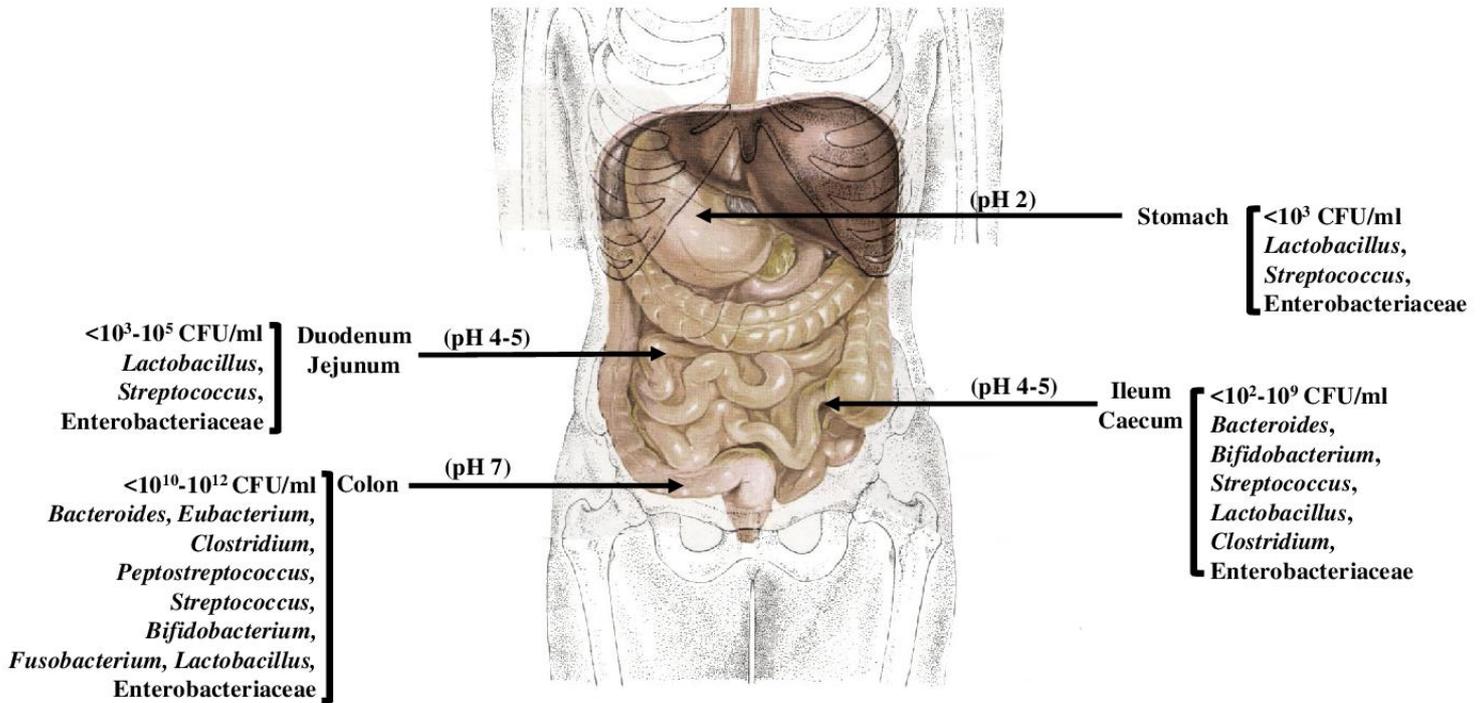


Figure 1: Schematic representation of the human gastro intestinal tract with the relative abundance of microorganisms between different compartments.

B. Implications of microbiota in the host's health

The gut microbiota plays a crucial role in host health during the host's entire life span. Perturbations in its composition during development and establishment in the first stages of life have been linked to pediatric disorders and the onset of disease in adulthood (Clemente et al., 2012). As mentioned above, the microbiota is involved in multiple interactions such as the protection against pathogens, the stimulation of the immune system, the promotion of host cells differentiation but it should be mentioned that also the host has an impact on the shaping of the microbiota. In fact, the mammalian GIT represents a selective environment, consisting of an assemblage of resources and conditions, allowing and/or supporting the growth of certain microorganisms with specific features for the survival and growth in the host (Milani et al., 2017). The composition of the gut microbiota is also influenced by the lifestyle, different ages, as well as physiologic states of the host and cultural/geographic locations. In this context, the geographical area could be responsible for the disappearance of low-abundance species (e.g., antibiotic sensitive strains) or for the acquisition of others during industrialization/urbanization (Mancabelli et al., 2017a).

Human health and the gut microbiota

The gut microbiota is a dynamic system and there is not a stable configuration of "healthy" microbiota. Very recently, the term "enterotypes", support the notion that the human microbiota forms three distinct host-microbial symbiotic conditions characterized by groups of co-occurring species/genera. These three enterotypes are driven by higher abundance of the genera *Bacteroides* (Enterotype I), *Prevotella* (Enterotype II), or *Ruminococcus* (Enterotype III), respectively. Arumugam et al., provides evidences on how the abundances of each of the three discriminating taxa strongly correlate with those of other genera, suggesting that the enterotypes are driven by groups of bacteria that contribute to the preferred community compositions, thus structuring the microbiota (Arumugam et al., 2011).

Notably, such different enterotypes are reflecting key functional roles exploited by their members such as *Prevotella* genus, associated with a high-fiber diet or *Bacteroides* members, that are abundant with a long-term diet rich in animal proteins, saturated fats and several amino acids.

In addition, the gut microbiota possesses important features linked to host health. It provides several compounds interacting with host epithelial cells, such as lipids and proteins. Notably, epithelial cells

have evolved separate mechanisms dedicated to sense the presence of different molecules, such as lipopolysaccharide (LPS), which is a principal component of Gram-negative cell surfaces and the peptidoglycan that is a major component of Gram-positive bacteria's cell wall. In this context, a higher abundance of Gram-negative bacteria is linked to several inflammatory and metabolic disease states (Abe et al., 2010; Milani et al., 2017).

The microbiota of animals

In the last years, there has been a growing number of studies involving the characterization of the bacterial composition of the GIT of non-human animals, particularly farm animals for commercial/agricultural purpose. Similarly to what described for humans, the gut microbiota of healthy animals is presumed to possess a pivotal role in the prevention of pathogens colonization and in the nutrients assimilation.

So far, pigs (*Sus scrofa domestica*), cattles (*Bos Taurus*) and chickens (*Gallus gallus domestica*) are the most studied animals because of the larger farmed populations in economically developed regions (Choi et al., 2015; Mao et al., 2015; Niu et al., 2015; Pourabedin and Zhao, 2015; Fecteau et al., 2016; Heinritz et al., 2016; Wang et al., 2016; Bessegatto et al., 2017) (Fig.2). The GI tract of these animals offers many niches for bacterial members, for example pigs harbor a complex gut microbiota, comprising Firmicutes and Proteobacteria in the ileum lumen whereas at the phylum-level profiles of the caecum and mid-colon include also Bacteroidetes and Spirochetes (Kim et al., 2012; Arnal et al., 2014; Looft et al., 2014).

Conversely, in dairy and beef cattle's gut microbiota, the genus *Prevotella* is the most abundant taxon typically related to the diet of the ruminant. These animals revealed to possess a predominance of Firmicutes and Bacteroidetes, followed by the Proteobacteria phylum (Gomez et al., 2017).

On the other hand, the most dominant bacteria in the chicken microbiota are lactobacilli, facultative and microaerophilic bacteria. Other abundant species belong to Clostridiaceae and Enterobacteriaceae families as well as *Enterococcus* and *Bifidobacterium* genus (van der Wielen et al., 2002; Gong et al., 2007; Deusch et al., 2015).

Recently, the microbiota of wildlife species has been analyzed including tasmanian devil (*Sarcophilus harrisii*), wild red pandas (*Ailurus fulgens*) or koala (*Phascolarctos cinereus*) (Fig.2). The Tasmanian devil possesses a higher abundance of Firmicutes, showing a high Firmicutes-Bacteroidetes ratio which appears to be a common feature of many carnivorous mammals (Cheng et al., 2015).

Besides, the comparison between captive red pandas and wild red panda revealed that the first ones are characterized by bacterial communities mostly dominated by Firmicutes, whereas wild red pandas

possess community members more evenly distributed amongst Firmicutes, Proteobacteria and Bacteroidetes (Kong et al., 2014).

Finally, the composition of koala's gastrointestinal tract unveils the main presence of Bacteroidetes, Firmicutes and Proteobacteria phyla, followed by Actinobacteria and Fusobacteria (Alfano et al., 2015; Bahrndorff et al., 2016).

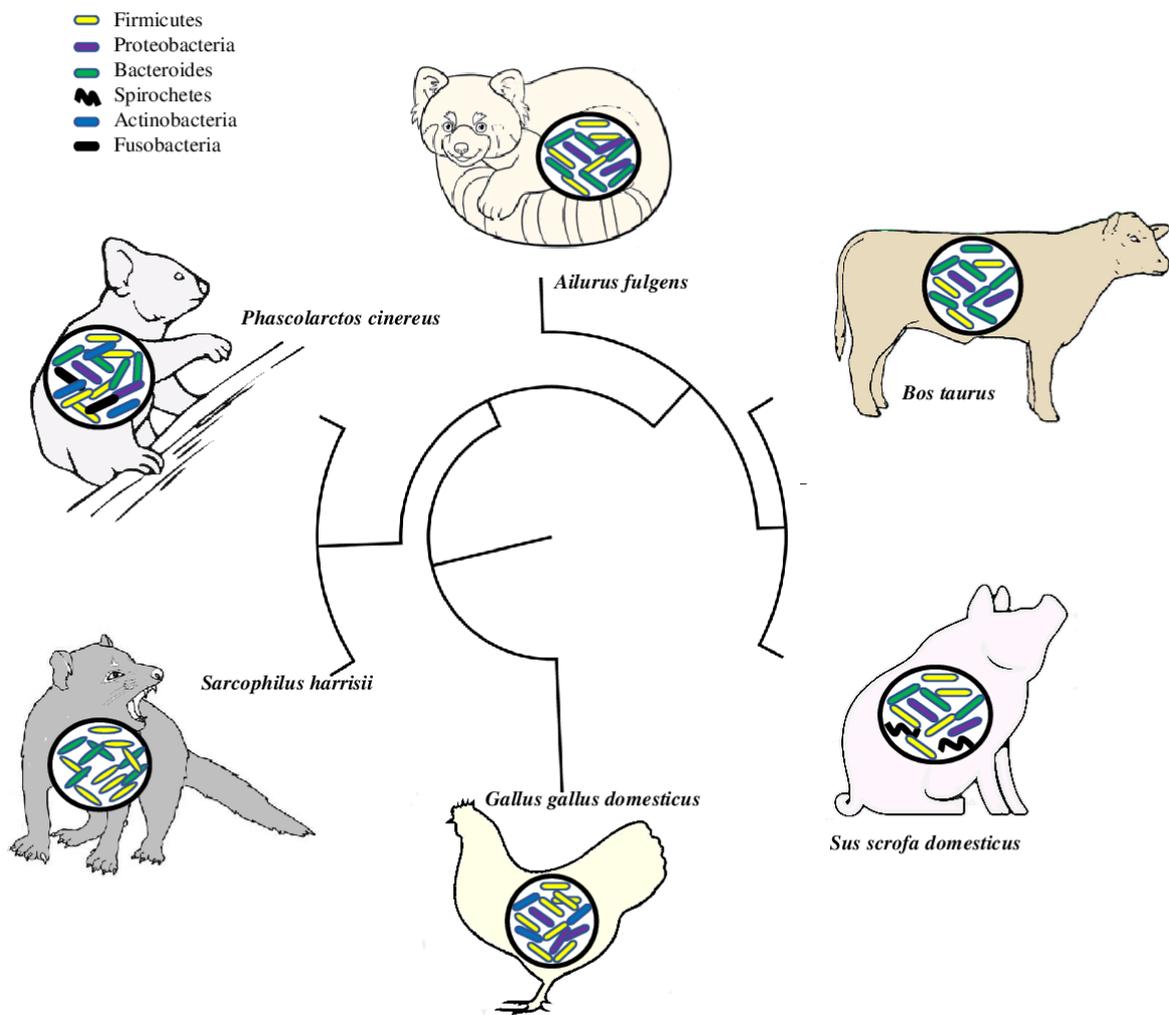


Figure 2: Overview of the dominant phyla in the gut microbiota of different mammalian species, from domesticated to wild animals.

C. Correlation between microbiota and disease

As underlined previously, the development of such a microbial assemblage represents the establishment of a balanced equilibrium between microbial components and the host (Fig.3).

Several factors, such as the environment and the lifestyle, may cause shifts in the composition of the microbiota, and thus causing dysbiosis.

Gut dysbiosis is typically related to a negative impact on host health with long-term consequences, being associated with several diseases or disorders, such as Inflammatory Bowel Disease (IBD), metabolic syndrome, obesity, diabetes, neurological disorders and cancer (Milani et al., 2017) (Fig.3). Improved sequencing technologies and other “omics” technologies (such as proteomics and metabolomics), together with metabolic network modeling (Borenstein et al., 2008; Freilich et al., 2009), can explain how host or environmental factors affect gut microbial ecology over the host lifespan. Among these factors, nutrition, delivery mode, intestinal permeability, pathogenic infections, and antibiotic use, have a large impact on microbiota development (Kerr et al., 2015) (Gohir et al., 2015). In this regard, alteration of the gut microbiota during infancy has been associated with obesity and obesity-related disorders. For example, recent analysis showed how *Bacteroides fragilis* levels at one month of age are significantly correlated with a higher body mass index (BMI) in children (Milani et al., 2017).

Dysbiosis in early life has also been related to type 1 diabetes (T1D), which is an autoimmune disease characterized by the destruction of insulin-producing beta cells. Studies performed in patients with T1D have shown a lower diversity and significant differences in the ratios of the most abundant gut commensal phyla, such as Firmicutes and Bacteroidetes, as well as a decreased abundance of the butyrate producer *Faecalibacterium prausnitzii* in diabetic children (Gulden et al., 2015). In this context, butyrate-producing species are more important in supporting host health, and it has been suggested that these bacteria play a crucial role in reducing the risk of developing T1D (de Goffau et al., 2014).

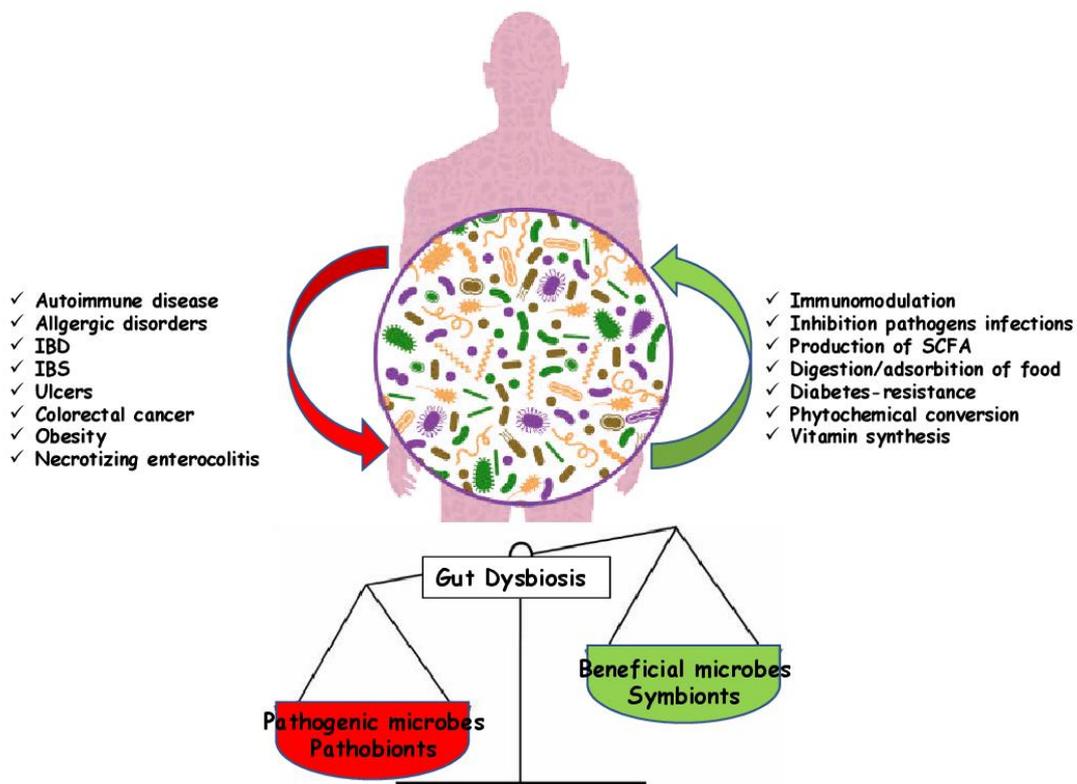


Figure 3: Schematic representation of several factors responsible for shifts in the composition of the microbiota, and thus causing dysbiosis.

Gut microbiota in Inflammatory Bowel Disease

A wide range of factors can cause shifts in the gut microbiota balance. In this context, a study performed in adults suggested how geographical origin has an important impact in developing dysbiosis (Loftus, 2004), where the incidence of IBD is higher in industrialized Western societies rather than in traditional agricultural cultures (Loftus, 2004). Inflammatory Bowel Disease (IBD) includes a group of inflammatory conditions of the colon and small intestine (Baumgart and Sandborn, 2007). The most important types of IBD encompass Crohn's disease (CD) and ulcerative colitis (UC). So far, the correlation between dysbiosis and intestinal inflammation is not completely clear. It seems that one of the cause is the reduced bacterial diversity, characterized by the loss of commensal species such as *Clostridium leptum* as well as *Eubacterium* and bifidobacteria (Favier et

al., 1997; Mangin et al., 2004). These observations are consistent with a role of the gut microbiota in the development of IBD. In detail, CD can occur in all sites of the GI tract. Investigations of the CD pathology has shown an increase of genera belonging to Proteobacteria phylum in the gut of CD-patients (Li et al., 2012).

On the other side, UC is generally characterized by inflammation and ulceration confined to the colorectal region and to the mucosal layer of the GIT. The analysis of samples affected by UC highlighted a positive correlation of this pathology to bacteria belonging to the Enterobacteriaceae family (Garrett et al., 2010; Gerritsen et al., 2011). Besides, analyses of gut microbiota composition through 16S rRNA microbial profiling revealed that members of the *Faecalibacterium* taxon are heavily compromised in UC patients (Morgan et al., 2015; Reshef et al., 2015).

Additionally, the gut microbiota of patients affected by UC, displayed a significant depletion of bifidobacteria. In this context, the lack of this genus, and especially *Bifidobacterium bifidum*, is directly connected to disease progression, underling the importance of this taxon for the prevention and treatment of the intestinal inflammation (Duranti et al., 2016).

Recently, it has been proposed that bacterial products such as butyrate possess a regulatory effect on inflammation in IBD (Segain et al., 2000; Sanderson, 2004). In details, butyrate is known to downregulates mucosal inflammatory responses (Segain et al., 2000; Maslowski et al., 2009), and it has been shown to be an effective topical treatment of ulcerative colitis (Scheppach et al., 1992; Hallert et al., 2003; Song et al., 2006). Besides, patients with CD showed a loss of butyrate-producing species (van Nuenen et al., 2004; Manichanh et al., 2006; Marchesi et al., 2007; Nemoto et al., 2012): this reduction suggests that changes in the metabolic activity, as well as composition of intestinal microbiota are important in the development of IBD (Spor et al., 2011; Kim, 2015; Matsuoka and Kanai, 2015; Baothman et al., 2016). Interestingly, a number of clinical interventions aimed to counterbalance the gut dysbiosis through fecal microbiota transplantation and probiotics, provide encouraging results in the treatment of IBD (Matsuoka and Kanai, 2015).

Colorectal cancer

In the last years, considerable attention has been drawn to the investigation of the association between gut dysbiosis and colorectal cancer (CRC). In fact, the gut microbiota seems to exploit an impact on colorectal carcinogenesis (Huycke and Gaskins, 2004; Garrett, 2015). In this context, it has been shown that perturbations on the composition of the gut microbiota influence the incidence and progression of CRC (Zitvogel et al., 2017). Several by-products of the gut microbiota can directly target intestinal epithelial cells and either mediate oncogenic effects (e.g., hydrogen sulfide and the *Bacteroides fragilis* toxin) or suppress tumorigenesis (as demonstrated for SCFA) (Zitvogel et al., 2015).

Several studies suggest that certain bacterial taxa can directly affect tumor development by producing virulence factors like toxins or by manipulating inflammatory status of the tumor microenvironment (Wu et al., 2009; Kostic et al., 2013; Gagniere et al., 2016; Rea et al., 2018).

According to this view, a human colonic bacterium, the Enterotoxigenic *Bacteroides fragilis* (ETBF), which secretes *B. fragilis* toxin (BFT) causes human inflammatory diarrhea and strongly induces colonic tumors in multiple intestinal neoplasia (MIN) mice by triggering a T helper type 17 (Th17) inflammatory response (Wu et al., 2009). Furthermore, recent investigations suggest how *Fusobacterium nucleatum* is a potential candidate for CRC susceptibility. In this regard, *F. nucleatum* has been detected on the surface of over 50% of colon rectal adenomas (Kostic et al., 2013). Interestingly, clinically isolated *F. nucleatum* is able to accelerate onset of colonic tumors in (*Apc*)^{Min/+} mice by promoting intestinal inflammation within the tumor microenvironment (Kostic et al., 2013).

Besides, *Helicobacter pylori* contributes to carcinogenesis through epithelial injury and inflammation. Supporting this finding, the pathogenic bacteria *H. pylori* has been classified as a carcinogen by the International Agency for Research on Cancer (IARC) (Rea et al., 2018).

In last years, there has been a growing interest in understanding the differences in the inter-individual microbiota composition, because it could explain variations in gut metabolic processes, such as metabolism of drugs and dietary substrates (Clayton et al., 2009). In this regard, many health benefits of soy-rich diets, linked to positive outcomes for osteoporosis, prostate cancer and cardiovascular disease, have been attributed to S-(-)equol produced from the soy isoflavone diadzein by bacterial rather than human enzymes. Recent studies suggest that only 25–30% of the adult population of Western countries produce S-(-)equol when fed soy foods, compared to a 50–60% production in adults from Japan, Korea, or China (Setchell and Clerici, 2010; Jackson et al., 2011). Interestingly, it has been underlined that cancer-protective effects of soy described in Asian populations might not

generalize to Western population because of differences in key components of the microbiota (Lozupone et al., 2012).

Modulation of the microbiota composition by prebiotics

Modifications occur in the gut microbiota throughout host lifespan, especially during childhood, and the modulation of the microbial composition could have important consequences for pathogens colonization and invasion (Relman, 2012).

This concept has fueled the development of novel strategies shaping and influencing the infant microbiota composition by using nutraceutical products (e.g., prebiotics and/or probiotics).

In fact, immediately after delivery, HMOs are considered natural prebiotic compounds because they actively stimulate the growth of specific members of the newborns gut microbiota. In this context, HMOs specifically enhance the growth of certain bifidobacteria, able to efficiently metabolize HMOs (Locascio et al., 2009). In detail, *Bifidobacterium longum* subsp. *infantis* is capable to hydrolyze the core structure (lacto-N-tetraose [LNT] or lacto-N-neo-tetraose [LNnT]) of HMOs. According to this view, the prebiotic effects exerted by HMOs are structure specific, in fact, some bacteria are able to metabolize type 1 structures (terminal Gal1-3GlcNAc). Whereas other bacteria prefer type 2 structures (Gal1-4GlcNAc) or they may be able to metabolize branched HMOs. There are multiple mechanisms by which HMOs help to shape microbial communities in the infant gut. HMOs may exploit antimicrobial effects toward certain bacterial groups like group B *Streptococcus* (GBS) which stops growing in the presence of HMOs (Lin et al., 2017). Furthermore, HMOs have an impact also on certain yeast cells. In this latter case, HMOs are able to alter hyphal morphology and length, and epithelial cells attachment of *Candida albicans* (Gonia et al., 2015).

This bacteriostatic effect seems to be related to specific HMO structures that disrupt proper bacterial membrane glycosylation. Usually, pathogens need to attach to epithelial surfaces in order to colonize and invade the host. This attachment is often facilitated by protein-glycan interactions through specific proteins encoded by pathogens or when pathogens are covered by glycans binding to proteins on epithelial cells. HMOs are able to resemble some of these glycans and serve as soluble analogs that block pathogen attachment (Gonia et al., 2015).

In addition, prebiotics can be used to modify the gut microbiota of bottle-fed infants. In this context, when prebiotics are added to infant formula, they promote the development of a neonatal gut microbiota very close to that of breastfed infants (Guaraldi and Salvatori, 2012) showing an increase of bifidobacteria and lactobacilli abundance (Vandenplas et al., 2015; Jinno et al., 2017). The main prebiotics used in infant formula are galacto-oligosaccharides (GOS), fructo-oligosaccharides (FOS),

and polydextrose (PDX). Supplementation of infant formula with the prebiotic PDX resulted in higher abundance of bifidobacteria (Scalabrin et al., 2012).

Many studies support the notion that GOS and inulin-type fructans are beneficial for digestive and immune health (Kunz et al., 2000; Roberfroid et al., 2010). Besides, addition of GOS or a GOS/FOS mixture to infant formula has a positive effect on bifidobacterial and lactobacilli abundance (Knol et al., 2005; Rinne et al., 2005; Ben et al., 2008; Vandenplas et al., 2015). Notably, in infants fed with GOS, *Bifidobacterium breve* levels were higher, whereas those of *Clostridium difficile* were lower compared to fed control formula (Sierra et al., 2015).

Moreover, non-digestible oligosaccharides added to formula have shown results similar to those of breastfeeding in reducing the colonic pH and increasing the production of SCFAs and lactate.

Finally, high-risk preterm born infants formula-fed supplemented with FOS, were shown to have an increased number of bifidobacteria present in fecal samples and a significant reduction of *Escherichia coli* and enterococci (Kapiki et al., 2007).

According to this view, there are different strategies to modify the development of microbiota in order to prevent latter diseases, but additional studies are needed to clarify functional differences between the effects of different types of prebiotics and/or the combination of different ones.

D. Gut microbiota and microbial biomarkers

Perturbations in the gut microbiota composition have been correlated with multiple diseases and disorders. There is a growing interest in understanding the precise mechanisms and causalities of these alterations, because they could be utilized for novel diagnostic and prognostic tests, also known as microbial biomarkers. Identification of specific microbial biomarkers might help to predict the risk of chronic diseases such as metabolic syndrome, colorectal cancer and intestinal inflammations. Moreover, microbiota interventions, i.e. bacterial therapies, might offer a new opportunity for addressing chronic diseases in the future. In the gut microbiota of adults, a decrease in microbial diversity is considered to represent a biomarker associated with disease. Notably, a decrease in richness of the microbiota population is described as a biomarker for metabolic syndrome, but it is not considered to be sufficient to reliably identify and confirm the presence of an ongoing pathological state (Milani et al., 2017).

Bifidobacteria as a marker for health

Bifidobacteria represent important commensals of the human gut especially in the first stages of life, where they represent the dominant members of the infant gut microbiota (Hill et al., 2012; Turrone et al., 2012; Arrieta et al., 2014; Bergstrom et al., 2014). Bifidobacterial levels become lower after weaning but remain relatively stable in adulthood. The presence and abundance of bifidobacteria is associated with a healthy status of the host and they can be considered as a potential microbial biomarker. In this context, aberrant bifidobacterial numbers is one of the most frequently observed intestinal microbiota alterations in infants associated diseases (Shean and Faia, 1975; Kerckhoffs et al., 2009).

Notably, colic and necrotizing enterocolitis (NEC) in newborns, as well as development of obesity, celiac disease, and autoimmune diseases in adulthood, are correlated with lower levels of bifidobacteria in the infant intestine (Butel et al., 2007; Tojo et al., 2014).

Moreover, this taxon has a pivotal role in terms of immune regulation through specific immune stimulation and acidification of the intestinal environment by the production of SCFAs and lactate.

In detail, many studies suggested that *Bifidobacterium longum* subsp. *infantis* is associated with normal development of immune tolerance, and it is able to normalize the permeability of the intestinal mucosa (Fukuda et al., 2011; Chichlowski et al., 2012). Furthermore, a lower abundance of *B. longum*

is typically related to malnutrition in children and it goes through an increased risk for impaired learning ability and physical stunting in later life (Gordon et al., 2012).

A recent investigation, show how *Bifidobacterium bifidum* PRL2010 provides opportunities for the prevention and treatment of the ulcerative colitis directly through the re-establishment of the gut microbiota homeostasis (Duranti et al., 2016).

In this latter case, *B. bifidum* PRL2010 is a well characterized species, possessing a genetic repertoire able to metabolize mucin, another host-produced glycan that constitutes one of the main barriers covering the GIT mucosa (Turroni et al., 2010). Besides, pilus-like structures produced by *B. bifidum* PRL2010 are pivotal in modulating the interaction between PRL2010 cells and human enterocytes as well as with other members of the gut microbiota (Turroni et al., 2013; Turroni et al., 2014). In addition, cross-feeding actions exploited by PRL2010 cells provide simple sugars that can in turn be utilized to other microbial gut inhabitants (Egan et al., 2014b; Egan et al., 2014a; Turroni et al., 2015). In this context, microbe-microbe interactions can either positively or negatively influence the fitness of the affected organisms (Pande et al., 2015) by the release of compounds in the environment (Phelan et al., 2011; Morris et al., 2013). As mention before, *B. bifidum* PRL2010 exerts a protective role toward colitis ensues not only from the regulation of the host innate immunity, but also from the enhancement of the intestinal barrier. Duranti et al., provide evidence that pre-treatment of TNBS-induced colitis mice with *B. bifidum* PRL2010 as well as with the major subunit of one of the sortase-dependent pili encoded by this microorganism, FimP_{PRL2010}, impacts on the status of innate and T-cell polarization responses in induced colitis (Duranti et al., 2016).

Moreover, mice with mesenteric ischemia/reperfusion, a clinical emergency with high mortality associated with several pathological conditions, such as organ transplantation, or bowel strangulation, treated with *B. bifidum* PRL2010, showed attenuate mesenteric-induced changes. *B. bifidum* PRL2010 exerted a slight beneficial effect on neutrophils infiltration and lipid peroxidation in the gut of treated mice (Duranti et al., 2018).

Finally, another important aspect worth of consideration is that bifidobacteria have been associated with reduced abdominal pain in healthy adult and are capable to alleviate gastrointestinal symptoms of adult celiac patients. Increasing levels of this taxa in the intestine could be considered a target to prevent and/or alleviate microbiota-associated disease

The *Lactobacillus* genus as a microbial biomarker

In recent years, Next Generation Sequencing (NGS) has allowed the discrimination of intestinal bacteria and their correlations with diseases state, but the identification of universal gut microbial biomarkers linked to a disease/healthy state is still at the beginning.

Lactobacillus genus is widespread in different ecological niches; this taxon is a member of host-associated gut and vaginal microbiota as well as of functional and fermented foods

(Walter, 2008; Duar et al., 2017). One of the main niche of *Lactobacillus* spp. is the female genital tract and the most frequently isolated species are *Lactobacillus crispatus*, *Lactobacillus gasseri* and *Lactobacillus jensenii*. A *Lactobacillus*-dominated microbiota appears to be a good biomarker for the vaginal health because this genus can create a barrier against pathogen invasion by producing components that can play an important role in the inhibition of bacterial and viral infections (Petrova et al., 2015).

Besides, lactobacilli are known to be present in the infant gut microbiota at lower numbers than other bacterial taxa. Vaginally delivery infant possesses higher abundance of *Lactobacillus* species such as, *L. gasseri*, *Lactobacillus ruminis*, *Lactobacillus breve*, *Lactobacillus reuteri*, *Lactobacillus sakei*, *Lactobacillus plantarum*, and *Lactobacillus casei* if compared to C-section newborns (Nagpal et al., 2016). Supporting the vertical transmission of some *Lactobacillus* species, a recent investigation demonstrates that *L. casei* possesses an enzymatic arsenal for HMOs metabolism (Bidart et al., 2016). Interestingly, a recent study aimed to identify potential universal intestinal microbial biomarkers, provides that individuals affected by gut diseases (DS) exhibited higher abundance of *Lactobacillus* compared to healthy subjects (Mancabelli et al., 2017b). Probably, the observed results of the non-pathogenic *Lactobacillus* taxa in DS may reflect lower niche competition caused by simplification of the dysbiotic gut microbiota (Walter, 2008).

Notably, lactobacilli are fastidious organisms with nutritional requirements and they would consider disadvantageous persist in distal regions to host nutrient absorption. Lactobacilli require amino acids, nucleic acid derivatives, peptides, vitamins, salts and fermentable carbohydrates for growth, and they possess very limited capabilities to utilize complex carbohydrates. Nonetheless, in a dysbiotic gut, the higher abundance of this taxon might reflect a lack of competitiveness in the gut ecosystem (Walter, 2008).

However, more investigations and studies are needed in order to fully characterize new potential microbial biomarkers and their correlation with host health.

E. Technical approaches for microbiota determination

Although microorganisms are ubiquitous and abundant, we currently lack a fundamental mechanistic understanding of many of the key roles played by microorganisms in nature, including those that reside in the human and animal body. Until the recent development of novel culturomics approaches, only a very small fraction of the human and animal gut microbiota had been isolated and studied in pure culture (Biteen et al., 2016). So far, because of a large proportion of the human gut microbiota was uncultured several culture-independent approaches have been developed, i.e., metagenomics, metatranscriptomics, and metaproteomics, to discover the identities, functional roles and activities of the so-far-uncultivated members of the gut microbiota (Rajilic-Stojanovic and de Vos, 2014).

In this regard, high-throughput sequencing of (a portion of) the 16S rRNA gene (i.e., 16S rRNA gene-based microbial profiling analysis) as a conserved phylogenetic marker represents the current standard technique for profiling complex microbial communities, although shotgun metagenomics is progressively replacing 16S rRNA gene-based microbial profiling analysis.

The 16S rRNA gene-based microbial profiling approach exploits universal primers for amplification of single or multiple hypervariable regions of the 16S rRNA gene (Hamady and Knight, 2009). Reads of the obtained amplicons, gained from a next-generation sequencing (NGS) platform, are processed through bioinformatic pipelines, such as the popular Qiime software suite (Caporaso et al., 2010) or Mothur (Schloss et al., 2009), thus generating taxonomic profiling of the microbiota of the analyzed environmental samples (Hamady and Knight, 2009).

16S rRNA gene-based microbial profiling also facilitates identity assignment for unknown bacterial members through discrimination based on the sequences of their unique hypervariable regions (Clarridge, 2004). Moreover, the sequencing of the gut microbiome, through metagenomics approaches, has been developed to confirm both the phylogenetic and the functional gene repertoire of the gut microbiota (Gilbert and Dupont, 2011). Anyway, one of the limitations of this approach is that do not provide information on which genes are expressed or not at any given time. Other -omics approaches overcome these limitations, such as the sequencing of the whole microbial RNA pool of a given sample, i.e., metatranscriptomics, or analysis of the overall protein content, i.e., metaproteomics. However, it is worth mentioning that as well as for the metagenomics approach, the limitations of these two latter technologies is that many genes, as well as their products, are not still functionally characterized. Finally, the evaluation of the microbial produced metabolites, i.e., metabolomics, will provide an overall signature representing microbial activities.

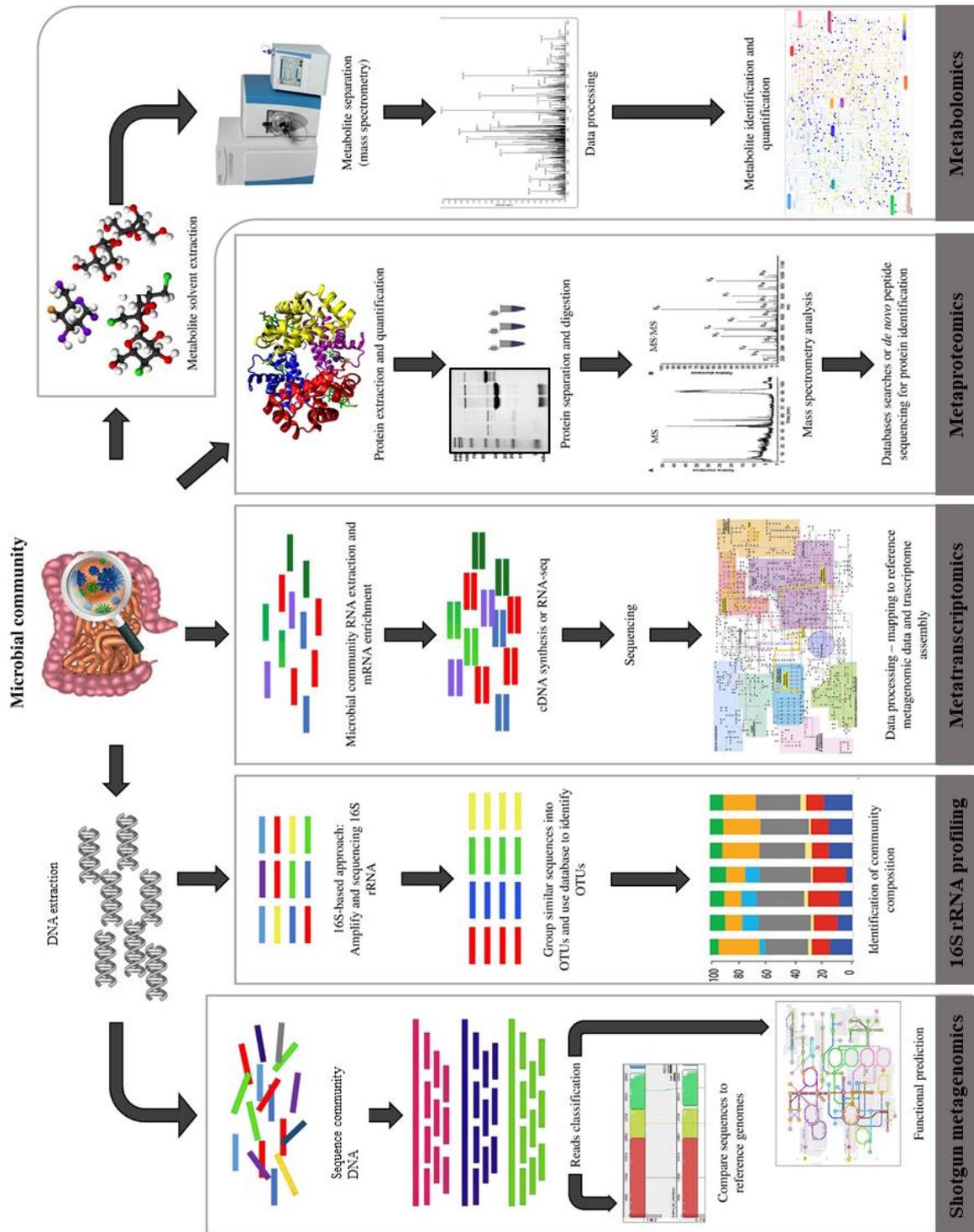


Figure 4: An illustration with the main steps in metagenomic, metatranscriptomic, metaproteomic and metabolomic approaches.

Methodologies for the investigation of the microbiota composition

Several studies investigating the human gut microbiota relied on 16S rRNA gene-based microbial profiling analyses. The 16S rRNA gene encompasses nine different variable regions, i.e., V1 to V9, each flanked by highly conserved DNA sequences that are suitable for specific primer binding and very often the decision to employ a specific primer pair is based on historic use or/and current literature (Sundquist et al., 2007; Claesson et al., 2010; Turrone et al., 2012; Milani et al., 2013).

In addition, none of the currently available DNA sequencing technologies offers full-length gene sequencing at sufficient depth for cost-effective multiplexing of multiple samples in a single run.

As mentioned before, an alternative to human gut microbiota cataloging that is replacing 16S rRNA gene microbial profiling is represented by shotgun metagenomic sequencing allowing to profile the entire gut microbiota and thus bypassing gene-specific amplification. This approach decoded all (fragmented) microbial DNA, which was extracted from the analyzed environmental sample, including also unclassified bacteria and viruses. It provides more information, such as insights into functional aspects of the microbial community and it does not suffer from the potential bias of the amplification reaction required for 16S rRNA gene-based profiling.

Deep in detail, shotgun data can be used to explore the repertoire of genes involved in a wide range of metabolic processes, such as those employed in biosynthesis of compounds, e.g., short-chain fatty acids, or in the catabolism of nutrients, e.g., carbon sources.

Furthermore, functional classification of the shotgun metagenomic reads through the use of customized databases allow to assess antibiotic resistance, presence of (pro)phages, degradation of conjugated bile salts, extracellular structures responsible for adhesion, and immunomodulation. Moreover, an assembly-based approach can be useful to reconstruct complete or partial genomes of uncultivated taxa (Rinke et al., 2013). However, the enormous amount of data obtained from DNA sequencing of complex bacterial communities, such as those residing in the gastrointestinal tract (GIT), requires processing power and bioinformatics pipelines for sequence information management, interrogation, and administration (Biteen et al., 2016). It should be mentioned that underpopulated reference databases and poor functional characterization of many genes considerably limit the usefulness of the metagenomic approaches employed to investigate the microbiota composition.

Approaches to complete a high-definition image of the gut microbiota composition

16S rRNA gene-based microbial profiling analyses provide insights into the composition of the gut bacterial community at a higher taxonomic level than species (Chakravorty et al., 2007). To overcome this limitation and to obtain a deeper image of the composition of the human gut microbiota, i.e., at the species or even subspecies level, it is required to target a molecular marker that is much more variable at the interspecies level with respect to 16S rRNA gene.

The internally transcribed spacer (ITS) sequence, represent the spacer region between the 16S rRNA and the 23S rRNA genes, within the rRNA locus. It represents a valuable genetic marker for such a purpose, in fact, an ITS-based protocol known as ITS bifidobacterial profiling analysis was used to obtain a detailed image of bifidobacterial communities. This approach can discriminate bifidobacteria at the subspecies level and thus to disentangle bifidobacterial community composition in complex ecosystems, including the human and animals gut (Milani et al., 2014; Milani et al., 2016). In this regard, the ITS bifidobacterial profiling approach is helpful for taxonomic information about strain identity and also supply data about the genetic of the organism, thus providing metabolic and evolutionary insights (Milani et al., 2014). Complete genome analysis of the gut microbiome means decoding the complete genome sequence of each strain. This aim requires an important effort because of the complexity of the gut microbiota, which may include hundreds of operational taxonomic units (OTUs).

Moreover, the frequent incapability to simulate the natural growth conditions, under *in vitro* experiments, renders the cultivation of most members of the gut microbiota even more difficult. Single-cell genomics can efficiently contribute to the genomic characterization of the microbiome. Standard approaches to single-cell analyses include the physical isolation of the microbial cell, followed by extraction of chromosomal DNA from each cell and amplification of its genomic content (Dean et al., 2001).

Remarkably, single-cell genome sequences can be obtained directly from crude samples, hence generating reference genome sequences for those gut microorganisms that are recalcitrant to cultivation (Fodor et al., 2012; McLean et al., 2013). Unfortunately, the currently available single-cell approaches are still not particularly efficient, while the quality of the attained data and the possibility of contamination may skew output data compared to that obtained by standard genomic methods.

Currently, single-cell data sets enable the recovery of only about 35% of the genomic data. However, if this technique can be further improved, is expected to fill important knowledge gaps of the contents

and structure of the human gut microbiome. Nonetheless, despite promising developments of microfluidic technologies for microbial single-cell investigations, actual implementation of this approach remains very challenging.

A recently applied approach to unveil the gut microbiota composition at high resolution down to the strain level without performing any isolation and cultivation of bacterial strains involves the reconstruction of a genome sequence of an individual microbiota member from shotgun metagenomic data. This approach provides taxonomic information about strain identity and very useful data related to the genetic composition of the organism, thereby providing metabolic and evolutionary insights (Lugli et al., 2017).

An interesting tool aimed at determining the composition of the human gut microbiota at high resolution (down to the strain level) is named MetaPhlAn (Segata et al., 2012). This software relies on read mapping to a precomputed database of strain-specific marker genes generated through comparative analysis of all publicly available bacterial genome sequences. The main limit of this tool is that only previously sequenced species can be profiled, thus ignoring the presence of as-yet-unknown/uncultured members in the population.

Culturomics approaches

In the last years, the culture-independent approaches have been applied mostly in order to reveal the human and animals gut microbiota composition, thus neglecting microbial cultivation techniques. In this regard, there is a substantial knowledge gap between the gut bacterial species that have been cultivated and those that have not yet been isolated and thus characterized (Hugon et al., 2015). It has been reported that approximately 56% of gut microbiota detected by NGS approaches have cultured representatives (Goodman et al., 2011; Walker et al., 2014). Recently, with the advent of so-called culturomic approaches, this gap is being shrinking. Culturomics is based on high-throughput cultivation conditions to explore the human gut microbiota. Recently, several culturomics studies of human stool samples involved the formulation of more complex growth media, which allowed the isolation and cultivation of a large number of novel gut microorganisms (Lagier et al., 2012; Browne et al., 2016; Lagier et al., 2016).

Chapter 2

Outline of the thesis

The aim of this PhD thesis is to provide insights into the microbial composition of the gut microbiota in animal and human beings. In detail, it is now well known that gut microbiota may impact on host health during the entire host lifespan.

Perturbations of the normal gut microbiota, i.e. dysbiosis, have been associated with several disease and disorders. Thanks to the advances of the next generation sequencing era it was possible to perform several analyses to better understand the host-microbial interactions.

Chapter 3 illustrates an in-depth cataloguing of the bifidobacterial communities that inhabits several mammalian species by means of an Internal Transcribed Spacer (ITS) sequences for bifidobacterial profiling (Bifidobacterial-ITS profiling) approach.

Chapter 4 and 5, describes the identification and characterization, through a culture-dependent approach, of seven bifidobacterial isolates, which have been taxonomically and genetically characterized as novel bifidobacterial species.

Chapter 6 discusses the role of bifidobacteria in the gastrointestinal tract of mammals and how they can persist and colonize the human gut. Bifidobacteria are important gut commensals, however, the molecular mechanisms by which these microorganisms establish themselves in this environment is largely unknown. In this context, an analysis of the genetic diversity of the predicted arsenal of sortase-dependent pili belong to known members of the *Bifidobacterium* genus has been performed.

Chapter 7 describes a 16S-rRNA microbial profiling analysis of the human gut microbiome comparing healthy mucosa vs mucosa with gastrointestinal polyps from the same patient and subsequently comparing the same samples with mucosa from patients with colorectal cancer (CRC). In particular, a metagenomics analysis has been performed with the aim to identify possible microbial biomarker for the early identification of CRC.

Chapter 8 describes a novel metagenomic pipeline, referred to as Genetic Identity Card (GIC), able to reveal the microbiota composition of probiotic supplements and the genetic/genomic content of the strains included in the analyzed probiotic supplements. This pipeline encompasses next-generation sequencing analyses based on 16S rRNA-associated, combined with a shotgun metagenomics approach.

Chapter 9 illustrates the development of an ITS-based profiling method allowing to accurately profile lactobacilli communities at species-level. This approach encompasses a genus-specific primer pair combined with a database of ITS sequences retrieved from all available *Lactobacillus* genomes and a bioinformatic pipeline performing all required steps to reconstruct a lactobacilli species-level profile.

Chapter 3

Unveiling bifidobacterial biogeography across the mammalian branch of the tree of life

Milani C, Mangifesta M, Mancabelli L, Lugli GA, James K, Duranti S,
Turrone F, Ferrari C, Ossiprandi MC, van Sinderen D, Ventura M.

This chapter was published in:

ISME journal. 2017 Dec;11(12):2834-2847. Epub 2017 Aug 22.

Reprinted with permission from Springer Nature.

Unveiling gut microbiota biogeography across the mammalian branch of the tree of life

Key words: gut microbiota, *Bifidobacterium*, tree of life, mammals, gastro intestinal tract, vertical transmission

Christian Milani¹, Marta Mangifesta^{1,2}, Leonardo Mancabelli¹, Gabriele Andrea Lugli¹, Kieran James³, Sabrina Duranti¹, Francesca Turrone¹, Chiara Ferrario¹, Maria Cristina Ossiprandi⁴, Douwe van Sinderen³ and Marco Ventura¹

Laboratory of Probiogenomics, Department of Life Sciences, University of Parma, Italy¹; GenProbio srl, Parma, Italy²; APC Microbiome Institute and School of Microbiology, Bioscience Institute, National University of Ireland, Cork, Ireland³; Department of Medical-Veterinary Science, University of Parma, Parma, Italy⁴

Corresponding author. Mailing address for Marco Ventura Department of Life Sciences, University of Parma, Parco Area delle Scienze 11a, 43124 Parma, Italy. Phone: +39-521-905666. Fax: +39-521-905604. E-mail: marco.ventura@unipr.it

Abstract

Internally Transcribed Spacer (ITS) rRNA profiling is a novel tool for detailed analysis of microbial populations at low taxonomic ranks. Here, we exploited this approach to explore species-level biogeography of the *Bifidobacterium* genus, across 291 adult animals that fall within the mammalian branch of the tree of life. These include humans and 13 other primates, domesticated animals, such as dogs, cats, cows, sheeps, goats, horses and pigs, and 46 additional mammalian species. The collected profiles revealed the presence of 89 putative novel bifidobacterial taxa in addition to 45 previously described species. Remarkable, in contrast to what currently known for many gut commensals, we did not observe host-specialization among bifidobacterial species, but rather their widespread distribution across mammals. Moreover, ITS rRNA profiling of wild relatives of domesticated dogs, rabbits and pigs clearly indicates that domestication and close contact with humans have impacted on the composition of the fecal bifidobacterial population. These data were complemented by analysis of bifidobacterial communities in milk of eight mammalian families, showing that bifidobacteria represent prototypical early gut microbiota members which are inherited by new-borns from their lactating mother. Thus, this study highlights the role of bifidobacteria as pioneering gut colonizers of a wide range of mammals.

Introduction

The development of the mammalian gut microbiota starts at birth, with colonization of the sterile gastro-intestinal tract (GIT) of the newborn by

bacteria that are harbored by the mother or present in the surrounding environment (Yatsunenکو et al 2012). In this regard, recent studies have highlighted a combined effect of bacterial acquisition from the surrounding environment, including that of the mother, and of gut commensal selection based on human genetic heritage (van Opstal and Bordenstein 2015). In the first six to twelve months following birth, the gut microbiota is in a continuous compositional flux and its biodiversity progressively expands as the infant diet evolves from solely milk-based to that typical of their parent (s) or carer (s) (De Filippo et al 2010, Yatsunenکو et al 2012). Despite its biological relevance, the high level of complexity and the absence of cost-effective tools to assess the gut microbiota at strain-level have so far prevented a detailed profiling of its transmission route (s).

Being among the first gut colonizers of newborns, bifidobacteria are known to be an abundant component of the gut microbiota of various animals (including humans), that provide parental care to their offspring (Avershina et al 2016, Bunesova et al 2014, Milani et al 2014a, Turrone et al 2009a). Moreover, bifidobacteria have been shown to exert crucial roles in a variety of biological processes such as development of the GIT, induction of mucus layer production, protection against opportunistic pathogens, maturation/stimulation of the immune system, as well as expansion of the gut glyco-biome and participation in the processing of (from a host perspective) indigestible food components (Arbolea et al 2016, Milani et al 2015c, Milani et al 2016, O'Callaghan and van Sinderen 2016, Ventura et al 2012). We have recently developed an Internal Transcribed Spacer (ITS)-based

metagenomic tool for bifidobacterial community profiling (Milani et al 2014b), allowing (i) accurate taxon identification down to phylotype level, (ii) tracking vertical transmission events, and (iii) discovery of putative novel bifidobacterial species (Milani et al 2014b, Milani et al 2015b). Here, we describe the composition of the bifidobacterial gut community present in (fecal samples of) various animals that span the (phylogenetic) mammalian tree of life, and provide clues as to how these gut commensals establish themselves in the newborn gut through a vertical transmission route involving maternal milk as both a vehicle and a biological reservoir of early colonizing bacteria.

Materials and methods

Sample collection. Human fecal samples were collected at the Parma University Hospital (Parma, Italy) while stool of other mammals was retrieved at several Italian zoological parks. Samples consisted of 10 gr of fresh fecal material, which is considered to be a sufficient amount to represent the overall biodiversity of the fecal microbiota as outlined in previously published reports (Ley et al 2008). In order to be sure that the fecal material was fresh and of a certified origin, it was collected immediately following defecation. Furthermore, this amount of stool sample is sufficient for both bacterial DNA extraction and isolation purposes (see below for protocol details). Only samples of verifiable origin were retained. In case of aquatic mammals, sampling was performed during a veterinary examination routine and fecal samples were collected with rectal swabs to avoid contamination (Harper et al 2003). Conservation and shipment were performed following the “Stool Packaging Instructions” of the Core Microbiome

Sampling Protocol A reported in the Manual of Procedures of the Human Microbiome Project (http://www.hmpdacc.org/doc/HMP_MOP_Versi on12_0_072910.pdf). In detail, all samples were cooled to 4°C in a fridge under anaerobic conditions immediately after collection and within 24 hours packed with ice and sent under anaerobic conditions to the laboratory where they were processed. To ensure refrigeration during transfer, samples were shipped using the World Courier Cold Chain Logistics. Notably, it has previously been established that refrigeration at 4°C for up to 72 hours does not impact on microbial composition, as has also been demonstrated for long term storage at -80°C (Choo et al 2015). Upon arrival, approximately 4 gr was immediately processed by the DNA isolation protocol, while the remaining ~6 gr was maintained at -80°C until DNA extraction. These storage conditions were chosen to avoid alterations of the relative abundance of taxa caused by bacterial growth or DNA degradation, while also retaining viability for isolation attempts (Cardona et al 2012, Fouhy et al 2015). Samples were subjected to DNA extraction using the QIAamp DNA Stool Mini kit following the manufacturer's instructions (Qiagen).

Ethical statement. This study was carried out in accordance with the recommendations of the ethical committee of the University of Parma. The protocol (n. prot. 161/OPBA/2016 N. INT 22/2016) was approved by the “Comitato di Etica Università degli Studi di Parma”, Italy. All animal procedures were performed according to national guidelines (Decreto legislativo 26/2014)

Bifidobacterial isolation. The bifidobacterial population of fecal samples was assayed using mupirocin-based medium (BSM), which has

previously been described to be selective for bifidobacteria (Serafini et al 2011, Turrone et al 2009b) (For details refer to Supplementary Text).

Fermentation profiles on milk-glycans. Details regarding growth conditions and determination of growth profiles are provided in the Supplementary Text.

Evaluation of the bifidobacterial cell density by qPCR. Evaluation of the presence of bifidobacteria in animal fecal samples was performed by quantitative real-time (qRT-PCR) as described in the Supplementary Text.

Bifidobacterial ITS profiling. Bifidobacterial ITS profiling was performed as described previously (Milani et al 2014b). Additional details are reported in the Supplementary Text.

Statistical Analyses, network representations and hierarchical clustering. Co-variance statistical analyses were performed with SPSS software v. 22 (IBM SPSS Statistics for Windows, Version 22.0. Armonk, NY: IBM Corp.) and force-driven network representations were obtained with Gephy software (Bastian 2009). Network clusters were predicted using the Modularity statistical function of Gephy. Hierarchical clustering was performed using the software MeV v.4.9.0 (<http://mev.tm4.org/>) and the Paerson correlation distance metric.

Genomes sequencing and analysis. All genomes used for this study were determined by GenProbio srl (Parma, Italy) using the MiSeq Illumina (Illumina, USA) and the MEGAnnotator pipeline (Lugli et al 2016) following the protocols described in the Supplementary Text.

Data Deposition. Bifidobacterial ITS profiling data were deposited in SRA database under the study accession number PRJNA350611. Genome

sequences for *B. sp. new_taxa_70*, *B. sp. new_taxa_71*, *B. breve* 7E and *B. pseudocatenulatum* 1E were deposited under accession numbers: MVOG00000000, MVOH00000000, MNLA00000000 and MNLB00000000, respectively.

Supplementary information are available at ISME Journal's website.

Results and discussion

Distribution of bifidobacteria across the mammalian tree of life. To precisely map the presence of bifidobacterial species in complex bacterial communities that reside in the gastrointestinal tract (GIT) of mammals, we collected 291 fecal samples of (adult) animals, corresponding to 67 different mammalian species and 28 mammalian families, and representing a major branch on the phylogenetic tree of life (Table S1). Bacterial DNA extracted from these samples was then profiled by means of a recently developed pipeline based on genus-specific primers targeting the hypervariable ITS region and a customized bifidobacterial ITS database (Milani et al 2014b, Milani et al 2015b). Illumina sequencing of the obtained amplicons generated more than 3.75 million quality-filtered reads that were grouped in clusters of identical sequences (OTUs, Operational Taxonomic Units) and then taxonomically classified. These data sets were employed to obtain the distribution patterns and relative abundance profiles of all currently known bifidobacterial species among the examined mammals. In addition, we also evaluated the proportion of the 67 mammalian species that harbor each bifidobacterial taxa, defined as prevalence. Interestingly, *Bifidobacterium longum* and

Bifidobacterium adolescentis were shown to be present at a prevalence of 95.5 % and 91 % in the profiled mammalian species, respectively, followed by *Bifidobacterium pseudolongum* and *Bifidobacterium bifidum* with a prevalence of 85 % (Fig. 1). These ubiquitous species thus appear to have evolved an ability to colonize a wide range of mammalian hosts. In contrast, 15 bifidobacterial species, despite showing the most limited distribution with a prevalence in mammalian species lower than 10 %, were nonetheless found to be present in an average of 3 mammalian orders and 4 mammalian families (Fig. 1). Interestingly, other species, which were previously thought to be highly specialized to colonize the insect gut, such as *Bifidobacterium actinocoloniiforme*, *Bifidobacterium asteroides*, *Bifidobacterium bohemicum*, *Bifidobacterium bombi* and *Bifidobacterium indicum* (Milani et al 2014a), were shown to be widely distributed among various mammalian hosts (Fig. 1). Notably, despite the low average relative abundance at which these bifidobacterial species were detected (ranging from 0.01% to 1.72%), they were not solely identified in mammals that eat insects, demonstrating that their broad distribution is not due to accidental and possibly transient colonization. These data indicate that, even if these species do not possess the same adaptive abilities of bifidobacterial taxa with a high prevalence level, the absence of a strict host-specific specialization behavior is a common feature of (currently known) bifidobacteria. The apparent wide-spread distribution of bifidobacterial species in different mammalian hosts may be a common characteristic of other members of the gut microbiota, as has been suggested by previous genus-level overviews

performed by 16S rRNA-based profiling of feces collected from animals kept in captivity (Ley et al 2008).

Furthermore, members of the mammalian families *Cercopithecidae*, *Hominidae*, *Bovidae*, *Suidae*, *Equidae* and *Canidae* were shown to be colonized by a relatively high number of bifidobacterial species, harboring on average at least 29 of the 45 currently known members of the genus *Bifidobacterium* (Fig. 1). Interestingly, these families encompass primates and domesticated mammalian species such as cattle, pigs, horses and dogs that live in close contact with humans (Fig. 1). This finding highlights that changes in gut microbiota composition may be triggered by a shift from a wild towards a domesticated life style, thus supporting previous studies on *Suidae* (Ushida et al 2016) and birds (Wang et al 2016). Moreover, the families *Hominidae* and *Canidae* were shown to harbor the highest number of known bifidobacterial species, i.e., 31 and 30, respectively, of which they had 22 in common (Fig. 1), suggesting that cohabitation of these hosts may have favored the spread and exchange of bifidobacterial microbiota members colonizing these animals. This idea has previously been put forward to explain the finding of shared microbiota members between companion animals and their owners (Lax et al 2014, Misisic et al 2015, Song et al 2013).

Effect of domestication and artificial selection on the bifidobacterial population. To further investigate the impact of domestication on the biogeography of bifidobacterial communities, we profiled the bifidobacterial population of additional undomesticated species, i.e., hares, boars and wolves, and compared the obtained profiles to those retrieved from their domesticated relatives

living in close contact with humans, i.e. rabbits, pigs and dogs, respectively. (Table S1 and Fig. S1). Notably, the obtained data revealed a higher bifidobacterial biodiversity in the domesticated species as compared to wild animals (Fig. S1), thus corroborating the hypothesis that frequent interactions with humans and a domesticated lifestyle favored acquisition of additional bifidobacterial taxa by these mammals through niche shift.

Furthermore, we evaluated if artificial selection may have impacted on gut microbiota composition. For this purpose, faecal samples were collected from 10 wolves living in the same nature reserve and from seven dog breeds living in the same kennel and with an identical diet (Table S1). The particular canine hosts were selected to represent the main phylogenetic cluster of breeds (Asian/Ancient, Hunting and Mastiff) as identified by Parker et al. based on microsatellite analysis (Parker et al 2004). Hierarchical clustering based on the average bifidobacterial profiles obtained from the fecal samples of the included wolves and dog breeds reflected that of host phylogeny (Parker et al. 2004) (Fig. S2). This finding therefore indicates that domestication and artificial selection of the wild wolf and/or the associated period of close contact with humans have gradually shaped the gut microbiota composition of the mammalian species *Canis lupus*.

Quantitative profiling of bifidobacteria among mammals. ITS-profiling of members of the *Bifidobacterium* genus in the 291 analyzed mammalian fecal samples also generated information regarding the abundance of known members of the genus *Bifidobacterium* relative to the total bifidobacterial population (Fig. 1). In

order to avoid bias of our analyses due to the variable rRNA locus copy number in bifidobacterial genomes (Lugli et al 2014), we estimated the number of rRNA gene clusters based on available bifidobacterial genome data and used this information to normalize the ITS taxonomic profiling results (Table S2 and Fig. 1).

Notably, the bifidobacterial species that exhibit high prevalence are also those that, where present, show high average relative abundance, e.g. *B. adolescentis* (32 %), *B. longum* (11.6 %), and *B. pseudolongum* (13.1 %), though with *B. bifidum* (3.6 %) representing an exception to this observation (Fig. 1). These species represent dominant bifidobacterial taxa, which may be characterized by effective and broad-ranging adaptive capabilities to colonize the gut of these mammalian hosts compared to other (co-colonizing) bifidobacteria. In contrast, of the remaining 42 bifidobacterial species, 26 exhibited, where present, an average relative abundance below 1 %, with only *Bifidobacterium eulemuris*, *Bifidobacterium magnum*, *Bifidobacterium breve* and *Bifidobacterium merycicum* representing on average > 5 % of the whole bifidobacterial population (Fig. 1). The observed low average relative abundance at which the majority of bifidobacterial species were observed reflects the general presence of multiple taxa in each mammalian species analyzed (on average 9.6 [currently recognized] bifidobacterial taxa being present at a relative abundance of > 0.01 % per mammalian species) (Fig. 1).

The above results confirm previous genus-level overviews (Ley et al 2008) and reinforce the view that bifidobacterial development was characterized by extensive selective pressure towards the

colonization of a wider range of different hosts and the development of extensive microbe-microbe co-operation (Milani et al 2015a, Turrone et al 2016). In order to assess the quantitative contribution of bifidobacteria to the overall microbiota diversity of each of the analyzed ecological niches, the total bacterial load as well as the bifidobacterial cell count were evaluated by means of qPCR. Such analyses showed that in adult mammals the gut microbiota concentration ranges between 1E+06 and 4.7E+10 genome copy number per gr of feces, with bifidobacteria representing on average 3.5 % of this microbiota population. Furthermore, bifidobacteria were detected in all profiled mammalian species, of which 53.7 % were shown to exhibit a relative abundance of the genus *Bifidobacterium* of >0.1 % (Fig. 1). Thus, while bifidobacteria do not represent one of the dominant genera in the gut environment of mammals, they appear to be widespread and ubiquitous across the entire mammalian branch within the tree of life.

Identification of putative novel members of the genus *Bifidobacterium*. ITS-profiling revealed the presence of OTUs with an identity level of <93 % with respect to any of the currently recognized bifidobacterial species. This identity cut-off represents the lowest identity level observed between subspecies when comparing the 213 publicly available ITS sequences of known bifidobacterial taxa, focusing on the hypervariable region corresponding to the Probio-bif_uni/Probio-bif_rev amplicon used for bifidobacterial ITS rRNA profiling (Table S3). Interestingly, these unclassified bifidobacterial OTUs correspond on average to 21.5 % of the bifidobacterial population in the 67 mammalian species analysed, ranging from a minimum of 0.02 % to a maximum of 99.2

% (Figs. 1 and 2). Alignment and manual evaluation of all identified unclassified OTUs resulted in the identification of 89 putative new (sub)species (named new_taxa_1 through to new_taxa_89) with OTU sequence identity \leq 92 % with respect to all currently known bifidobacterial species. These putative novel taxa were added to the bifidobacterial ITS database and profiled in all 291 samples included in this study (Table S4 and Fig. 2). Profiling data sets were normalized assuming that the putative novel bifidobacterial taxa possessing two ribosomal loci, which is the average observed for the genus based on data of currently known bifidobacterial species (Table S2). Remarkably, each of the analysed mammalian species appear to harbour an average of 16 novel bifidobacterial species with an average abundance, when present, of 1.66 % with respect to the overall bifidobacterial population (Fig. 2). Evaluation of the occurrence of the putative novel taxa among the analysed mammals failed to reveal specific distribution patterns linked to host taxonomy or diet, yet again highlighting that *Primates*, *Cercopithecidae* and *Hominidae*, as well as mammalian families encompassing domesticated species harbor the highest number of (novel) bifidobacterial species (Fig. 2).

Moreover, while six novel taxa (new_taxa_4, 10, 13, 26, 34 and 50) were shown to be prevalent at an average of 63.9 % of the mammalian species (ranging from 53.7 % to 82.1 %), none of these exhibit a relative abundance higher than 4.7 % respect to the overall bifidobacterial population. In fact, evaluation of the average relative abundance of the 89 putative novel bifidobacterial taxa, calculated including only data of the mammalian species where they are detected, revealed that only

new_taxa_74 and new_taxa_20 were shown to elicit an average relative abundance over 8 % (17.9 % and 12.4 %, respectively) (Fig. 2). These findings indicate that these suspected, and as yet undescribed, novel members of the genus *Bifidobacterium* are widespread among mammals, though their presence has until now remained undetected, probably due to their low abundance (Fig. 2). To validate the *in silico* identification of novel taxa, we isolated and genomically characterized two of these novel bifidobacterial species by means of a previously described cultivation-based approach (Turrioni et al 2009a, Turrioni et al 2012) (Table S5) (Supplementary Text).

Co-variance of bifidobacterial species among mammals. Availability of profiling data for all known and putative novel bifidobacterial species allowed the evaluation of their co-occurrence or co-exclusion across the mammalian tree of life by means of the Pearson correlation index (Table S6). Statistically significant (p -value < 0.05) co-variance data was subsequently used to construct a force-driven network (Fig. 3). Interestingly, significant (p -value < 0.05) co-exclusion appears to be very limited among members of the *Bifidobacterium* genus and was only observed among the ubiquitous *B. adolescentis*, *B. longum* and *B. pseudolongum* taxa, probably due to their high prevalence in the mammalian gut environment, as underlined by the collected profiling data (see above) (Table S6). In contrast, all other co-variance interactions with p -value < 0.05 were observed to be co-occurrences, which supports the idea of extensive co-evolution among bifidobacterial members that are present in the mammalian gut microbiota (Table S6).

Furthermore, the constructed force-driven co-variance network highlights the presence of eight clusters of bifidobacterial species with a high degree of co-occurrence among the analysed mammals (Fig. 3). Notably, four clusters involve just a small number of species (≤ 8), while the four remaining clusters were formed due to extensive co-occurrence between 16, 23, 24 and 26 species (Fig. 3). In contrast, very limited co-occurrence interactions (≤ 5) were shown to exist between different clusters (Fig. 3), with the exception of two clusters that were linked by 18 connections, of which 3, 4 and 6 involve new_taxa_60, new_taxa_10 and new_taxa_78, respectively (Fig. 3). Notably, these data indicate that bifidobacteria have developed specific sets of co-operative behaviour between co-colonizers which act as powerful evolutionary drivers in the mammalian gut microbiota. This assumption is supported by previous studies that have reported the existence of cross-feeding between specific bifidobacterial species (Milani et al 2015a, Turrioni et al 2016).

In contrast, correlation analyses between the presence/absence of bifidobacterial species and the taxonomy or dietary life styles (vegetarian, carnivorous or omnivorous) of the mammalian hosts failed to identify any statistically significant correlation. This finding reinforces the notion that bifidobacterial species did not evolve individually toward adaptation to specific environments but instead relied on microbe-microbe co-operation to achieve colonization of a wider range of different hosts.

Exploring vertical (mother-to-child) transmission of bifidobacteria. Vertical transmission of gut microbiota members from mother to offspring during delivery and subsequent

parental care is believed to play a key role in colonization and persistence of bacteria in humans (Milani et al 2015b). In order to evaluate transmission of specific bacterial phylotypes among mammals, we collected faecal samples of four human mother-child and 12 animal mother-child pairs, together representing seven different mammalian species and five mammalian orders, and encompassing humans and domesticated animals (Table S7). In addition, we re-analyzed two datasets from previous work (Milani et al 2015b), listed in Table S7 as Human_5 and Human_6 mother-child dyads. Re-analysis yielded improved results due to the use of the updated version of our extended, customized bifidobacterial ITS database. Where possible, we also sampled the milk of the mother and faeces of her multiple newborns (Table S7). The resulting qPCR evaluation of the relative abundance of the genus *Bifidobacterium* showed that, despite considerable variability among samples, in mammals bifidobacteria represent on average 0.7 % and 8.9 % of the whole faecal microbiota of sampled mothers and corresponding offspring, respectively (Fig. S3). Interestingly, comparison of data collected from human and non-human species revealed that, on average, human infants harbor a higher abundance of bifidobacteria (32.3 % compared to 0.7 % of non-human offspring, ANOVA p-value < 0.001) (Fig. S3). This finding reflects previous observations that human milk differs in oligosaccharides composition from that of other mammals. More specifically, the milk of humans contains higher abundance of type I oligosaccharides, i.e. presenting Lacto-N-biose type I structures, respect to milk from non-primate

mammals (Albrecht et al 2014, Urashima et al 2012).

Taxonomic classification of the predicted ITS-based OTUs allowed us to reconstruct the composition of the bifidobacterial population in the analyzed mother-child pairs and to identify phylotypes that they had in common (Fig. S4). Interestingly, the large majority such common OTUs in mother-offspring pairs correspond in all cases by a single species constituting on average 82.9 % of the profile retrieved for the pool of common OTUs (Fig. S4). This finding suggests that specific members of the microbiota possess enhanced capabilities supporting vertical transmission, including the ability to reach a high relative abundance in the mother and/or a genetic repertoire that supports transmission and subsequent colonization.

The retrieved species-level profiles of the offspring were also implied in the construction of a Principal coordinate analysis (PCoA) plot based on Bray-Curtis dissimilarity (Fig. S5). This analysis revealed that siblings born in the same litter from the same mother cluster together, thus indicating the development of a similar gut bifidobacterial population, as also confirmed by bar plot representation of the taxonomic profiles (Fig. S4). Homing in on the phylotype level, analysis of the retrieved bifidobacterial ITS-based OTUs (as calculated based on 100 % identity) revealed the presence of identical phylotypes in faecal samples of mother and corresponding offspring for all analyzed mammalian families (Fig. 4). Notably, the OTUs shared by mother-child dyads represent an average of 69.4 % of the total bifidobacterial population of the corresponding single or multiple offspring (Fig. 4). This demonstrates extensive

mother-child transmission of gut bifidobacterial phylotypes among mammalian representatives, and underlines the key role exerted by vertical transmission in defining the biodiversity and composition of the bifidobacterial population (and perhaps other microbiota components) colonizing the newborn gut.

Furthermore, the most represented bifidobacterial taxa in the shared OTUs pool of each mother-offspring pair were observed to represent, on average, the 81.7 % and 92.3 % of the total bifidobacterial population in the respective mother and offspring samples (Fig. 4).

Moreover, evaluation of the three most abundant OTUs of the mother representing the shared OTUs pool revealed that they constitute, on average, 64.7 % and 81.1 % of the population of their species in the mothers and their children, respectively. (Fig. 4). Such findings support the hypothesis that competition for vertical transmission does not only occur at species level, but also at phylotype level, meaning that only the small number of phylotypes for each species able to reach high relative abundance will be successfully transmitted from mother to child.

Role of milk as modulator of gut microbiome development. Production of milk as the sole infant nutrition during the first weeks and months following birth is one of the distinctive features of Mammalia. Moreover, mammary glands have been proposed to act as a bacterial reservoir that supports transmission of infant gut microbial colonizers during suckling (Rodriguez 2014).

The availability of 11 milk samples from representatives of four mammalian orders and six species (Table S7) showed the presence of bifidobacteria in all milk samples, thus suggesting

a specific adaptation of bifidobacteria to colonize and/or persist in the mammary gland (Fig. S3) ((Supplementary Text). Moreover, ITS profiling of milk samples allowed us to assess the ability of milk to act as a vector for the migration of gut bacteria from mother to offspring. Similar to previous observations for human samples (Martin et al 2012, Milani et al 2015b), species level profiling of mother's milk and faecal samples, as well as of corresponding newborn faecal samples revealed the presence of apparently identical bifidobacterial strains in each of the analyzed mammalian mother-child dyads. Our analysis also indicates that the taxonomic profile of the phylotypes that are common between mother-offspring resemble, in terms of composition and species abundance, that of OTUs commonly found in offspring-milk and mother-milk samples (Fig. S6). In fact, the OTUs shared by milk and offspring represent on average 74.6 % of the bifidobacterial population of the new-born, while 58.0 % is also shared by the mother (Fig. S7). Intriguingly, these data further substantiate the notion of maternal milk (and perhaps the mammary gland) being a reservoir of (bifido)bacteria which may be transmitted to the suckling infant.

Isolation and genomic characterization of two phylotypes corresponding to OTUs shared by Human_2 and Cattle_2 families, i.e. identified in both mother and her offspring, confirmed the *in silico* prediction. In fact, isolates obtained from mother and the respective child, and corresponding milk sample, displayed an ANI value >99.99 % as well as 100% ITS identity (Supplementary Text). Altogether, data obtained by dissection of the genome of these two isolates (Table S8) and growth experiments on HMO and HMO-derived

sugars (Fig. S7) indicated that (these vertically transmitted strains have evolved to use carbon sources typically present in milk, thus pointing at mammary glands as an effective reservoir and milk as an optimal vector for vertical transmission of (seemingly specifically adapted) bifidobacterial members of the gut microbiota from mother to offspring.

Conclusions

Bifidobacteria represent a prototype for the investigation of the ecological and functional roles exerted by the mammalian gut microbiota, especially during the crucial developmental phase immediately following birth when gut colonization takes place.

Nevertheless, current ecological knowledge on bifidobacteria is essentially restricted to the human gut, and in particular the period during which the infant's nutrition is exclusively milk (Avershina et al 2016, Martin et al 2016, Nuriel-Ohayon et al 2016, Roger et al 2010, Turrone et al 2012, Yassour et al 2016). In the current study, we explored the bifidobacterial ecology in 67 mammalian species, representing an exhaustive array of species within this class of animals. Notably, our analyses demonstrate that bifidobacteria are universally distributed across the mammalian branch of the tree of life irrespective of host's taxonomy or diet, thus revealing that these gut commensals have evolved a broad colonization potential, rather than that they engaged in host-specific specialization.

Collection and analysis of fecal samples from wild relatives of common domesticated animals was performed to explore the effect of domestication and close contact with humans. The obtained data clearly indicate that domestication of animals has

promoted a genetic diversification which followed or accompanied such human interventions. Notably, the comparison of the bifidobacterial biogeography between domesticated vs. wild animals revealed an increase in the number of colonizing species following domestication. Moreover, further exploration of the impact of artificial selection and interaction with humans was also performed by means of profiling of seven dog breeds. These data revealed that the composition of the bifidobacterial gut population of dog breeds differs from that of wolves, and corresponds to their phylogenetic distance (as based on microsatellite typing. Even if profiling of bifidobacteria across mammals excluded a major role of host genetics, data retrieved from wild specimens and dog breeds indicates that close contact with humans has progressively shaped the bifidobacterial population. In particular, the increase in species richness may have resulted from prolonged co-existence with humans (characterized by high bifidobacterial biodiversity) suggestive of a niche shift. Another key ecological feature of bifidobacteria is that they are numerically relevant during the first weeks and months following birth in the human gut (Avershina et al 2016), a feat that appears to be a general trait of bifidobacteria across the mammalian tree of life. Altogether, these data, as well as the presence of bifidobacteria in the distinctive mammalian secretory fluid, milk, suggest an intriguing genetic specialization of bifidobacteria to the mammalian gut, possibly re-enforced by the direct transmission of bifidobacterial consortia from mother to offspring. Previous studies have shown that certain human-derived bifidobacterial species are shared between

mothers and their children (Avershina et al 2016, Milani et al 2015b), and our findings imply that this scenario extends to many other mammals. Moreover, we show that the gut bifidobacterial population of siblings is rather similar, implying a key role exerted by the mother in defining the gut bifidobacterial community. The establishment of the gut microbiota by means of vertical transmission points to a key functional role exploited by the transmitted microorganisms in the gut of the host. In this context, the role exerted by the gut microbiota in priming and maturing the immune system, energy harvesting and nervous system development in human neonates is well documented (Cox et al 2014, Diaz Heijtz et al 2011, Matsuki et al 2016).

Thus, bifidobacteria represent ideal prototypes to investigate gut microbiota establishment and development. The developmental trajectory of bifidobacteria in humans includes a decrease in the relative abundance during adult and elderly stages of life, but their maintenance in the core microbiota, even at very low levels, has been noticed in many metagenomic studies involving healthy individuals.

Previous metagenomic analyses revealed the high prevalence of bifidobacteria in the human milk microbiome (Collado et al 2009, Martin et al 2009, Turrone et al 2011), even though the origin of these microbes is unclear. In this study, we have shown that bifidobacteria are (sometimes highly) prevalent in the milk microbiome not only in humans, as previously described (Collado et al 2009, Martin et al 2009), but also in various other mammalian species. Furthermore, bifidobacterial phylotypes that appear to be commonly present in stool samples of mother-offspring dyads, were

sometimes also detected in the corresponding mother's milk sample, thus reinforcing the notion that milk may also serve as an inoculum for breast-feeding infants (Jeurink et al 2013, Martin et al 2009, Martin et al 2012).

The genomic characterization of two vertically transmitted bifidobacterial strains (identified in the milk microbiome as well as in stool samples of the corresponding mother-infant pair) revealed a rich repertoire of genes encoding glycosyl hydrolases, as well as transporters, supporting an active metabolism of lactose and milk oligosaccharides in a strain isolated in humans, whose milk is rich of these glycans. In contrast, limited adaptation has been observed in a strain isolated in cattle milk, where these complex sugars are less abundant. These results indicate that certain bifidobacterial taxa have genetically adapted to specific milk environments, thus resulting in improved fitness for colonization of mammary glands and the gut of breast-fed newborns. Notably, genomic characterization of additional isolates is needed to confirm that genetic adaptation to the milk environment is a distinctive feature of vertically transmitted strains. ITS profiling data revealed that the genus *Bifidobacterium* may exist of many more species than the currently 45 taxonomically recognized species, with 89 predicted novel (sub)species identified in this work, two of which were isolated and genetically characterized. Notably, and in contrast to what has previously been proposed (Lugli et al 2014, Milani et al 2014a), the broad distribution of these 134 bifidobacterial sub (species) in mammals does not reveal any host-specific specialization behavior, nor does it seem to correspond to the taxonomy or diet of the host. Arguably, bifidobacteria-host

specialization may have occurred at strain- rather than species- level , as also pointed out by a previous study (Oh et al 2010) concerning diversification of *Lactobacillus reuteri* strains isolated in different hosts. These data together with the identification of common complex bifidobacterial populations and eight groups of bifidobacterial taxa showing strict co-occurrence demonstrate that bifidobacteria have evolved to achieve extensive co-colonization which may depend on functional co-operation.

Acknowledgments

Part of this research is conducted using the High Performance Computing (HPC) facility of the University of Parma.

We thank GenProbio srl for financial support of the Laboratory of Probiogenomics. This work was supported by the EU Joint Programming Initiative – A Healthy Diet for a Healthy Life (JPI HDHL, <http://www.healthydietforhealthylife.eu/>) to DvS (project reference 15/JP/HDHL/3280) and to MV (MIUR Prot. 001988). DvS is a member of The APC Microbiome Institute, which is funded by Science Foundation Ireland (SFI), through the Irish Government’s National Development Plan (Grant number SFI/12/RC/2273).

We thank “Fondazione Bioparco di Roma” (Viale del Giardino Zoologico 20, 00197, Rome, Italy), “Zoomarine” (via dei Romagnoli, 00040 Torvaianica, Pomezia, Rimini, Italy), “Giardino zoologico di Pistoia” (Via Pieve a Celle, 160, 51100, Pistoia, Italy) and “Centro Tutela e Ricerca Fauna Esotica e Selvatica Monte Adone” (Via Brento, 9, 40037, Sasso Marconi, Bologna, Italy) for their support in the recruitment of the samples.

Author Contributions

C.M. designed all experiments, performed bioinformatics analyses and wrote the manuscript. G.A.L. and L.M. performed genomics bioinformatics analyses. M.V and D.v.S. conceived the study, revised and approved the manuscript. M.M. performed sample collection qPCR experiments and bifidobacterial ITS profiling. S.D., F.T. and C.F. participated in qPCR experiments and bifidobacterial ITS profiling. K.J. performed growth analyses on HMOs. M.C.O. participated in data analysis. All authors reviewed and approved the final version of the manuscript.

Additional Information

Competing financial interests: The authors declare no competing financial interests.

References

- Albrecht S, Lane JA, Marino K, Al Busadah KA, Carrington SD, Hickey RM *et al* (2014). A comparative study of free oligosaccharides in the milk of domestic animals. *The British journal of nutrition* **111**: 1313-1328.
- Arbolea S, Watkins C, Stanton C, Ross RP (2016). Gut Bifidobacteria Populations in Human Health and Aging. *Frontiers in microbiology* **7**: 1204.
- Avershina E, Lundgard K, Sekelja M, Dotterud C, Storro O, Oien T *et al* (2016). Transition from infant- to adult-like gut microbiota. *Environmental microbiology* **18**: 2226-2236.
- Bastian MH, S.; Jacomy, M. (2009). Gephi: an open source software for exploring and manipulating networks. *International AAAI Conference on Weblogs and Social Media*.
- Bunesova V, Vlkova E, Rada V, Killer J, Musilova S (2014). Bifidobacteria from the gastrointestinal tract of animals: differences and similarities. *Beneficial microbes* **5**: 377-388.
- Cardona S, Eck A, Cassellas M, Gallart M, Alastrue C, Dore J *et al* (2012). Storage conditions of intestinal microbiota matter in metagenomic analysis. *BMC microbiology* **12**: 158.
- Choo JM, Leong LE, Rogers GB (2015). Sample storage conditions significantly influence faecal microbiome profiles. *Scientific reports* **5**: 16350.
- Collado MC, Delgado S, Maldonado A, Rodriguez JM (2009). Assessment of the bacterial diversity of breast milk of healthy women by quantitative real-time PCR. *Letters in applied microbiology* **48**: 523-528.
- Cox LM, Yamanishi S, Sohn J, Alekseyenko AV, Leung JM, Cho I *et al* (2014). Altering the intestinal microbiota during a critical developmental window has lasting metabolic consequences. *Cell* **158**: 705-721.
- De Filippo C, Cavalieri D, Di Paola M, Ramazzotti M, Poullet JB, Massart S *et al* (2010). Impact of diet in shaping gut microbiota revealed by a comparative study in children from Europe and rural Africa. *Proceedings of the National Academy of Sciences of the United States of America* **107**: 14691-14696.
- Diaz Heijtz R, Wang S, Anuar F, Qian Y, Bjorkholm B, Samuelsson A *et al* (2011). Normal gut microbiota modulates brain development and behavior. *Proceedings of the National Academy of Sciences of the United States of America* **108**: 3047-3052.
- Fouhy F, Deane J, Rea MC, O'Sullivan O, Ross RP, O'Callaghan G *et al* (2015). The effects of freezing on faecal microbiota as determined using MiSeq sequencing and culture-based investigations. *PLoS One* **10**: e0119355.
- Harper CG, Whary MT, Feng Y, Rhinehart HL, Wells RS, Xu S *et al* (2003). Comparison of diagnostic techniques for *Helicobacter cetorum* infection in wild Atlantic bottlenose dolphins (*Tursiops truncatus*). *Journal of clinical microbiology* **41**: 2842-2848.
- Jeurink PV, van Bergenhenegouwen J, Jimenez E, Knippels LM, Fernandez L, Garssen J *et al* (2013). Human milk: a source of more life than we imagine. *Beneficial microbes* **4**: 17-30.
- Lax S, Smith DP, Hampton-Marcell J, Owens SM, Handley KM, Scott NM *et al* (2014). Longitudinal analysis of microbial interaction between humans and the indoor environment. *Science* **345**: 1048-1052.

Ley RE, Hamady M, Lozupone C, Turnbaugh PJ, Ramey RR, Bircher JS *et al* (2008). Evolution of mammals and their gut microbes. *Science* **320**: 1647-1651.

Lugli GA, Milani C, Turrone F, Duranti S, Ferrario C, Viappiani A *et al* (2014). Investigation of the evolutionary development of the genus *Bifidobacterium* by comparative genomics. *Applied and environmental microbiology* **80**: 6383-6394.

Lugli GA, Milani C, Mancabelli L, van Sinderen D, Ventura M (2016). MEGAnnotator: a user-friendly pipeline for microbial genomes assembly and annotation. *FEMS microbiology letters*.

Martin R, Jimenez E, Heilig H, Fernandez L, Marin ML, Zoetendal EG *et al* (2009). Isolation of bifidobacteria from breast milk and assessment of the bifidobacterial population by PCR-denaturing gradient gel electrophoresis and quantitative real-time PCR. *Applied and environmental microbiology* **75**: 965-969.

Martin R, Makino H, Cetinyurek Yavuz A, Ben-Amor K, Roelofs M, Ishikawa E *et al* (2016). Early-Life Events, Including Mode of Delivery and Type of Feeding, Siblings and Gender, Shape the Developing Gut Microbiota. *PLoS One* **11**: e0158498.

Martin V, Maldonado-Barragan A, Moles L, Rodriguez-Banos M, Campo RD, Fernandez L *et al* (2012). Sharing of bacterial strains between breast milk and infant feces. *Journal of human lactation : official journal of International Lactation Consultant Association* **28**: 36-44.

Matsuki T, Yahagi K, Mori H, Matsumoto H, Hara T, Tajima S *et al* (2016). A key genetic factor for fucosyllactose utilization affects infant gut microbiota development. *Nature communications* **7**: 11939.

Milani C, Lugli GA, Duranti S, Turrone F, Bottacini F, Mangifesta M *et al* (2014a). Genomic encyclopedia of type strains of the genus *Bifidobacterium*. *Applied and environmental microbiology* **80**: 6290-6302.

Milani C, Lugli GA, Turrone F, Mancabelli L, Duranti S, Viappiani A *et al* (2014b). Evaluation of bifidobacterial community composition in the human gut by means of a targeted amplicon sequencing (ITS) protocol. *FEMS microbiology ecology* **90**: 493-503.

Milani C, Andrea Lugli G, Duranti S, Turrone F, Mancabelli L, Ferrario C *et al* (2015a). Bifidobacteria exhibit social behavior through carbohydrate resource sharing in the gut. *Scientific reports* **5**: 15782.

Milani C, Mancabelli L, Lugli GA, Duranti S, Turrone F, Ferrario C *et al* (2015b). Exploring vertical transmission of bifidobacteria from mother to child. *Applied and environmental microbiology*.

Milani C, Turrone F, Duranti S, Lugli GA, Mancabelli L, Ferrario C *et al* (2015c). Genomics of the Genus *Bifidobacterium* Reveals Species-Specific Adaptation to the Glycan-Rich Gut Environment. *Applied and environmental microbiology* **82**: 980-991.

Milani C, Ticinesi A, Gerritsen J, Nouvenne A, Lugli GA, Mancabelli L *et al* (2016). Gut microbiota composition and *Clostridium difficile* infection in hospitalized elderly individuals: a metagenomic study. *Scientific reports* **6**: 25945.

Misic AM, Davis MF, Tyldsley AS, Hodkinson BP, Tolomeo P, Hu B *et al* (2015). The shared microbiota of humans and companion animals as evaluated from *Staphylococcus* carriage sites. *Microbiome* **3**: 2.

Nuriel-Ohayon M, Neuman H, Koren O (2016). Microbial Changes during Pregnancy, Birth, and Infancy. *Frontiers in microbiology* **7**: 1031.

O'Callaghan A, van Sinderen D (2016). Bifidobacteria and Their Role as Members of the Human Gut Microbiota. *Frontiers in microbiology* **7**: 925.

Oh PL, Benson AK, Peterson DA, Patil PB, Moriyama EN, Roos S *et al* (2010). Diversification of the gut symbiont *Lactobacillus reuteri* as a result of host-driven evolution. *The ISME journal* **4**: 377-387.

Parker HG, Kim LV, Sutter NB, Carlson S, Lorentzen TD, Malek TB *et al* (2004). Genetic structure of the purebred domestic dog. *Science* **304**: 1160-1164.

Rodriguez JM (2014). The origin of human milk bacteria: is there a bacterial entero-mammary pathway during late pregnancy and lactation? *Advances in nutrition* **5**: 779-784.

Roger LC, Costabile A, Holland DT, Hoyles L, McCartney AL (2010). Examination of faecal *Bifidobacterium* populations in breast- and formula-fed infants during the first 18 months of life. *Microbiology* **156**: 3329-3341.

Serafini F, Bottacini F, Viappiani A, Baruffini E, Turrone F, Foroni E *et al* (2011). Insights into physiological and genetic mupirocin susceptibility in bifidobacteria. *Applied and environmental microbiology* **77**: 3141-3146.

Song SJ, Lauber C, Costello EK, Lozupone CA, Humphrey G, Berg-Lyons D *et al* (2013). Cohabiting family members share microbiota with one another and with their dogs. *eLife* **2**: e00458.

Turrone F, Foroni E, Pizzetti P, Giubellini V, Ribbera A, Merusi P *et al* (2009a). Exploring the diversity of the bifidobacterial population in the human intestinal tract. *Applied and environmental microbiology* **75**: 1534-1545.

Turrone F, Marchesi JR, Foroni E, Gueimonde M, Shanahan F, Margolles A *et al* (2009b). Microbiomic analysis of the bifidobacterial population in the human distal gut. *The ISME journal* **3**: 745-751.

Turrone F, Foroni E, Serafini F, Viappiani A, Montanini B, Bottacini F *et al* (2011). Ability of *Bifidobacterium breve* to grow on different types of milk: exploring the metabolism of milk through genome analysis. *Applied and environmental microbiology* **77**: 7408-7417.

Turrone F, Peano C, Pass DA, Foroni E, Severgnini M, Claesson MJ *et al* (2012). Diversity of *Bifidobacteria* within the Infant Gut Microbiota. *PLoS One* **7**: e36957.

Turrone F, Milani C, Duranti S, Mancabelli L, Mangifesta M, Viappiani A *et al* (2016). Deciphering bifidobacterial-mediated metabolic interactions and their impact on gut microbiota by a multi-omics approach. *The ISME journal* **10**: 1656-1668.

Urashima T, Asakuma S, Leo F, Fukuda K, Messer M, Oftedal OT (2012). The predominance of type I oligosaccharides is a feature specific to human breast milk. *Advances in nutrition* **3**: 473S-482S.

Ushida K, Tsuchida S, Ogura Y, Toyoda A, Maruyama F (2016). Domestication and cereal feeding developed domestic pig-type intestinal microbiota in animals of suidae. *Animal science journal = Nihon chikusan Gakkaiho* **87**: 835-841.

van Opstal EJ, Bordenstein SR (2015). MICROBIOME. Rethinking heritability of the microbiome. *Science* **349**: 1172-1173.

Ventura M, Turrone F, Motherway MO, MacSharry J, van Sinderen D (2012). Host-microbe interactions that facilitate gut colonization by commensal bifidobacteria. *Trends in microbiology* **20**: 467-476.

Wang W, Zheng S, Sharshov K, Cao J, Sun H, Yang F *et al* (2016). Distinctive gut microbial community structure in both the wild and farmed Swan goose (*Anser cygnoides*). *Journal of basic microbiology*.

Yassour M, Vatanen T, Siljander H, Hamalainen AM, Harkonen T, Ryhanen SJ *et al* (2016). Natural history of the infant gut microbiome and impact of antibiotic treatment on bacterial strain diversity and stability. *Science translational medicine* **8**: 343ra381.

Yatsunenko T, Rey FE, Manary MJ, Trehan I, Dominguez-Bello MG, Contreras M *et al* (2012). Human gut microbiome viewed across age and geography. *Nature* **486**: 222-227.

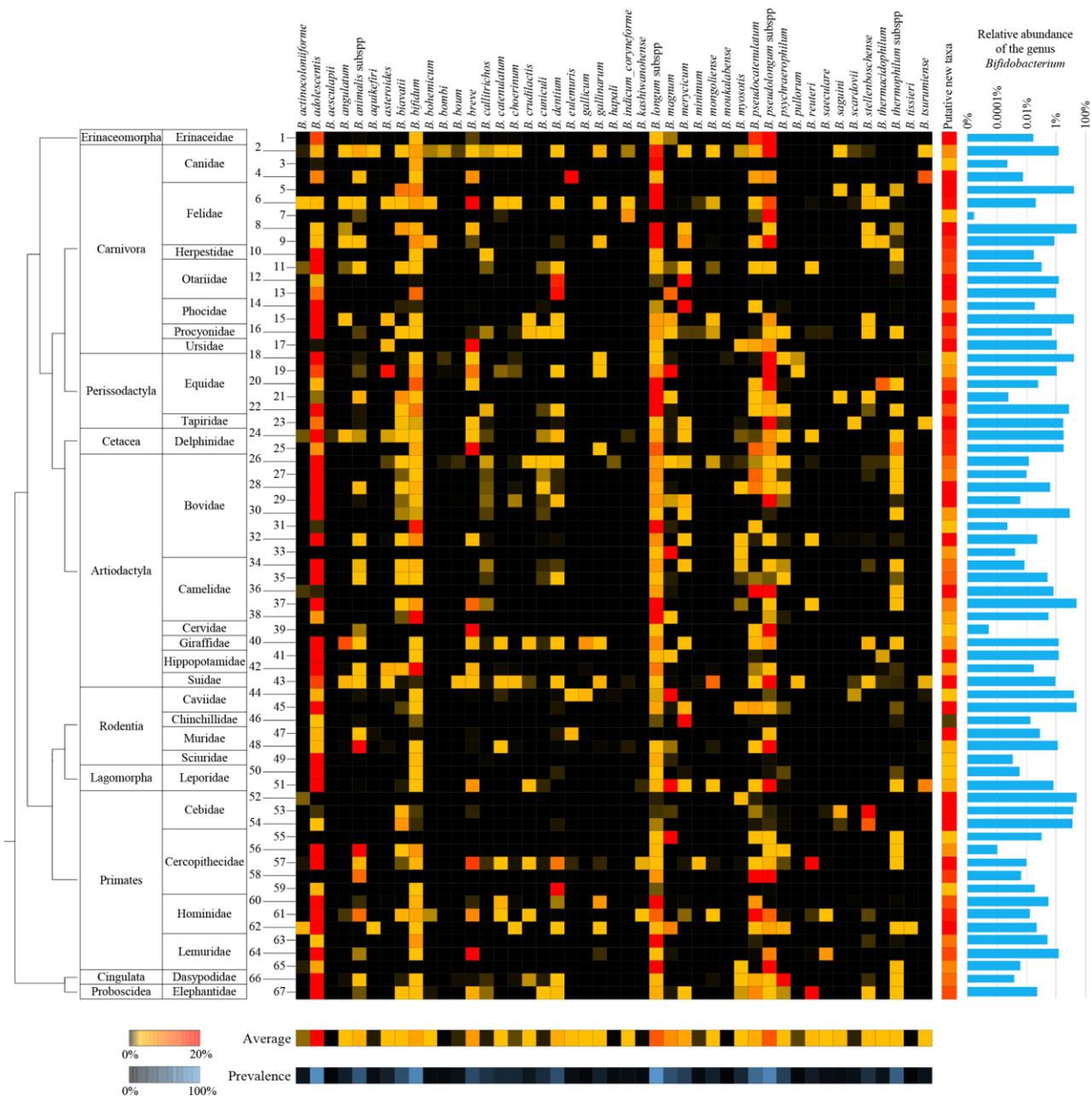


Figure 1. Taxonomic profile of the currently known 45 bifidobacterial species in the mammalian tree of life. The heat map shows the observed relative abundance of 45 known bifidobacterial species in the 291 analyzed samples of adult mammals numbered according to Table S9. On the left-hand side, the taxonomy of the samples at Order and Family level, as well as a phylogenetic tree of the mammalian Orders covered by this study is reported. On the right-hand side of the image, a heat map shows data regarding the relative abundance of putative novel bifidobacterial species respect to the total bifidobacterial population, while in the lower part two heat maps present the average abundance of the profiled bifidobacterial species respect to the total bifidobacterial population and their prevalence in the 291 mammalian samples. Bar plots display qPCR data on total bacterial load and bifidobacterial load for each sample in logarithmic scale.

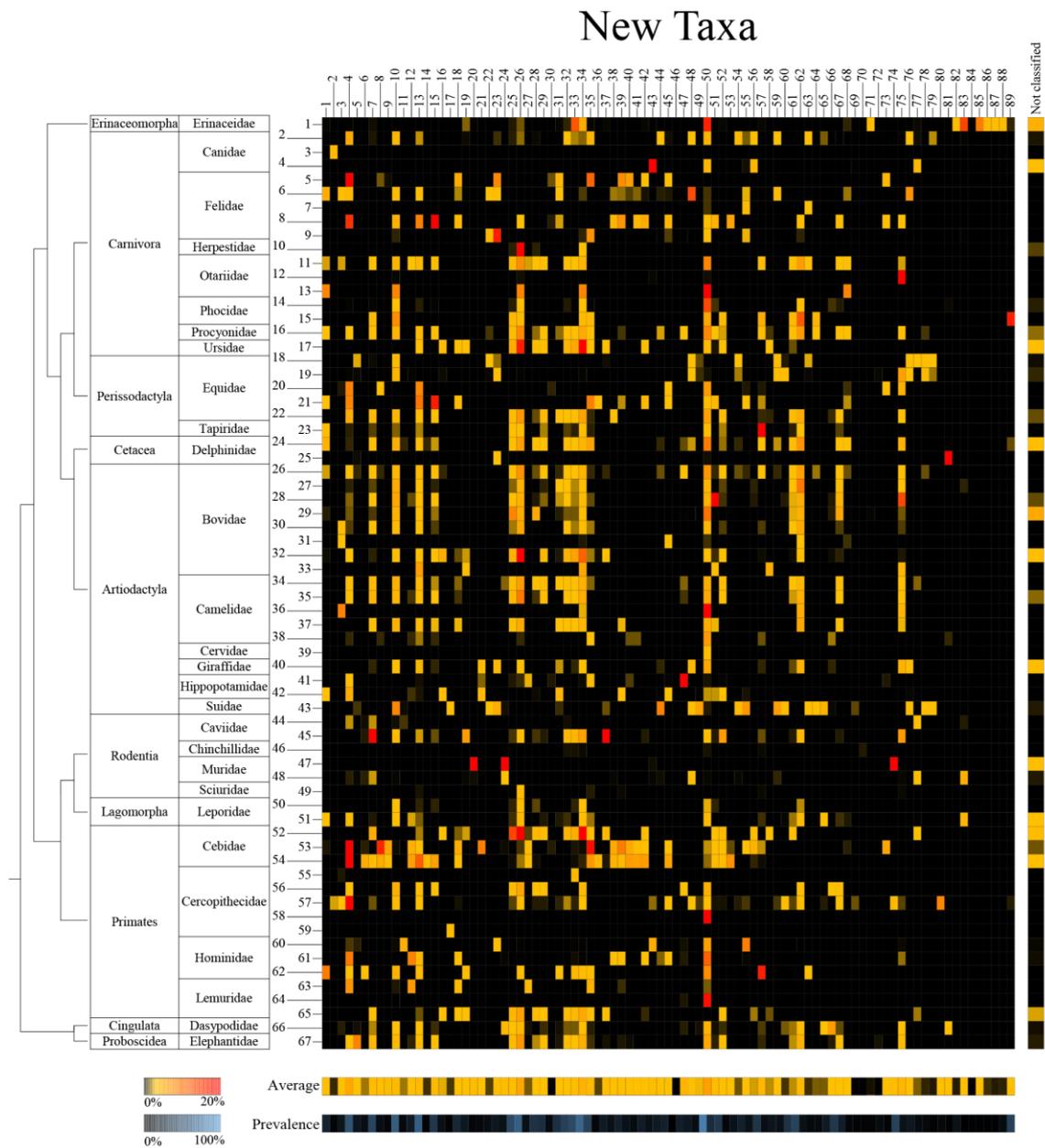


Figure 2. Taxonomic profile of 89 putative novel bifidobacterial species in the mammalian branch of the tree of life. The heat map shows the relative abundance respect to the total bifidobacterial population observed for the 89 putative bifidobacterial species, numbered progressively from 1 to 89, in the 291 analyzed samples of adult mammals numbered according to Table S9. On the left-hand side the taxonomy of the samples at Order and Family level is indicated, as well as a phylogenetic tree of the mammalian Orders covered by this study. The vertical heat map on the right-hand side shows data regarding the relative abundance of OTUs not classified as new bifidobacterial species but displaying similarity with other members of the *Bifidobacteriaceae* family, while in the lower part of the image two horizontal heat maps present the average abundance respect to the total bifidobacterial population of the profiled putative bifidobacterial species and their prevalence in the 291 mammalian samples.

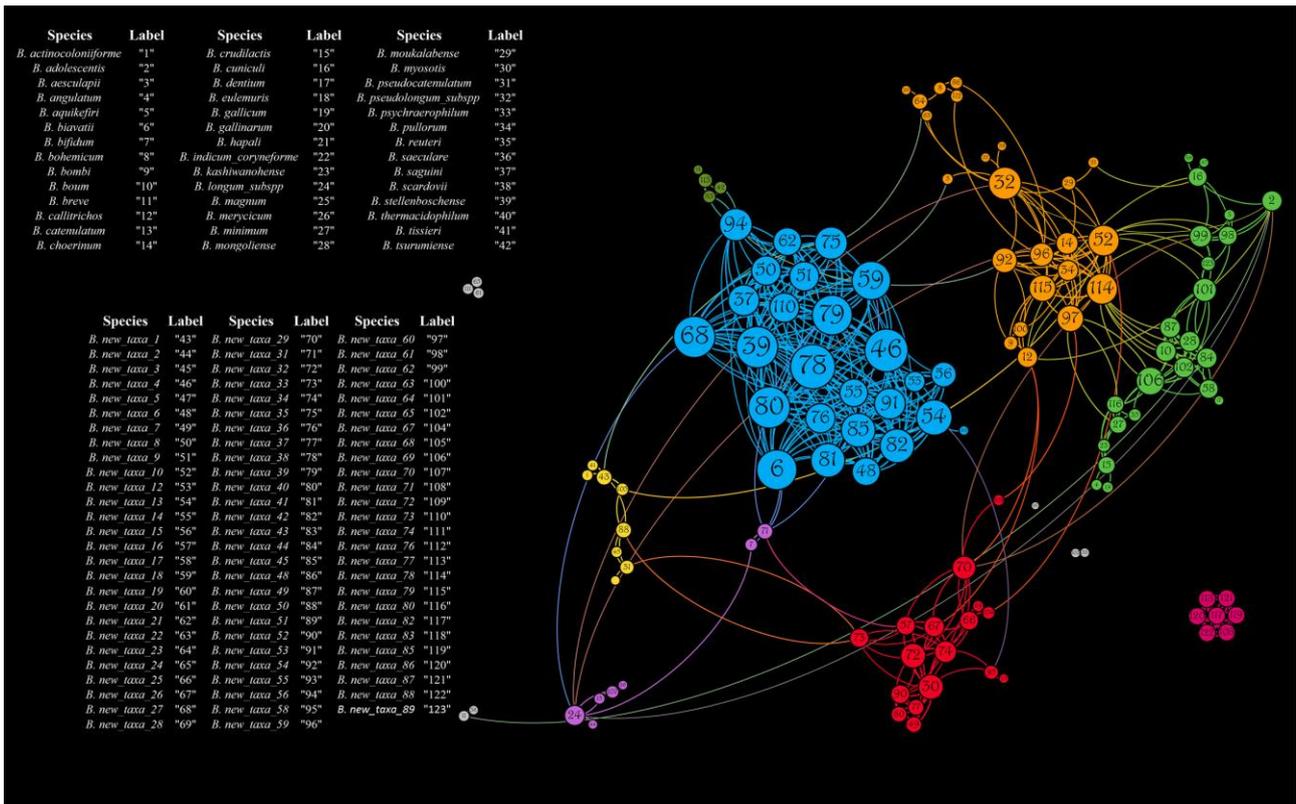


Figure 3. Co-variance network of the 45 known and 89 putative novel bifidobacterial species present in mammals. The force-driven network shows the predicted co-variances with p-values < 0.05 between the profiled bifidobacterial species. The node size is proportional to the number of co-variances, while the node color indicates the eight clusters observed.

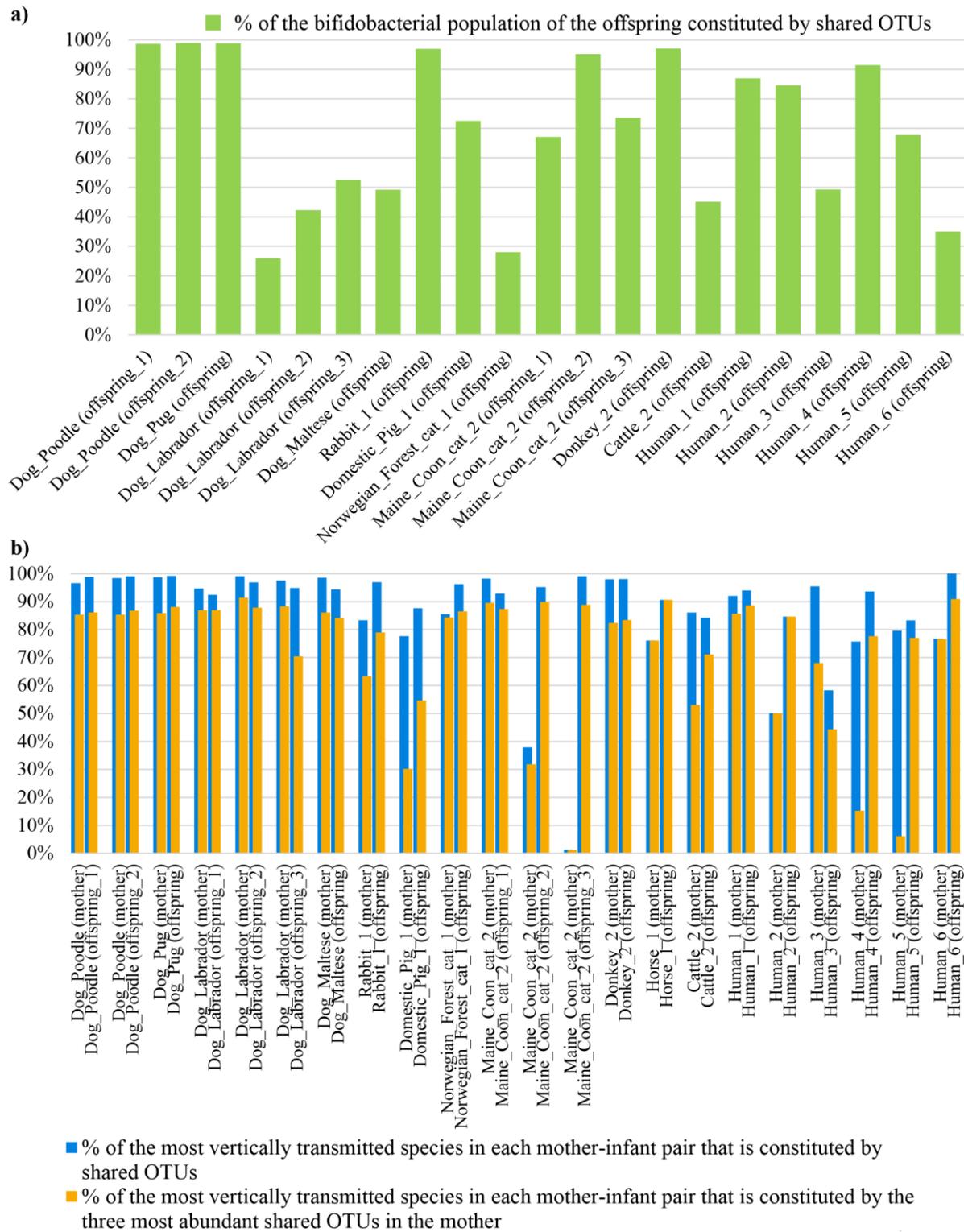


Figure 4. Impact of vertical transmission events on the offspring microbiota. Panel a shows a bar plot representing for each mother-offspring set the percentage of the total bifidobacterial population of the offspring that is constituted by OTUs also found in the mother. Panel b displays a bar plot depicting for each mother-offspring set: in blue the percentage of the most frequently observed vertically transmitted bifidobacterial strains as represented by OTUs that were in common between mother-offspring, and in orange the percentage of the most frequently observed vertically transmitted strains represented by the three most abundant OTUs from the mother that were also found in her offspring.

Chapter 4

Bifidobacterium vansinderenii sp. nov.,
isolated from faeces of emperor tamarin
(*Saguinus imperator*)

Duranti S[‡], Mangifesta M[‡], Lugli GA[‡], Turrone F, Anzalone R, Milani C,
Mancabelli L, Ossiprandi MC, Ventura M.

This chapter was published in:

Int J Syst Evol Microbiol. 2017 Oct;67(10):3987-3995. Epub 2017 Sep 12.

Reprinted with permission from CCC Republication.

‡ these authors contribute equally

***Bifidobacterium vansinderenii* sp. nov. isolated from faeces of emperor tamarin**

(*Saguinus imperator*).

Sabrina Duranti*¹, Marta Mangifesta*^{1,2}, Gabriele Andrea Lugli*¹, Francesca Turrone¹, Rosaria Anzalone¹, Christian Milani¹, Leonardo Mancabelli¹, Maria Cristina Ossiprandi³, Marco Ventura¹

Laboratory of Probiogenomics, Department of Chemistry, Life Sciences and Environmental Sustainability, University of Parma, Parma, Italy¹; GenProbio srl, Parma, Italy²; Department of Veterinary Medical Science, University of Parma, Parma, Italy³.

*These authors contributed equally

* Corresponding author. Mailing address for Marco Ventura laboratory of Probiogenomics, Department of Chemistry, Life Sciences and Environmental Sustainability, University of Parma, Parco Area delle Scienze 11a, 43124 Parma, Italy. Phone: ++39-521-905666. Fax: ++39-521-905604. E-mail: marco.ventura@unipr.it

Abstract

A novel *Bifidobacterium* strain Tam10B^T, i.e. LMG 30126^T, was isolated from Emperor tamarin (*Saguinus imperator*). Cells are Gram-positive, non-motile, non-sporulating, non-haemolytic, facultative anaerobic and fructose 6-phosphate phosphoketolase-positive. Phylogenetic analyses based on 16S rRNA genes as well as multilocus sequences (representing *hsp60*, *rpoB*, *dnaJ*, *dnaG* and *clpC* genes) and the core genome revealed that *Bifidobacterium* Tam10B^T exhibits close phylogenetic relatedness with *Bifidobacterium tissieri* DSM 100201^T. Comparative analysis of 16S rRNA gene sequences confirmed the phylogenetic results showing the highest gene sequence identity with strain *B. tissieri* DSM 100201^T (96.5 %). Furthermore, genotyping based on the genome sequence of Tam 10B and combined with phenotypic analyses, clearly shows that strain Tam10B^T is distinct from each of the type strains of the so far recognized *Bifidobacterium* species. The type strain Tam10B^T (=LMG 30126^T=CCUG 70655^T) represents a novel species, for which the name *Bifidobacterium vansinderenii* sp. nov is proposed.

Keywords: Bifidobacteria, *Bifidobacterium*, *Bifidobacterium vansinderenii*, Emperor tamarin, *Saguinus imperator*, Next generation sequencing.

Introduction

Bifidobacteria is one of the most important bacterial group belonging to the *Actinobacteria* phylum (1, 2). Members of the *Bifidobacterium* genus represent Gram-positive, anaerobic, non-spore forming and non-motile bacteria. Currently, the genus *Bifidobacterium* includes 57 taxa, representing 48 species and 9 subspecies (3-7).

Notably, bifidobacteria have been isolated from six different ecological niches, including the gastrointestinal tract (GIT) of humans, (other) animals and insects, the oral cavity, human blood, sewage, raw milk and water kefir fermentation process (7-11). Bifidobacteria are considered non-pathogenic and purported to elicit beneficial health effects onto their host (9, 10). As part of a project aimed at exploring the gut microbiota biodiversity in animals, a small number of putative novel taxa belonging to the genus *Bifidobacterium* were identified (unpublished data). Among the analysed mammalian hosts, we included the primate emperor tamarin (*Saguinus imperator*), a primate that belongs to the Callitrichidae family, *Saguinus* genus, and emperor species, and that lives in the Amazonian lowland. Tamarin follow a broad dietary regime, which includes fruits, flowers, nectar, plant exudates (gums, saps and latex) and animal prey (including frogs, snails, lizards, spiders and insects). This work explored the presence and diversity of cultivable bifidobacterial strains from a faecal sample of this primate by means of a cultivation-dependent approach and *de novo* genome sequencing. The (*Saguinus imperator*) faecal sample was serially diluted with Phosphate Buffered Saline (PBS) solution supplemented with cysteine hydrochloride (0.5 g/L), and one ml of each dilution was inoculated in de Man-Rogosa-Sharpe (MRS) agar medium supplemented with mupirocin (50 µg/ml). Plates were incubated under anaerobic conditions (2.99 % H₂, 17.01 % CO₂, and 80 % N₂) in a chamber (Concept 400; Ruskin) at 37°C for 48 h. After incubation, morphologically different colonies were randomly picked and re-streaked in order to isolate purified bacterial strains. A total of seven

isolates were thus obtained, which were cultivated under anaerobic conditions as mentioned above. Chromosomal DNA of the isolated strains was extracted as described previously (12). The isolates were subjected to molecular typing through the use of Enterobacterial Repetitive Intergenic Consensus sequences (ERIC) PCR. ERIC PCR was carried out using ERIC-1 (5'-ATGTAAGCTCCTGGGGATTCAC-3') and ERIC-2 (5'-AAGTAAGTGACTGGGGTGAGCG-3') primers (13) following the procedure previously described (13) (Fig. S1). During this study, an isolate, called Tam10B^T, belonging to probably unknown *Bifidobacterium* species was identified. Genome analysis, molecular characterization, growth parameters and morphological profiles were evaluated for Tam10B^T. The genome of Tam10B^T was decoded through a Next Generation Sequencing (NGS) approach using a MiSeq platform (Illumina, United Kingdom) at GenProbio srl (Parma, Italy) following a previously described protocol (3, 14, 15). The generated reads were depleted of adapter sequences, quality filtered and assembled through the MEGAnnotator pipeline (16). Furthermore, open reading frame (ORF) identification and functional annotation of ORFs was performed as previously described (16). The genome sequence of Tam10B^T consists of 3,111,005 bp, corresponding to 2,522 predicted protein-encoding ORFs. The general genome features of Tam10B^T are summarized in Table 1. *In silico* analyses of the genome of Tam10B^T allowed the estimation of its G+C content to be 62.5 % (Table 1), a value consistent with the G+C content range of the *Bifidobacterium* genus (17). This Whole Genome Shotgun project has been

deposited at DDBJ/ENA/GenBank under the accession NEWD00000000. The version described in this paper is version NEWD01000000. We investigated the phylogenetic relatedness of Tam10B^T strain with other currently recognized bifidobacterial (sub)species by inferring the nucleotide sequences of 16S rRNA, *hsp60*, *rpoB*, *dnaJ*, *dnaG* and *clpC* genes as well as the genes constituting the core genome of the genus *Bifidobacterium* (1, 18, 19). Such analyses allowed the reconstruction of 16S rRNA gene-based phylogenetic tree (Fig. 1), a multigene-based tree (Fig. 2) and a phylogenomic tree (Fig. 3), the latter based on the core genome sequences of the *Bifidobacterium* genus (19). In order to identify the bifidobacterial core genome sequences of the *Bifidobacterium* genus, a comparative genome analysis involving the known 55 bifidobacterial (sub)species, including Tam10B^T, was performed to unveil core genome sequences. The ORF content of each bifidobacterial genome was organized in functional gene clusters using the gene family (GF) method of the pangenome analysis pipeline (PGAP) (20), involving the basic local alignment search tool (BLAST) (E value cutoff of 1×10^{-5} and 50 % identity across at least 80 % of sequence length). Sequences were then clustered into protein families, using a graph theory-based Markov clustering algorithm (MCL) (21). This analysis allowed the identification of 262 clusters of orthologous genes (COGs) shared by all 55 (sub)species, representing the core genome of currently sequenced *Bifidobacterium* representatives. Thus, a phylogenetic core genome tree was constructed based on the concatenation of 237 protein sequences that represent the *Bifidobacterium* core genome sequences with the

exclusion of paralogs identified in each genome (Fig. 3). Each phylogenetic tree showed the same phylogenetic positioning of Tam10B^T with respect to the other members of the *Bifidobacterium* genus, revealing a close phylogenetic relatedness of Tam10B^T with *Bifidobacterium tissieri* DSM 100201^T (Figs. 1, 2 and 3). Furthermore, the bifidobacterial pangenome analysis involving all the current members of the *Bifidobacterium* genus allowed the identification of 550 unique genes belonging to Tam10B^T genome (Fig. 4). Comparative genomics analyses involving both Tam10B^T and *B. tissieri* DSM 100201^T chromosome sequences, highlighted the occurrence 794 genes in the genome of Tam10B^T that appear to lack any orthologous in *B. tissieri* DSM 100201^T chromosome (Fig. 4). Thus, such findings further supported the genetic diversity of Tam10B^T respect to any of the current recognized bifidobacterial species. In order to investigate the genetic difference between Tam10B^T and other *Bifidobacterium* strains, a nucleotide identity value was evaluated based on gene sequences that are considered to act as molecular clocks in bacterial taxonomy (18, 22, 23) including 16S rRNA, *hsp60*, *rpoB*, *dnaJ*, *dnaG* and *clpC* (Table S1). These analyses revealed that the 16S rRNA gene of Tam10B^T exhibits the highest level of sequence identity (96.5 %) to that of *B. tissieri* DSM 100201^T (Table S1). Notably, a 97 % identity value between 16S rRNA genes is considered to represent the cut-off value for bacterial species assignment (24). Moreover, the five housekeeping genes selected for sequence comparison highlight the highest identity values (ranging from 93 % to 99.4 %) with the corresponding *B. tissieri* DSM 100201^T genes (Table S1). Furthermore, the genetic similarity at

genomic level of strain Tam10B^T with respect the other currently recognized bifidobacterial (sub)species was calculated by Average Nucleotide Identity (ANI) analysis (25). This analysis showed that the Tam10B^T genome displays the highest level of identity (89 %) to the chromosomal sequences of *B. tissieri* DSM 100201^T (Table S2). In this context it should be noted that two strains displaying an ANI value ≤ 95 % are considered to belong to two distinct species (26). Tam10B^T cells possess a rod-shape morphology as observed by phase-contrast microscopy (Fig. 5). Optimal growth conditions of Tam10B^T cells compared with *B. tissieri* DSM 100201^T and *B. stollenboschense* DSM 23968^T, were determined in MRS broth under anaerobic growth conditions. Growth was tested at different pH values (at pH 3.5, 4.0, 4.5, 5.0, 5.5 and 6.0) and in aerobic and anaerobic conditions at different temperatures, such as 20, 25, 30, 35, 37, 40 and 45 °C (Table 2). Gram staining was performed with LIVE BacLight™ Bacterial Gram Stain Kit (ThermoFisher Scientific) using cells grown on MRS broth at 37°C under anaerobic conditions for 48h the manufacturer's instructions (ThermoFisher Scientific). Fermentation profiles experiments on different carbon sources, including arabinose, arabinogalactan, cellobiose, fructo-oligosaccharides, fructose, fucose, galactose, galacto-oligosaccharides, glucose, glycogen, inulin, lactose maltodextrin, maltose, maltotriose, mannitol, mannose, melibiose, N-acetyl-D-galactosamine, N-acetyl-D-glucosamine, pullulan, raffinose, rhamnose, ribose, sorbitol, starch, sucrose, trehalose, turanose and xylose, were performed using 96-well microtiter plate. Bacterial cells were grown in each well containing a different

sugar, and incubated in an anaerobic cabinet. Growth was evaluated by monitoring the OD₆₀₀ of cultures in the microtiter plates on a plate reader every 12 h. Cultures were grown in biologically independent triplicates, and the resulting growth data were expressed as a means of these replicates (Table 2). The novel isolate, *B. tissieri* DSM 100201^T and *B. stellenboschense* DSM 23968^T were also investigated for enzymatic activities, using API 50 CHL and Rapid ID 32 test kits (BioMérieux) (Table 2). The cell-wall peptidoglycan composition was performed following the protocols previously described by Schumann (27). The total hydrolysate of the peptidoglycan showed the presence of the amino acid lysine, alanine, glutamic acid and muramic acid. The peptidoglycan type of Tam10B^T was L-Glu-L-Ala-L-Lys, with the presence of Ala, Glu and Lys (ratio 0.9 : 1.0 : 0.6), while for the *B. tissieri* DSM 100201^T and *B. stellenboschense* DSM 23968^T the ratio of Ala, Glu and Lys were 1.1 : 1.0 : 0.7 and 1.1 : 1.0 : 1.6., respectively (Table 3). Cellular fatty acids methyl esters were determined accordingly to the protocol described by Miller (28) with the same modifications (29) using Agilent model 6890N gas chromatograph. The results were analyzed using MIS Standard Software (Microbial ID) that determined automatically the peaks and the percentage of fatty acids detected. These data showed that palmitic, myristic and oleic acid were the dominant fatty acid (Table 4). The fructose 6-phosphate phosphoketolase (F6PPK) is considered a key enzyme of phosphoketolase pathway, a taxonomic marker for identification of *Bifidobacterium* and related genera (30). The fructose 6-phosphate phosphoketolase can be identified in genome

sequences through the use of Pathway Tools software (31). Notably, *in silico* analysis of Tam10B^T genome showed that this strain possesses the complete metabolic pathway for F6PPK activity.

Description of *Bifidobacterium vansinderenii* sp. nov.

Bifidobacterium vansinderenii (van.sin.de.ren'i.i. N.L. gen. n. *vansinderenii* of van Sinderen; named after Professor Douwe van Sinderen, in recognition of his extensive contributions to our current knowledge about bifidobacterial biology). Cells are Gram-positive, non-motile, non-sporulating and F6PPK-positive. Cells grow in anaerobic and aerobic conditions. The diameters of each colony, when grown on MRS agar, ranges from 1.0 to 2.5 mm. Furthermore, the colonies grown on the surface of MRS agar plate are white and circular, while the embedded colonies are white and elliptical. Moreover, the cells of *B. vansinderenii* sp. nov., as well as *B. tissieri* DSM 100201^T and *B. stellenboschense* DSM 23968^T are able to grow from 25 to 42°C under anaerobic condition. Whereas, under aerobic conditions these bacterial taxa were able to grow between 35-42°C. In contrast, *B. stellenboschense* DSM 23968^T was not able to grow at 42°C (Table 2). In all cases the cells do not grow at 20°C and 45°C (Table 2). Moreover, the cells grown at the range of pH from 3.5 to 6, with an optimum growth at pH 6 after incubation of 48h as for the *B. tissieri* DSM100201^T, while *B. stellenboschense* DSM 23968^T the range of pH is from 4.0 to 6.0 (Table 2). Fermentation profiles of *B. vansinderenii* LMG 30126^T revealed that it is able to ferment a wide range of simple and complex carbohydrates, including cellobiose, fructose, galactose, glucose, maltodextrin, mannitol,

mannose, melibiose, N-acetyl-D-glucosamine, raffinose, rhamnose, ribose, sorbitol, sucrose, trehalose and turanose (i.e., reaching a final OD₆₀₀ of ≥ 0.51). In contrast, *B. vansinderenii* LMG 30126 displays little if any growth on arabinose, fucose, inulin, N-acetyl-D-galactosamine, pullulan and xylose (i.e., final OD₆₀₀ is ≤ 0.5) (Table 2). Noticeable differences in growth on cellobiose, glycogen, maltotriose, mannose, N-acetyl-D-galactosamine, rhamnose, sorbitol and xylose were evident among the *B. vansinderenii* sp. nov., *B. tissieri* DSM 100201^T and *B. stellenboschense* DSM 23968^T. Positive enzymatic activity is observed for arginine, α -glucosidase, N-acetyl- β -glucosaminidase, alkaline phosphatase, arginine arylamidase and leucyl-glycine arylamidase. Whereas, a weak reaction occurs for α -galactosidase, β -galactosidase, α -arabinosidase and glutamyl glutamic acid arylamidase (Table 2). Moreover, β -glucosidase, β -glucuronidase, glutamic acid decarboxylase, pyroglutamic acid

arylamidase and serine arylamidase activity were not detected (Table 2). The major fatty acids identified are palmitic, myristic and oleic acid. The peptidoglycan type is L-Glu-L-Ala-L-Lys. The type strain Tam10B^T (=LMG 30126^T=CCUG 70655^T) was isolated from the stool sample of Emperor tamarin (*Saguinus imperator*). The DNA G+C content is 62.5 %.

Acknowledgments

This work was funded by the EU Joint Programming Initiative – A Healthy Diet for a Healthy Life (JPI HDHL, <http://www.healthydietforhealthylife.eu/>) and the MIUR to MV. We thank GenProbio srl for financial support of the Laboratory of Probiogenomics. Part of this research is conducted using the High Performance Computing (HPC) facility of the University of Parma. The authors declare that they have no competing interests.

References

1. **Milani C, Lugli GA, Duranti S, Turrone F, Bottacini F, Mangifesta M, Sanchez B, Viappiani A, Mancabelli L, Taminiu B, Delcenserie V, Barrangou R, Margolles A, van Sinderen D, Ventura M.** 2014. Genomic encyclopedia of type strains of the genus *Bifidobacterium*. *Appl Environ Microbiol* **80**:6290-6302.
2. **Zhang G, Gao B, Adeolu M, Khadka B, Gupta RS.** 2016. Phylogenomic Analyses and Comparative Studies on Genomes of the Bifidobacteriales: Identification of Molecular Signatures Specific for the Order Bifidobacteriales and Its Different Subclades. *Front Microbiol* **7**:978.
3. **Milani C, Lugli GA, Duranti S, Turrone F, Bottacini F, Mangifesta M, Sanchez B, Viappiani A, Mancabelli L, Taminiu B, Delcenserie V, Barrangou R, Margolles A, van Sinderen D, Ventura M.** 2014. Genomic Encyclopedia of Type Strains of the Genus *Bifidobacterium*. *Appl Environ Microbiol* **80**:6290-6302.
4. **Michelini S, Oki K, Yanokura E, Shimakawa Y, Modesto M, Mattarelli P, Biavati B, Watanabe K.** 2016. *Bifidobacterium myosotis* sp. nov., *Bifidobacterium tissieri* sp. nov. and *Bifidobacterium hapali* sp. nov., isolated from faeces of baby common marmosets (*Callithrix jacchus* L.). *Int J Syst Evol Microbiol* **66**:255-265.
5. **Michelini S, Modesto M, Filippini G, Spiezio C, Sandri C, Biavati B, Pisi A, Mattarelli P.** 2016. *Bifidobacterium aerophilum* sp. nov., *Bifidobacterium avesanii* sp. nov. and *Bifidobacterium ramosum* sp. nov.: Three novel taxa from the faeces of cotton-top tamarin (*Saguinus oedipus* L.). *Syst Appl Microbiol* **39**:229-236.
6. **Praet J, Meeus I, Cnockaert M, Aerts M, Smagghe G, Vandamme P.** 2015. *Bifidobacterium commune* sp. nov. isolated from the bumble bee gut. *Antonie Van Leeuwenhoek* **107**:1307-1313.
7. **Laureys D, Cnockaert M, De Vuyst L, Vandamme P.** 2016. *Bifidobacterium aquikefiri* sp. nov., isolated from water kefir. *Int J Syst Evol Microbiol* doi:10.1099/ijsem.0.000877.
8. **Ventura M, van Sinderen D, Fitzgerald GF, Zink R.** 2004. Insights into the taxonomy, genetics and physiology of bifidobacteria. *Antonie Van Leeuwenhoek* **86**:205-223.
9. **Turrone F, van Sinderen D, Ventura M.** 2011. Genomics and ecological overview of the genus *Bifidobacterium*. *Int J Food Microbiol* **149**:37-44.
10. **Ventura M, Turrone F, Lugli GA, van Sinderen D.** 2014. Bifidobacteria and humans: our special friends, from ecological to genomics perspectives. *J Sci Food Agric* **94**:163-168.
11. **Delcenserie V, Gavini F, Beerens H, Tresse O, Franssen C, Daube G.** 2007. Description of a new species, *Bifidobacterium crudilactis* sp. nov., isolated from raw milk and raw milk cheeses. *Syst Appl Microbiol* **30**:381-389.
12. **Ventura M, Elli M, Reniero R, Zink R.** 2001. Molecular microbial analysis of *Bifidobacterium* isolates from different environments by the species-specific amplified ribosomal DNA restriction analysis (ARDRA). *Fems Microbiol Ecol* **36**:113-121.
13. **Ventura M, Zink R.** 2002. Rapid identification, differentiation, and proposed new taxonomic classification of *Bifidobacterium lactis*. *Appl Environ Microbiol* **68**:6429-6434.
14. **Milani C, Andrea Lugli G, Duranti S, Turrone F, Mancabelli L, Ferrario C, Mangifesta M, Hevia A, Viappiani A, Scholz M, Arioli S, Sanchez B, Lane J, Ward DV, Hickey R, Mora D, Segata N, Margolles A, van Sinderen D, Ventura M.** 2015. Bifidobacteria exhibit social behavior through carbohydrate resource sharing in the gut. *Sci Rep* **5**:15782.
15. **Duranti S, Milani C, Lugli GA, Mancabelli L, Turrone F, Ferrario C, Mangifesta M, Viappiani A, Sanchez B, Margolles A, van Sinderen D, Ventura M.** 2016. Evaluation of genetic diversity among strains of the human gut commensal *Bifidobacterium adolescentis*. *Sci Rep* **6**:23971.
16. **Lugli GA, Milani C, Mancabelli L, van Sinderen D, Ventura M.** 2016. MEGAnnotator: a user-friendly pipeline for microbial genomes assembly and annotation. *Fems Microbiol Lett* **363**.

17. **Ventura M, Canchaya C, Tauch A, Chandra G, Fitzgerald GF, Chater KF, van Sinderen D.** 2007. Genomics of Actinobacteria: tracing the evolutionary history of an ancient phylum. *Microbiol Mol Biol Rev* **71**:495-548.
18. **Ventura M, Canchaya C, Del Casale A, Dellaglio F, Neviani E, Fitzgerald GF, van Sinderen D.** 2006. Analysis of bifidobacterial evolution using a multilocus approach. *Int J Syst Evol Microbiol* **56**:2783-2792.
19. **Lugli GA, Milani C, Turrone F, Duranti S, Ferrario C, Viappiani A, Mancabelli L, Mangifesta M, Taminiau B, Delcenserie V, van Sinderen D, Ventura M.** 2014. Investigation of the evolutionary development of the genus *Bifidobacterium* by comparative genomics. *Appl Environ Microbiol* **80**:6383-6394.
20. **Zhao Y, Wu J, Yang J, Sun S, Xiao J, Yu J.** 2012. PGAP: pan-genomes analysis pipeline. *Bioinformatics* **28**:416-418.
21. **Enright AJ, Van Dongen S, Ouzounis CA.** 2002. An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res* **30**:1575-1584.
22. **Glaeser SP, Kampfer P.** 2015. Multilocus sequence analysis (MLSA) in prokaryotic taxonomy. *Syst Appl Microbiol* **38**:237-245.
23. **Gevers D, Cohan FM, Lawrence JG, Spratt BG, Coenye T, Feil EJ, Stackebrandt E, Van de Peer Y, Vandamme P, Thompson FL, Swings J.** 2005. Opinion: Re-evaluating prokaryotic species. *Nat Rev Microbiol* **3**:733-739.
24. **Stackebrandt E, Goebel BM.** 1994. Taxonomic Note: A Place for DNA-DNA Reassociation and 16S rRNA Sequence Analysis in the Present Species Definition in Bacteriology. *Int J Syst Evol Microbiol* **44**:846-849.
25. **Konstantinidis KT, Tiedje JM.** 2005. Genomic insights that advance the species definition for prokaryotes. *Proc Natl Acad Sci U S A* **102**:2567-2572.
26. **Richter M, Rossello-Mora R.** 2009. Shifting the genomic gold standard for the prokaryotic species definition. *Proc Natl Acad Sci U S A* **106**:19126-19131.
27. **Schumann P.** 2011. Peptidoglycan Structure, *Methods in Microbiology* **38**:101-129 doi:10.1016/B978-0-12-387730-7.00005-X.
28. **Miller LT.** 1982. Single derivatization method for routine analysis of bacterial whole-cell fatty acid methyl esters, including hydroxy acids. *J Clin Microbiol* **16**:584-586.
29. **Kuykendall LD, Roy MA, Oneill JJ, Devine TE.** 1988. Fatty-Acids, Antibiotic-Resistance, and Deoxyribonucleic-Acid Homology Groups of *Bradyrhizobium-Japonicum*. *Int J Syst Bact* **38**:358-361.
30. **Biavati B, Mattarelli P.** 2012. Genus *Bifidobacterium*, p 171-206, *Bergey's Manual of Systematic Bacteriology*.
31. **Karp PD, Paley S, Romero P.** 2002. The Pathway Tools software. *Bioinformatics* **18 Suppl 1**:S225-232.

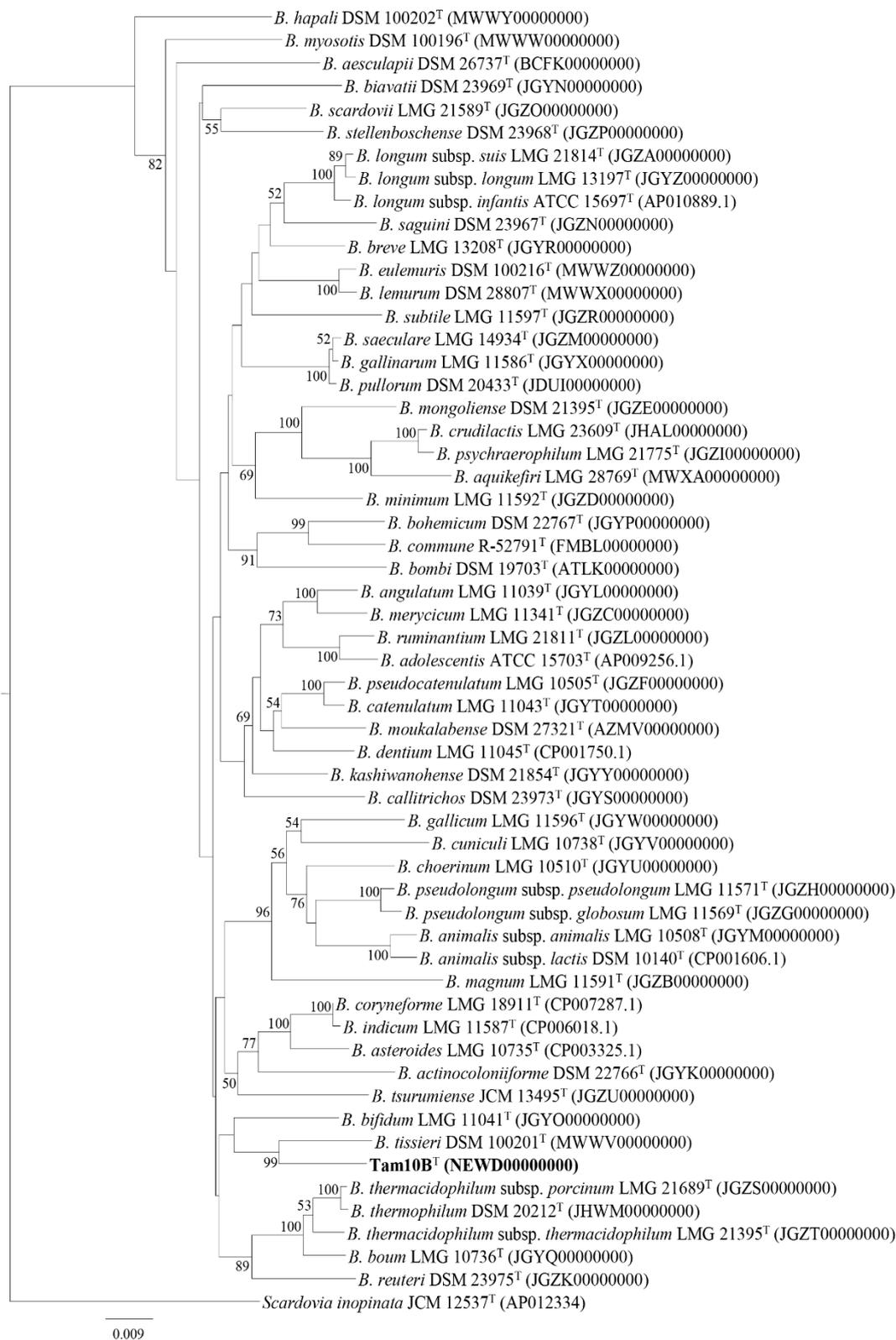


Figure 1. Phylogenetic relationship between strain Tam10B^T and members of the genus *Bifidobacterium* as based on comparative analysis of their respective 16S rRNA gene sequences. The 16S rRNA-based tree was constructed by the maximum likelihood method and the 16S rRNA gene sequence of *Scardovia inopinata* JCM 12537 was used as outgroup. Bootstrap percentages above 50 are showed at node points, based on 1000 replicates of the phylogenetic tree.

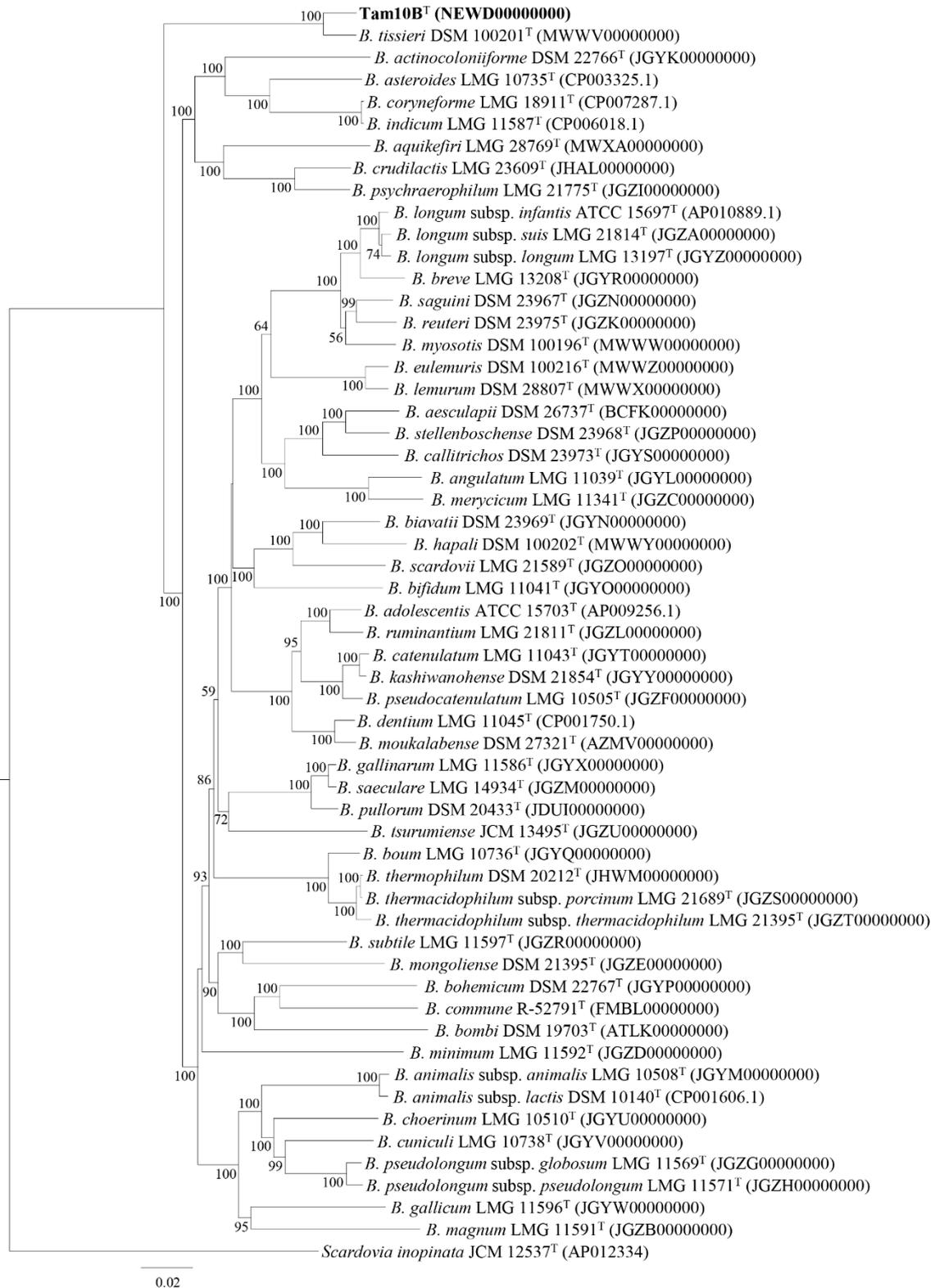


Figure 2. Phylogenetic tree based on the concatenation of proteins sequences deduced from the housekeeping genes *hsp60*, *rpoB*, *dnaJ*, *dnaG* and *clpC*, showing the relationship between strain Tam10B^T and members of the *Bifidobacterium* genus. The housekeeping-based tree was constructed by the maximum likelihood method and corresponding sequences of *Scardovia inopinata* JCM 12537 were used as outgroup. Bootstrap percentages above 50 are shown at node points, based on 1000 replicates of the phylogenetic tree.

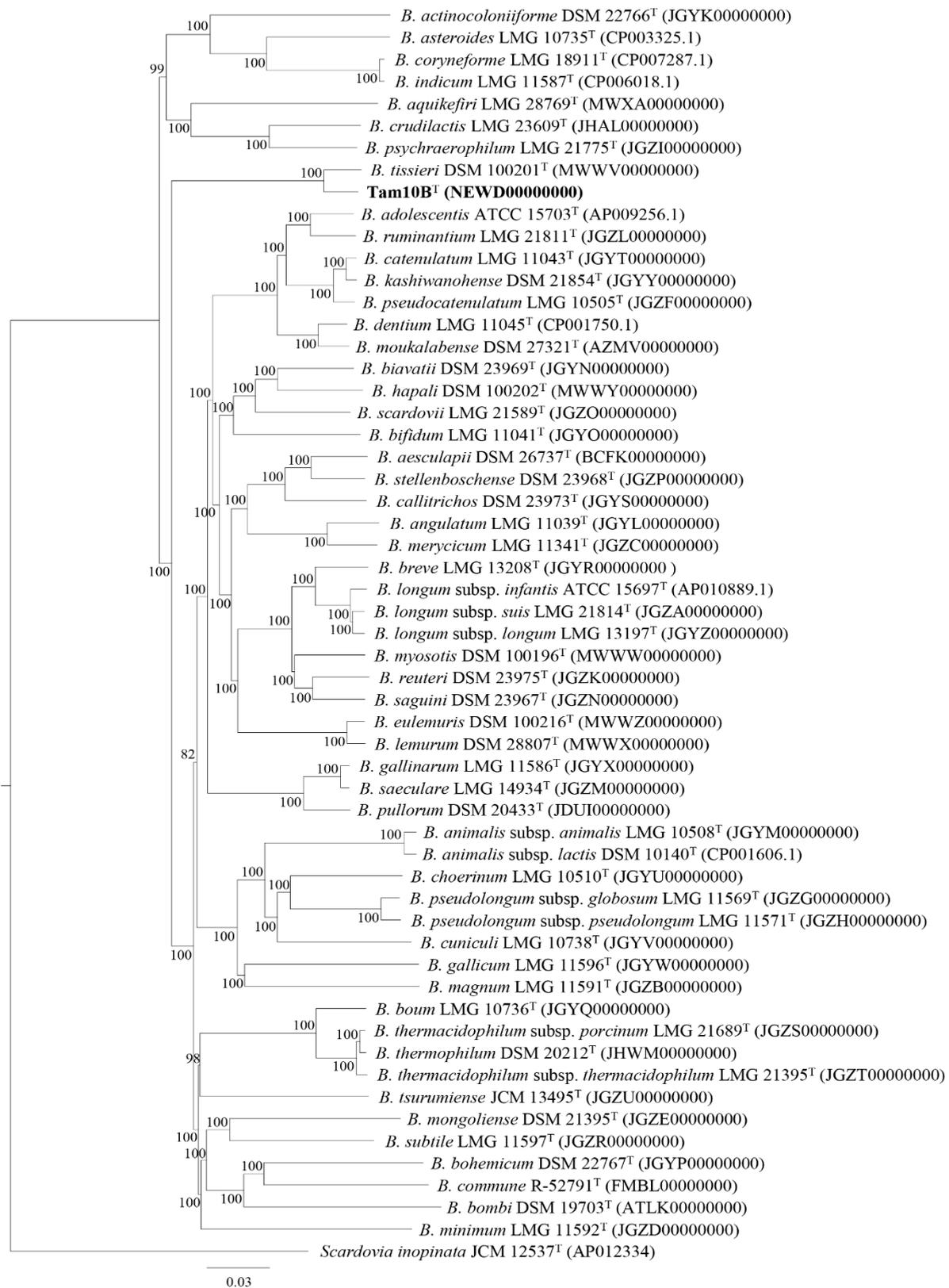


Figure 3. Phylogenetic tree of the *Bifidobacterium* genus based on the concatenation of 237 protein sequences that represent the *Bifidobacterium* core genome. The phylogenetic tree was constructed by the maximum likelihood method and the corresponding sequences of *Scardovia inopinata* JCM 12537 were used as outgroup. Bootstrap percentages above 50 are shown at node points, based on 1000 replicates of the phylogenetic tree.

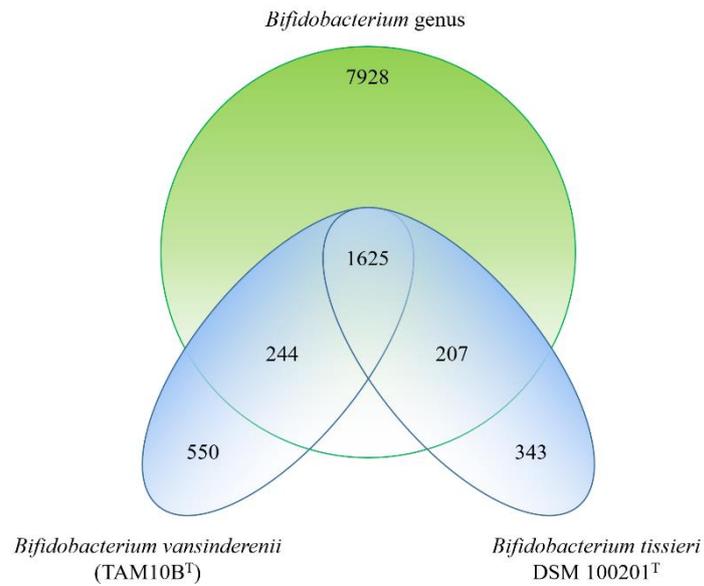


Figure 4. Genomic diversity between Tam10B^T and *B. tissieri* DSM 100201^T. The Venn diagram shows the homologues shared between Tam10B^T and *B. tissieri* DSM 100201^T as well as between these two strains and the pangenome of the *Bifidobacterium* genus.

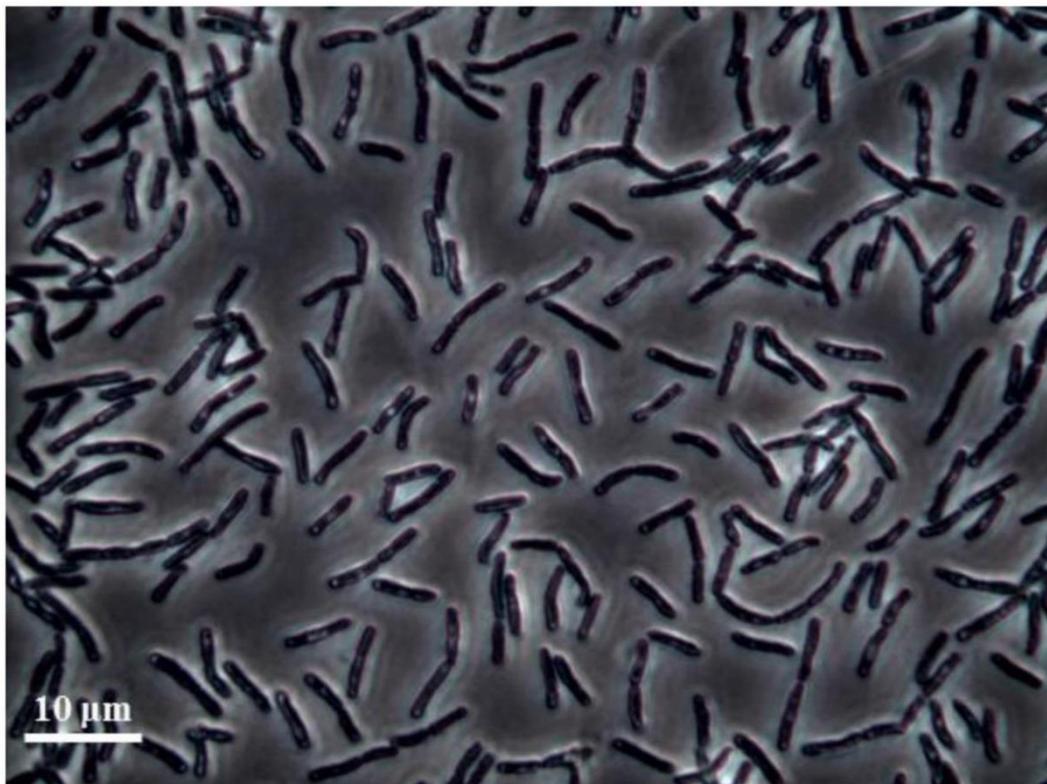


Figure 5. Tam10B^T cellular morphology by using phase-contrast microscopy. Bar, 10 μm.

Table 1. Summary of general genome features of Tam10B^T.

	Tam10B^T
Biological origin	Feces of tamarin
Average coverage	138
Assembled contigs	68
Genome length	3,111,005
Average GC percentage	62.5
Predicted ORFs	2,522
tRNA	60
rRNA*	4

*Predicted rRNA loci

Table 2. Phenotypic information for strains representing the putative novel species.

Fermentation	Tam10B ^T		<i>B. tissieri</i> DSM200201 ^T		<i>B. stellenboschense</i> DSM23968 ^T	
	24h	48h	24h	48h	24h	48h
Arabinose	+	+	+	+	+	+
Arabinogalactan	+	+	+	+	+	+
Cellobiose	++	++	+++	+++	+	+
Fructo-oligosaccharides	+++	+++	+++	+++	+++	+++
Fructose	+++	+++	+++	+++	+++	+++
Fucose	+	+	+	+	+	+
Galactose	+	+++	+++	+++	++	+++
Galacto-oligosaccharides	+++	+++	+++	+++	+++	+++
Glucose	+++	+++	+++	+++	+++	+++
Glycogen	+	+	+	++	+++	+++
Inulin	+	+	+	+	+	+
Lactose	+++	+++	+++	+++	+++	+++
Maltodextrin	+	++	+	++	++	+++
Maltose	+++	+++	+++	+++	+++	+++
Maltotriose	++	++	+	++	+++	+++
Mannitol	+	+++	+++	+++	+++	+++
Mannose	+	++	+	+	+	+
Melibiose	+	+++	+++	+++	+++	+++
N-Acetyl-D-galactosamine	++	+	+	+	-	-
N-Acetyl-D-glucosamine	++	++	+++	+++	-	+
Pullulan	+	+	+	+	+	+
Raffionose	+++	+++	+++	+++	+++	+++
Rhamnose	++	++	+	++	+	+
Ribose	+	+++	+++	+++	+++	+++
Sorbitol	++	++	+	+	+++	+++
Starch	++	+	+	+	+	+
Sucrose	+++	+++	+++	+++	+++	+++
Trehalose	++	+++	+++	+++	+++	+++
Turanose	+	+++	+++	+++	+++	+++
Xylose	+	+	+++	+++	++	+++
Enzymatic activity						
Arginine dihydrolase		+		+		-
α - Galactosidase		w		-		w
β - Galactosidase		w		-		w
α - Glucosidase		+		-		w
β - Glucosidase		-		-		-
α - Arabinosidase		w		-		+
β - Glucuronidase		-		-		+
N-acetyl-β -Glucosaminidase		+		-		-

Glutamic acid decarboxylase	-	-	-	-	+	+
Alkaline phosphatase	+	+	+	+	+	+
Arginine arylamidase	+	+	+	+	+	+
Leucyl-glycine arylamidase	+	+	+	+	+	+
Pyroglutamic acid arylamidase	-	-	-	-	+	+
Glutamyl glutamic acid arylamidase	w	-	-	-	+	+
Serine arylamidase	-	-	-	-	+	+
Temperature	Aerobic	Anaerobic	Aerobic	Anaerobic	Aerobic	Anaerobic
20°C	-	-	-	-	-	-
25°C	-	++	-	++	-	++
30°C	-	+++	-	++	-	++
35°C	++++	++++	++++	++++	++++	+++
37°C	++++	++++	++++	++++	++++	++++
40°C	++++	++++	++	++	+++	++
42°C	++++	++++	++++	++++	-	-
45°C	-	-	-	-	-	-
pH values	24h	48h	24h	48h	24h	48h
3.5	++	++	+	+	-	-
4	++	++	+	+	+	+
4.5	+	+	++	++	++	++
5	+	+++	++	+++	++	+++
5.5	+	+++	++	+++	++	+++
6.0	+	++++	++	+++	++	+++

Symbols: -, no growth; + poor growth as measured by $0.5 \leq OD_{600}$; ++ growth as measured by $0.51 \leq OD_{600} \leq 0.8$; +++ good growth as measured by $0.81 \leq OD_{600} \leq 1.9$; ++++ very good growth as measured by $OD_{600} \geq 1.91$.

Enzymatic activity: +, Positive; -, Negative; w, Weakly Positive

Table 3. Cell-wall amino acids composition of *B. vansinderenii* sp. nov and related type strains

	<i>B. vansinderenii</i> Tam10B ^T	<i>B. tissieri</i> DSM 100201 ^T	<i>B. stellenboschense</i> DSM 23968 ^T
Glutamic Acid	1.0	1.0	1.0
Serine	-	-	-
Alanine	0.9	1.1	1.1
Lysine	0.6	0.7	1.6

Table 4. Composition of cellular fatty acid of *B. vansinderenii* sp. nov. and related type strains.

Fatty acid	<i>B. vansinderenii</i> Tam10B ^T	<i>B. tissieri</i> DSM 100201 ^T	<i>B. stellenboschense</i> DSM 23968 ^T
C _{10:0}	-	-	-
C _{12:0}	1.7	1.6	1.8
C _{11:0} DMA	-	-	-
C _{14:0}	15.9	15.3	12.6
C _{14:0} DMA	7.2	6.9	7.4
C _{16:1} ^{ω^{9c}}	2.7	2.7	1.8
C _{16:0}	30.2	30.5	23.6
C _{16:0} DMA	-	-	1.4
C _{18:1} ^{ω^{9c}}	15.9	16.8	10.0
C _{16:1} ^{ω^{6c}}	-	-	-
C _{18:0}	1.2	1.8	1.7
C _{18:1} ^{ω^{9c}} DMA	11.7	11.6	19.8
C _{19:0} cyclo 9,10	2.3	2.6	1.6
C _{19:0} cyclo 9,10 DMA	2.4	2.1	1.9
C _{16:1} ^{2-OH}	-	-	1.0

DMA abbreviation indicates Dimethylacetal; - No signal; Values represents the percentages of total fatty acids.

Values represents molar ratio to glutamic acid.

Chapter 5

Phylogenetic classification of six novel species belonging to the genus *Bifidobacterium* comprising *Bifidobacterium anseris* sp. nov., *Bifidobacterium criceti* sp. nov., *Bifidobacterium imperatoris* sp. nov., *Bifidobacterium italicum* sp. nov., *Bifidobacterium margollesii* sp. nov. and *Bifidobacterium parmae* sp. nov.

Lugli GA[‡], Mangifesta M[‡], Duranti S[‡], Anzalone R, Milani C, Mancabelli L, Alessandri G, Turroni F, Ossiprandi MC, van Sinderen D, Ventura M. .

This chapter was published in:

Syst Appl Microbiol. 2018 May;41(3):173-183. Epub 2018 Jan 31.

Reprinted with permission from Elsevier.

[‡]these authors contribute equally

Phylogenetic classification of six novel species belonging to the genus
Bifidobacterium* comprising *Bifidobacterium anseris* sp. nov., *Bifidobacterium criceti
sp. nov., *Bifidobacterium imperatoris* sp. nov., *Bifidobacterium italicum* sp. nov.,
***Bifidobacterium margollesii* sp. nov. and *Bifidobacterium parmae* sp. nov.**

Running title: Identification of novel bifidobacterial species

Key words: phylogenetic, next generation sequencing, genomics, metagenomics, bifidobacteria,

Bifidobacterium

Gabriele Andrea Lugli*¹, Marta Mangifesta*², Sabrina Duranti*¹, Rosaria Anzalone¹, Christian Milani¹, Leonardo Mancabelli¹, Giulia Alessandri¹, Francesca Turrone¹, Maria Cristina Ossiprandi³,
Douwe van Sinderen⁴ and Marco Ventura¹

Laboratory of Probiogenomics, Department of Chemistry, Life Sciences and Environmental Sustainability, University of Parma, Parma, Italy¹; GenProbio srl, Parma, Italy²; Department of Veterinary Medical Science, University of Parma, Parma, Italy³; APC Microbiome Institute and School of Microbiology, National University of Ireland, Cork, Ireland⁴.

*These authors contributed equally

Corresponding author. Mailing address for Marco Ventura laboratory of Probiogenomics, Department of Chemistry, Life Sciences and Environmental Sustainability, University of Parma, Parco Area delle Scienze 11a, 43124 Parma, Italy. Phone: ++39-521-905666. Fax: ++39-521-905604. E-mail: marco.ventura@unipr.it

Abstract

Six *Bifidobacterium* strains, i.e., Goo31D, Ham19E, Rab10A, Tam1G, Uis4E and Uis1B, were isolated from domestic goose (*Anser domesticus*), European hamster (*Cricetus cricetus*), European rabbit (*Oryctolagus cuniculus*), emperor tamarin (*Saguinus imperator*) and pygmy marmoset (*Callithrix pygmaea*), respectively. Cells are Gram-positive, non-motile, non-sporulating, facultative anaerobic and fructose 6-phosphate phosphoketolase-positive. Phylogenetic analyses based on 16S rRNA and ITS sequences, multilocus sequences (including *hsp60*, *rpoB*, *dnaJ*, *dnaG* and *clpC* genes) and the core genome revealed that bifidobacterial strains Goo31D, Ham19E, Rab10A, Tam1G, Uis4E and Uis1B exhibit close phylogenetic relatedness to *Bifidobacterium choerinum* LMG 10510, *Bifidobacterium hapali* DSM 100202, *Bifidobacterium saguini* DSM 23967 and *Bifidobacterium stellenboschense* DSM 23968. Further genotyping based on the genome sequence of the isolated strains combined with phenotypic analyses, clearly show that these strains are distinct from each of the type strains of the so far recognized *Bifidobacterium* species. Thus, *Bifidobacterium anseris* sp. nov. (Goo31D = LMG 30189^T = CCUG 70960^T), *Bifidobacterium criceti* sp. nov. (Ham19E = LMG 30188^T = CCUG 70962^T), *Bifidobacterium imperatoris* sp. nov. (Tam1G = LMG 30297^T = CCUG 70961^T), *Bifidobacterium italicum* sp. nov. (Rab10A = LMG 30187^T = CCUG 70963^T), *Bifidobacterium margollesii* sp. nov. (Uis1B = LMG 30296^T = CCUG 70959^T) and *Bifidobacterium parmae* sp. nov. (Uis4E = LMG 30295^T = CCUG

70964^T) are proposed as novel *Bifidobacterium* species.

Introduction

Bifidobacteria are one of the most noteworthy microbial groups belonging to the *Actinobacteria* phylum, and are common inhabitants of the gastrointestinal (GIT) of mammals, birds and social insects [1-3]. Notably, this group of microorganisms is believed to confer a range of health-promoting properties to its host, such as modulation of immune response, modulation of mucosal physiology of the host and inhibition of pathogen proliferation [4-6]. Furthermore, bifidobacteria are known to be abundantly present in those animals that provide parental care to their offspring and several species belonging to this genus are reported to be among the first gut colonizers of newborns [7-10]. Currently 59 taxa, representing 50 species and nine subspecies have been formally recognized as members of the genus *Bifidobacterium* [11-18]. A recent metagenomics-based study, based on Internally Transcribed Spacer (ITS) rRNA profiling and aimed at exploring the gut microbiota biodiversity across the mammalian branch of the tree of life, identified a number of putative novel taxa belonging to the genus *Bifidobacterium* [19]. Interestingly, these data revealed the presence of 89 putative novel bifidobacterial taxa in addition to the previously described (sub)species [19]. The 291 analysed animal hosts included the domestic goose (*Anser domesticus*), a winged animal that belongs to the *Anatidae* family, the European hamster (*Cricetus cricetus*), a rodent that belongs to the *Cricetidae* family, the European rabbit (*Oryctolagus cuniculus*),

a lagomorph belonging to the *Leporidae* family, and the emperor tamarin (*Saguinus imperator*) and pygmy marmoset (*Callithrix pygmaea*), two primates harboring the *Callitrichidae* family that live in the Amazonian lowland. In the current study, we describe the identification of novel bifidobacterial species based on 16S rRNA and ITS profiling, followed by genomic comparison as based on whole genome sequencing. Genomic investigation as well as phylogenetic and phenotypic analyses allowed the identification of six proposed novel bifidobacterial species isolated from animal feces, being related to *Bifidobacterium choerinum* LMG 10510, *Bifidobacterium hapali* DSM 100202, *Bifidobacterium tissieri* DSM100201, *Bifidobacterium saguini* DSM 23967 and *Bifidobacterium stellenboschense* DSM 23968.

Materials and Methods

Bifidobacterial selection. Fecal samples of several animals were collected by biologist supervisors of Italian zoo's and Italian Natural Parks in collaboration with the Laboratory of Probiogenomics, University of Parma, Italy. One gram of fecal sample of a domestic goose (*Anser domesticus*), European hamster (*Cricetus cricetus*), European rabbit (*Oryctolagus cuniculus*), Emperor tamarin (*Saguinus imperator*) and Pygmy marmoset (*Callithrix pygmaea*) was mixed with nine ml of phosphate-buffered saline (PBS), pH 6.5. Serial dilution and subsequent plating were performed using de Man-Rogosa-Sharpe (MRS) agar, supplemented with 50 µg/ml mupirocin (Delchimica, Italy) and 0.05 % (wt/col) L-cysteine hydrochloride. Agar plates were incubated for 48 h at 37°C in a

chamber (Concept 400; Ruskin) with anaerobic atmosphere (2.99 % H₂, 17.01 % CO₂ and 80 % N₂). Morphologically different colonies that developed on MRS plates were randomly picked and re-streaked in order to isolate purified bacterial strains. All isolates were subjected to DNA isolation and characterized as previously described by Turrone *et al.* [20].

ERIC-PCR genotyping. The isolates were subjected to molecular typing through the use of Enterobacterial Repetitive Intergenic Consensus sequences (ERIC) PCR. ERIC-PCR was carried out using ERIC-1 (5'-ATGTAAGCTCCTGGGGATTAC-3') and ERIC-2 (5'-AAGTAAGTGACTGGGGTGAGCG-3') primers following a previously described procedure [21].

Amplification of 16S rRNA gene and associated ITS sequences. Partial 16S rRNA gene sequences were amplified from extracted DNA using primer pair Probio_Uni/Probio_Rev, which targets the V3 region of the 16S rRNA gene sequence [22]. Partial ITS sequences of the six novel bifidobacterial taxa were amplified using primer pair Probio-bif_Uni/Probio-bif_Rev [23].

Genome sequencing and assemblies. DNA extracted from the bifidobacterial isolates was subjected to whole genome sequencing using MiSeq (Illumina, UK) at GenProbio srl (Parma, Italy) following the supplier's protocol (Illumina, UK). Fastq files of the paired-end reads obtained from targeted genome sequencing of the isolated strains were used as input for the genome assemblies through the MEGAnnotator pipeline [24]. MIRA (version 4.0.2) software was used for *de novo* assembly of each bifidobacterial genome sequence

[25], while protein-encoding open reading frames (ORFs) were predicted using Prodigal [26].

Bifidobacterial core genome evaluation. For the 59 genomes of bifidobacterial type strains and the six novel sequenced genomes, a pan-genome calculation was performed using the PGAP pipeline [27]. The ORF content of each bifidobacterial genome was organized in functional gene clusters using the GF (Gene Family) method involving comparison of each protein to all other proteins using BLAST analysis (cut-off E-value of 1×10^{-5} and 50 % identity over at least 80 % of both protein sequences). Sequences were then clustered in protein families named clusters of orthologous genes (COGs), using MCL (graph-theory-based Markov clustering algorithm) [28] and protein families shared between all genomes, named core COGs, were defined by selecting the families that contained at least one single protein member for each genome.

Phylogenetic and phylogenomic comparisons. The collected bifidobacterial 16S rRNA gene sequences were aligned using MAFFT [29], as well as the ITS, housekeeping genes and core gene sequences. Bifidobacterial phylogenetic trees were constructed using the neighbor-joining method in Clustal W, version 2.1 [30] and were built using FigTree (<http://tree.bio.ed.ac.uk/software/figtree/>). For each genome pair, a value of average nucleotide identity (ANI) was calculated using the program JSpecies, version 1.2.1 [31]. The identity percentage between 16S rRNA gene or ITS sequences was calculated with MatGat, version 2.03 (Matrix Global Alignment Tool), using the BLOSUM 50 alignment matrix [32]. The Genome-to-Genome Distance Calculator (GGDC) version 2.1 was employed to estimate the

DNA-DNA hybridization (DDH) between bifidobacterial taxa, using the recommended “Formula 2” (identities / high-scoring segment pairs length) [33].

Phenotypic characterization. The morphology of six novel bifidobacterial taxa was determined using phase-contrast microscopy after incubation of each strain under anaerobic conditions at 37°C for 24h. Growth was assessed, using MRS broth, under aerobic or anaerobic conditions at different temperatures (20, 25, 30, 35, 37, 40 and 45 °C) and at different pH values (pH 3.5, 4.0, 4.5, 5.0, 5.5 and 6.0) (Table 1). Gram staining was performed with LIVE *BacLight*TM Bacterial Gram Stain Kit (ThermoFisher Scientific) as previously described by Duranti *et al.* [11]. The six identified, novel bifidobacterial taxa were grown on MRS medium without glucose supplemented with 1 % (wt/vol) of a particular sugar. Cultures were cultivated in the wells of a 96-well microtiter plate, with each well containing a different sugar, and incubated in an anaerobic cabinet at 37°C. The optical densities at 600nm (OD₆₀₀) was determined using a plate reader (Biotek, VT, USA) at two different time points, 24 and 48 h. Growth assays were carried out in duplicate and non-inoculated MRS medium was used as a negative control. Carbohydrates tested in this study include arabinose, arabinogalactan, cellobiose, fructo-oligosaccharides, fructose, fucose, galactose, galacto-oligosaccharides, glucose, glycogen, inulin, lactose, maltodextrin, maltose, maltotriose, mannitol, mannose, melibiose, N-acetyl-D-galactosamine, N-acetyl-D-glucosamine, pullulan, raffinose, rhamnose, ribose, sorbitol, starch, sucrose, trehalose, turanose and xylose (Table 2). The six novel isolates

and their closest related species, i.e., *B. choerinum* LMG10510, *B. tissieri* DSM100201, *B. hapali* DSM100202, *B. saguini* DSM23967 and *B. stellenboschense* DSM 23968^T, were also investigated for enzymatic activities, using API 50 CHL and Rapid ID 32 test kits (BioMérieux) (Table 2). The cell-wall peptidoglycan composition was analyzed following the protocols previously described by Schumann [34]. Cellular fatty acids methyl esters were determined according to the protocol described by Miller [35] with slight modifications [36]. The results analyzed using MIS Standard Software (Microbial ID) and showed that palmitic, myristic and oleic acid were the dominant fatty acids (Table S1).

Bifidobacterial genome sequences. Whole Genome Shotgun projects were deposited at DDBJ/ENA/GenBank under the accession numbers MVOG00000000, MVOH00000000, NMWT00000000, NMWU00000000, NMWV00000000 and NMYC00000000. The versions described in this paper are accessible under the following numbers: MVOG01000000, MVOH01000000, NMWT01000000, NMWU01000000, NMWV01000000 and NMYC01000000.

Results and Discussion

Novel bifidobacterial strain selection and *in vitro* characterization. Fecal samples were screened in order to assess the bifidobacterial population of the gut environment of a variety of animals, including goose, hamster, rabbit and monkeys (Table 3). A combination of culture-based, molecular-based and genomics-based methods was applied in order to

obtain a detailed characterization of each strain. A total of 32 bifidobacterial strains were isolated from stool samples from five different animals, including domestic goose (*Anser domesticus*), European hamster (*Cricetus cricetus*), European rabbit (*Oryctolagus cuniculus*), emperor tamarin (*Saguinus imperator*) and pygmy marmoset (*Callithrix pygmaea*). The genotypic characterization by ERIC-PCR revealed the presence of three known species, *B. choerinum*, *B. saguini* and *B. stellenboschense* by means of comparison to ERIC profiles of the type strains of the respective species (Fig. S1). Furthermore, a representative strain belonging to each putative novel species, was characterized in detail, i.e., strains Goo31D, Ham19E, Rab10A, Tam1G, Uis4E and Uis1B, which correspond to the bifidobacterial taxa previously identified with the names new_taxa_72, new_taxa_71, new_taxa_70, new_taxa_45, new_taxa_77 and new_taxa_61, respectively [19]. These strains represent Gram-positive bacteria, that are all non-motile, non-sporulating and F6PKK-positive. Growth of all isolates and known correlated species was examined on 30 different carbon sources (Table 2). As shown in Table 2, fermentation capabilities for all sugars, except for glucose, appear to be variable among the novel strains. Interestingly, metagenomic analyses performed by Milani *et al* [19]. identified the presence of these new taxa in other animals, i.e. Dolphin (*Delphinus delphis*), European hedgehog (*Erinaceus europaeus*), leopard (*Panthera pardus*) and Norwegian forest cat (*Felis catus*). This suggests that these bifidobacterial taxa can be found in other mammalian species.

Phylogenetic analyses based on 16S rRNA and ITS hypervariable regions.

A first characterization of the isolated bifidobacterial taxa was performed through 16S rRNA gene and ITS sequencing, followed by comparison with the same sequenced molecular markers retrieved from public databases. Notably, the ability to distinguish closely related bifidobacterial taxa based on an ITS sequence dataset of 48 bifidobacterial type strains has previously been described [23]. The 16S rRNA gene sequences of Goo31D, Ham19E, Rab10A, Tam1G, Uis4E and Uis1B showed identity values ranging from 96.2 % to 98.6 % with respect to *B. choerinum* LMG 10510, *Bifidobacterium saguini* DSM 23967, *Bifidobacterium stellenboschense* DSM 23968 and *Bifidobacterium tissieri* DSM 100201 (Table 3). Furthermore, the hypervariable ITS sequences of the six isolated strains displayed identity values ranging from 65.7 % to 89 % against *Bifidobacterium biavatii* DSM 23969 and *B. saguini* DSM 23967 (Table 3). Notably, identity values of hypervariable ITS sequences identified among the nine subspecies of the genus *Bifidobacterium* range from 90.4 % to 98.9 % between *Bifidobacterium animalis* and *Bifidobacterium longum* subspecies. Thus, data retrieved from the ITS analysis highlight a high degree of sequence diversity between the isolates from the classified members of the genus *Bifidobacterium*. Therefore, the hypervariable region of the ITS sequence corresponding to the region between primers Probio-bif_Uni/Probio-bif_Rev [23], was employed to build a phylogenetic tree to better assess the phylogenetic relatedness between strains (Fig. S2). In a similar fashion, a phylogenetic tree based on 16S rRNA gene sequences was

constructed (Fig. S3). These two phylogenetic trees displayed similar results for four of the six isolated strains. In this context, Goo31D and Rab10A share the same phylogenetic position in both trees with *B. choerinum* LMG 10510, while Tam1G and Uis4E are placed on the same phylogenetic branch together with *B. saguini* DSM 23967 and *B. stellenboschense* DSM 23968, respectively (Figs. S2 and S3). Notably, both Ham19E and Uis1B do not cluster with any bifidobacterial strain in the ITS-based phylogenetic tree. Furthermore, based on the 16S rRNA gene-based tree, both of these isolates are related to *B. choerinum* LMG 10510 and the *B. tissieri* DSM 100201-*Bifidobacterium vansinderenii* LMG 30126 pair [11-18], respectively (Fig. S3).

Phylogenetic analysis based on housekeeping genes.

In order to further explore the genetic differences between the six isolated strains and the currently recognized (sub)species of the *Bifidobacterium* genus, nucleotide identity between five genes, i.e. those of *hsp60*, *rpoB*, *dnaJ*, *dnaG* and *clpC*, which are considered to represent molecular clocks in (bifido)bacterial taxonomy, was evaluated [37, 38]. These five selected housekeeping genes showed identity values (at deduced amino acid level) ranging from 82.3 % to 100 % for Uis1B *dnaG* and Tam1G *clpC*, with respect to *B. tissieri* DSM 100201 and *B. saguini* DSM 23967 gene sequences (Table S2). Thus, in order to unveil the phylogenetic relatedness between isolates Goo31D, Ham19E, Rab10A, Tam1G, Uis4E and Uis1B on the one hand, and currently recognized (sub)species of the genus *Bifidobacterium* on the other, a phylogenetic tree was constructed based on the amino acid sequences of the above-mentioned house-keeping genes (Fig. S4).

The phylogenetic tree based on the five housekeeping genes displayed a higher robustness compared to that achieved by 16S rRNA gene sequences, as displayed by the bootstrap values (Figs. S3 and S4). While Goo31D, Ham19E and Rab10A displayed genetic relatedness with *B. choerinum* LMG 10510, Tam1G and Uis4E shared the same branches of *B. saguini* DSM 23967 and *B. stellenboschense* DSM 23968, respectively. Furthermore, among the so far identified bifidobacterial strains, Uis1B displayed the highest level of housekeeping gene diversity when compared with other bifidobacteria, sharing a phylogenetic position closer to that of *B. tissieri* DSM 100201 and *B. vansinderenii* LMG 30126. Thus, an *in silico* analysis based on genome sequences was performed in order to better classify these novel bifidobacterial isolates.

Phylogenomic characterization of the novel bifidobacterial taxa. To get insights into the genetic similarities between Goo31D, Ham19E, Rab10A, Tam1G, Uis4E and Uis1B, and publicly available genome sequences of the genus *Bifidobacterium*, the six isolated genomes were decoded by means of a Next Generation Sequencing (NGS) approach. The sequenced bifidobacterial strains were each shown to contain a single chromosome ranging in length from 2,155,882 to 2,820,211 for Ham19E and Uis4E, respectively, with an average fold coverage ranging from 55.52 to 139.27 (Table 3). The genome sequences of each isolated taxon were used to assess the genetic similarity at genomic level with respect to other currently recognized bifidobacterial (sub)species by Average Nucleotide Identity (ANI) analysis [39]. The assembled genome of Tam1G

exhibits the highest ANI value (94.55 %) against *B. saguini* DSM 23967, while the Uis4E genome sequence is strictly correlated with *B. stellenboschense* DSM 23968 (93.45 %) (Table 3). Furthermore, Uis1B displays the highest ANI value (88.04 %) with respect to *Bifidobacterium hapali* DSM 100202, while Goo31D, Ham19E and Rab10A exhibit ANI values that range from 91.8 % to 87.81 % with respect to *B. choerinum* LMG 10510 (Table 3). In this context, it should be noted that two strains displaying an ANI value <95 % are considered to belong to two distinct species [31]. Notably, these observed ANI values are below the assigned threshold value used for species recognition, supporting the notion that the investigated strains possess a unique genomic composition as compared to currently recognized bifidobacterial taxa (for which a genome sequence is available) [39]. Furthermore, strains Goo31D, Ham19E and Rab10A that exhibit a high level of genomic identity to *B. choerinum* LMG 10510, reveal the highest ANI value of 91.76 % between Ham19E and Rab10A, once again generating an ANI value below 95 %. Besides, Uis1B showed similar ANI values with respect to the phylogenetic correlated taxa *B. tissieri* DSM 100201 (86.81 %) and *B. vansinderenii* LMG 30126 (86.77 %), revealing similar genomic relatedness with three strains isolated from monkeys. Genome sequencing of Goo31D, Ham19E, Rab10A, Tam1G, Uis4E and Uis1B also allowed us to compare the genetic makeup between these strains and other members of the *Bifidobacterium* genus. A comparative study was undertaken to determine orthology between bifidobacterial strains. Notably, such an *in silico* analysis allows the identification of

259 clusters of orthologous genes (COGs) shared by all bifidobacterial taxa, representing the core genome of currently sequenced *Bifidobacterium* representatives. Of note, conserved genes, which represent paralogs within bifidobacterial genomes, were not considered, and a phylogenetic core genome tree was built based on the concatenation of 233 protein sequences (Fig. 1). The phylogenetic tree based on the core genome confirmed the positioning of the six isolated strains within the *Bifidobacterium* genus that were observed in the phylogenetic analyses based on five housekeeping genes (Fig. S4). Moreover, the subdivision in seven bifidobacterial phylogenetic groups allowed the distribution of the identified strains into two major groups, i.e., *Bifidobacterium pseudolongum* group for Goo31D, Ham19E and Rab10A, and *B. longum* group for Tam1G and Uis4E. In contrast, Uis1B does not belong to any of the so far identified bifidobacterial phylogenetic groups, yet shares a phylogenetic position close to other bifidobacterial strains that have been isolated from monkeys [11], [13].

According to phylogenetic analyses based on ITS, 16S rRNA and housekeeping gene sequences, and phylogenomic analyses based on the bifidobacterial core gene sequences, strains Goo31D, Ham19E, Rab10A, Uis4E and Uis1B are shown to be genetically different from any other bifidobacterial type strain characterized to date. Thus, the novel taxon, *Bifidobacterium anseris* sp. nov., *Bifidobacterium criceti* sp. nov., *Bifidobacterium italicum* sp. nov., *Bifidobacterium parmae* sp. nov. and *Bifidobacterium margollesii* sp. nov., are proposed here. Furthermore, a Genome-to-Genome Distance analysis was performed to examine in depth

the genomic relatedness between Tam1G and the phylogenomic correlated strain *B. saguini* DSM 23967 [33]. This analysis estimates a DNA-DNA hybridization (DDH) based on the recommended “Formula 2” (identities / high-scoring segment pairs length) of 55.2 % (confidence interval 52.5 - 57.9 %). The assignment of Tam1G to a novel bifidobacterial species was based on an ANI value below 95 % and an estimated DDH of well below 70 % when compared to its closest relative *B. saguini* DSM 23967, thus corroborating that *Bifidobacterium imperatoris* sp. nov. should be considered a novel bacterial species.

Description of *Bifidobacterium italicum* sp. nov.

Bifidobacterium italicum (i.ta.li.cum, N.L. gen. n. *italicum* pertaining to Italy; country where the type strain was isolated). Cells are Gram-positive, non-motile, non-sporulating and F6PPK-positive and, when grown in MRS broth under anaerobic condition, showed a small rod morphology (Fig. 2). Cells grow in both anaerobic and aerobic conditions. Colony diameter, when grown on MRS agar, ranges from 1.0 to 2.5 mm. Furthermore, colonies grown on the surface of an MRS agar plate are white and circular, while embedded colonies are white and elliptical. Moreover, after 48h the cells incubated under aerobic and anaerobic conditions are able to grow at temperatures that range from 25 to 42°C, but not at 20°C or 45°C (Table 1). Moreover, the cells grow at pH 5-6 (Table 1).

Fermentation profiles of *B. italicum* LMG 30187 revealed that it is able to ferment a wide range of simple and complex carbohydrates, including cellobiose, glucose, maltodextrin, maltose, melibiose and pullulan (i.e., reaching a final OD₆₀₀ of ≥ 0.51).

In contrast, *B. italicum* LMG 30187 displays little if any growth on arabinogalactan, fructo-oligosaccharides, galactose, galacto-oligosaccharides, glycogen, inulin, lactose, maltotriose, mannitol, mannose, N-acetyl-D-glucosamine, N-acetyl-D-galactosamine, raffinose, ribose, sorbitol, starch, sucrose, trehalose and turanose (i.e., final OD₆₀₀ is ≤ 0.5) (Table 2). Furthermore, the strain is unable to grow on arabinose, cellobiose, fructose, fucose, rhamnose and xylose. Positive enzymatic activities are observed for α-arabinosidase, β-glucuronidase, glutamic acid decarboxylase, alkaline phosphatase, arginine arylamidase, leucyl-glycine arylamidase, pyroglutamic acid arylamidase, glutamyl glutamic acid and serine arylamidase. In contrast, only weak reactions were observed for α-galactosidase, β-galactosidase and α-glucosidase activities (Table 2). Moreover, N-acetyl-β-glucosaminidase, arginine dihydrolase and β-glucosidase activities were not detected (Table 2). The peptidoglycan type is L-Lys-L-Ser-L-Ala₂. The major fatty acids identified are palmitic, myristic, oleic acid and stearic acid-dimethylacetal (Table S1). Phylogenetic analysis of *B. italicum* LMG 30187 highlights that this strain is closely related to *B. choerinum* LMG 10510. The type strain Rab10A (=LMG 30187^T=CCUG 70963^T) was isolated from a faecal sample of a European rabbit (*Oryctolagus cuniculus*). The DNA G+C content is 65.45 %. Digital Protologue Taxonumber: TA00239.

Description of *Bifidobacterium criceti* sp. nov.

Bifidobacterium criceti (cri.ce.ti, N.L. gen. n. *criceti* of *cricetus*, common scientific name of European hamster, *Cricetus cricetus*). Cells are Gram-positive,

non-motile, non-sporulating and F6PPK-positive and, when grown in MRS broth under anaerobic conditions, was shown to exhibit a small rod morphology (Fig. 2). Cells grow in anaerobic and aerobic conditions. The colony diameter, when grown on MRS agar, ranges from 1.0 to 2.5 mm. Furthermore, colonies grown on the surface of an MRS agar plate are white and circular, while the embedded colonies are white and elliptical. Moreover, after 48h the cells incubated in aerobic and anaerobic conditions are able to grow from 25 to 42°C, but do not grow at 20°C or 45°C (Table 1). Moreover, the cells grow at pH 5-6 (Table 1).

Fermentation profiles of *B. criceti* LMG 30188 revealed that it is able to ferment a wide range of simple and complex carbohydrates, including cellobiose, galacto-oligosaccharides, glucose, glycogen, maltodextrin, maltose and pullulan (i.e., reaching a final OD₆₀₀ of ≥ 0.51). In contrast, *B. criceti* LMG 30188 displays little if any growth on arabinogalactan, fructo-oligosaccharides, galactose, inulin, lactose, maltotriose, mannitol, mannose, melibiose, N-acetyl-D-glucosamine, N-acetyl-D-galactosamine, raffinose, rhamnose, ribose, sorbitol, starch, sucrose, trehalose, turanose and xylose. (i.e., final OD₆₀₀ is ≤ 0.5) (Table 2). Furthermore, the cells are not able to grow on arabinose, cellobiose, fructose and fucose. Positive enzymatic activities were observed for arginine dihydrolase, alkaline phosphatase, arginine arylamidase, glutamyl glutamic acid and serine arylamidase. Only a weak reaction was detected for leucyl-glycine arylamidase (Table 2). Moreover, α-galactosidase, β-galactosidase, α-glucosidase, β-glucosidase, α-arabinosidase, β-glucuronidase, N-acetyl-β-

glucosaminidase, glutamic acid decarboxylase and pyroglutamic acid arylamidase activities were not detected (Table 2). The peptidoglycan type is L-Lys-L-Ala₂-L-Ser. The major fatty acids identified are palmitic, myristic, oleic acid and stearic acid-dimethylacetal (Table S1). Phylogenetic analysis of *B. criceti* LMG 30188 shows that this strain is highly related to *B. choerinum* LMG 10510 and *B. italicum* LMG 30187. The type strain Ham19E (=LMG 30188^T=CCUG 70962^T) was isolated from the stool sample of a European hamster (*Cricetus cricetus*). The DNA G+C content is 62.53 %. Digital Protologue Taxonumber: TA00243.

Description of *Bifidobacterium anseris* sp. nov.

Bifidobacterium anseris (an.se.ris, N.L. gen. n. *anseris* of *anser*; common scientific name of domestic goose (*Anser domesticus*)). Cells are Gram-positive, non-motile, non-sporulating and F6PPK-positive and, when grown in MRS broth under anaerobic condition, was shown to exhibit a small rod morphology (Fig. 2). Cells grow under anaerobic and aerobic conditions. The colony diameter, when grown on MRS agar, ranges from 1.0 to 2.5 mm. Furthermore, colonies grown on the surface of an MRS agar plate are white and circular, while the embedded colonies are white and elliptical. Moreover, incubation under aerobic or anaerobic conditions show that this strain is able to grow from 25 to 42°C, yet incapable of growth at 20°C and 45°C (Table 1). Moreover, the cells grown at pH 5-6 (Table 1). Fermentation profiles of *B. anseris* LMG 30189 revealed that it is able to ferment a wide range of simple and complex carbohydrates, including galacto-oligosaccharides, lactose, glucose, glycogen, maltodextrin, maltose, melibiose, pullulan and

sucrose (i.e., reaching a final OD₆₀₀ of ≥ 0.51). In contrast, *B. anseris* LMG 30189 displays little if any growth on arabinogalactan, fructo-oligosaccharides, galactose, maltotriose, mannitol, mannose, raffinose, rhamnose, ribose, sorbitol, starch, turanose and xylose. (i.e., final OD₆₀₀ is ≤ 0.5) (Table 2). Furthermore, the strain does not exhibit growth on arabinose, cellobiose, fructose, fucose, inulin, N-acetyl-D-glucosamine, N-acetyl-D-galactosamine and trehalose. Positive enzymatic activities are observed for arginine dihydrolase, glutamic acid decarboxylase, alkaline phosphatase, arginine arylamidase, leucyl-glycine arylamidase and glutamyl glutamic acid (Table 2). Moreover, α -galactosidase, β -galactosidase, α -glucosidase, β -glucosidase, α -arabinosidase, β -glucuronidase, N-acetyl- β -glucosaminidase pyroglutamic acid arylamidase and serine arylamidase activity were not detected (Table 2). The peptidoglycan type is L-Lys-L-Ser-L-Ala. The major fatty acids identified are palmitic, myristic, oleic acid and stearic acid-dimethylacetal (Table S1). Phylogenetic analysis of *B. anseris* LMG 30189 highlights that this strain is highly related to *B. choerinum* LMG 10510 and *B. italicum* LMG 30187. The type strain Goo31D (=LMG 30189^T=CCUG 70960^T) was isolated from the stool sample of a domestic goose (*Anser domesticus*). The DNA G+C content is 64.3 %. Digital Protologue Taxonumber: TA00244.

Description of *Bifidobacterium parmae* sp. nov.

Bifidobacterium parmae (par.ma.e, N.L. gen. n. *parmae* of Parma; name of the city where the type strain was isolated and molecularly characterized). Cells are Gram-positive, non-motile, non-sporulating and F6PPK-positive and, when grown in MRS broth

under anaerobic condition, showed a rod morphology as well as cells with a branched or Y-shape, known as bifid-morphology (Fig. 2). Cells grow in anaerobic and aerobic conditions. The diameters of each colony, when grown on MRS agar, ranges from 1.0 to 2.0 mm. Furthermore, colonies grown on the surface of an MRS agar plate are white and circular, while embedded colonies are white and elliptical. Moreover, when incubated under aerobic or anaerobic conditions the strains is capable of growth from 25 to 42°C, yet is unable to grow at 20°C and 45°C (Table 1). Moreover, the strain was shown to grow at pH 4-6 (Table 1). Fermentation profiles of *B. parvae* LMG 30295 revealed that it is able to ferment a wide range of simple and complex carbohydrates, including cellobiose, fructo-oligosaccharides, fructose, galactose, galacto-oligosaccharides, glucose, glycogen, inulin, lactose, maltodextrin, maltose, maltotriose, mannitol, mannose, melibiose, N-acetyl-D-glucosamine, N-acetyl-D-galactosamine, raffinose, rhamnose, ribose, sorbitol, starch, sucrose, trehalose, turanose and xylose (i.e., reaching a final OD₆₀₀ of ≥ 0.51). In contrast, *B. parvae* LMG 30295 displays little if any growth on arabinose, arabinogalactan, fucose and pullulan (i.e., final OD₆₀₀ is ≤ 0.5) (Table 2). Positive enzymatic activities are observed for α -arabinosidase, glutamic acid decarboxylase, alkaline phosphatase, arginine arylamidase, leucyl-glycine arylamidase, pyroglutamic acid arylamidase and glutamyl glutamic acid. In contrast, weak activities were observed for β -galactosidase, β -glucuronidase and serine arylamidase (Table 2). Moreover, arginine dihydrolase, α -galactosidase, α -glucosidase, β -glucosidase, and N-acetyl- β -glucosaminidase

activities were not detected (Table 2). The peptidoglycan type is L-Lys-D-Asp. The major fatty acids identified are palmitic, myristic, oleic acid and stearic acid-dimethylacetal (Table S1). Phylogenetic analysis of *B. parvae* LMG 30295 highlights that this strain is highly related to *B. stellenboschense* DSM 23968. The type strain Uis4E (=LMG 30295^T=CCUG 70964^T) was isolated from the stool sample of a pygmy marmoset (*Callithrix pygmaea*). The DNA G+C content is 65.81 %. Digital Protologue Taxonumber: TA00242.

Description of *Bifidobacterium margollesii* sp. nov.

Bifidobacterium margollesii (mar.gol.les'i.i N.L. gen. n. *margollesii* of Margolles; named after Professor Abelardo Margolles, in recognition of his extensive contributions to our current knowledge on bifidobacterial biology). Cells are Gram-positive, non-motile, non-sporulating and F6PPK-positive and, when grown in MRS broth under anaerobic condition, were shown to exhibit a rod morphology as well as branched structures with a “Y” at the end, known as bifido-shaped (Fig. 2). Cells grow in anaerobic and aerobic conditions. The diameters of each colony, when grown on MRS agar, ranges from 1.0 to 2.5 mm. Furthermore, colonies grown on the surface of an MRS agar plate are white and circular, while embedded colonies are white and elliptical. Moreover, when grown under aerobic and anaerobic conditions for 48 hours the strain was shown to grow from 30 to 42°C, yet was not able to grow at 20°C, 25°C or 45°C (Table 1). Moreover, the cells grown at pH 4-6 (Table 1).

Fermentation profiles of *B. margollesii* LMG 30296 revealed that it is able to ferment a wide range of simple and complex carbohydrates, including

arabinogalactan, cellobiose, fructo-oligosaccharides, galactose, glucose, inulin, lactose, maltotriose, mannitol, mannose, melibiose, N-acetyl-D-glucosamine, N-acetyl-D-galactosamine, ribose, sorbitol, starch, trehalose, turanose and xylose (i.e., reaching a final OD₆₀₀ of ≥ 0.51). In contrast, *B. margollesii* LMG 30296 displays little if any growth on arabinose, fructose, fucose, galacto-oligosaccharides, glycogen, maltodextrin, maltose, pullulan, raffinose, rhamnose and sucrose (i.e., final OD₆₀₀ is ≤ 0.5) (Table 2). Positive enzymatic activities were observed for arginine dihydrolase, alkaline phosphatase, arginine arylamidase and glutamyl glutamic acid (Table 2). Moreover, α -galactosidase, β -galactosidase, α -glucosidase, β -glucosidase, α -arabinosidase, β -glucuronidase, N-acetyl- β -glucosaminidase, glutamic acid decarboxylase, leucyl-glycine arylamidase, pyroglutamic acid arylamidase and serine arylamidase activities were not detected (Table 2).

The peptidoglycan type is L-Lys-L-Ala-L-Ser. The major fatty acids identified are palmitic, myristic, oleic acid and stearic acid-dimethylacetal (Table S1). Phylogenetic analysis of *B. margollesii* LMG 30296 highlights that this strain is highly related to *B. hapali* DSM 100202, *B. tissieri* DSM100201 and *B. vansinderenii* LMG 30126.

The type strain Uis1B (=LMG 30296^T=CCUG 70959^T) was isolated from the stool sample of a Pygmy marmoset (*Callithrix pygmaea*). The DNA G+C content is 61.91 %. Digital Protologue Taxonumber: TA00241.

Description of *Bifidobacterium imperatoris* sp. nov. *Bifidobacterium imperatoris* (i.mpe.ra.to.ris N.L. gen. n *imperatoris* of imperator; common

scientific name of *Saguinus imperator*). Cells are Gram-positive, non-motile, non-sporulating and F6PPK-positive and, when grown in MRS broth under anaerobic conditions, exhibits a rod-shape morphology as well as a branched structure with a “Y” at the end, known as bifido-shaped (Fig. 2). Cells grow under anaerobic and aerobic conditions. The diameters of each colony, when grown on MRS agar, ranges from 1.0 to 3 mm. Furthermore, colonies grown on the surface of an MRS agar plate are white and circular, while embedded colonies are white and elliptical. Moreover, the strain when incubated under aerobic conditions for 48 hours was shown to grow from 30 to 42°C, while under anaerobic conditions it was shown to exhibit growth at a temperature range from 25°C to 42°C, yet are unable to grow at 20°C and 45°C (Table 1). Moreover, the cells exhibited growth at pH 5-6 (Table 1).

Fermentation profiles of *B. imperatoris* LMG 30297 revealed that it is able to ferment a wide range of simple and complex carbohydrates, including arabinose, fructo-oligosaccharides, galactose, glucose, glycogen lactose, galacto-oligosaccharides, glycogen, maltodextrin, maltose, maltotriose, mannitol, raffinose, ribose, sucrose, turanose and xylose (i.e., reaching a final OD₆₀₀ of ≥ 0.51). In contrast, *B. imperatoris* LMG 30297 displays little if any growth on arabinogalactan, cellobiose, fructose, fucose, inulin, mannose, N-acetyl-D-glucosamine, N-acetyl-D-galactosamine, pullulan, rhamnose, sorbitol, starch and trehalose (i.e., final OD₆₀₀ is ≤ 0.5) (Table 2). Positive enzymatic activities are observed for arginine dihydrolase, glutamic acid decarboxylase, alkaline phosphatase, arginine arylamidase, leucyl-glycine arylamidase, and

glutamyl glutamic acid. In contrast, little activity was observed for α -arabinosidase and serine arylamidase (Table 2). Moreover, α -galactosidase, β -galactosidase, α -glucosidase, β -glucosidase, β -glucuronidase, N-acetyl- β -glucosaminidase and pyroglutamic acid arylamidase activities were not detected (Table 2).

The peptidoglycan type is L-Lys-Gly. The major fatty acids identified are palmitic, myristic, oleic acid and stearic acid-dimethylacetal (Table S1).

Phylogenetic analysis of *B. imperatoris* LMG 30297 highlights that this strain is highly related to *B. saguini* DSM 23967. The type strain Tam1G (=LMG 30297^T=CCUG 70961^T) was isolated from the stool sample of an emperor tamarin (*Saguinus imperator*). The DNA G+C content is 56.13 %. Digital Protologue Taxonumber: TA00245.

Conclusions

A combination of phylogenetic, genomic and phenotypic analyses allowed to differentiate strains Goo31D, Ham19E, Rab10A, Tam1G, Uis4E and Uis1B from each of the type strains of the so far recognized *Bifidobacterium* species. In this context, the combined phylogenetic analyses involving 16S rRNA gene, ITS, selected housekeeping genes and core genome sequences reveal that the isolated strains exhibit close phylogenetic relatedness with *Bifidobacterium choerinum* LMG 10510, *Bifidobacterium hapali* DSM 100202, *Bifidobacterium saguini* DSM 23967 and *Bifidobacterium stellenboschense* DSM 23968. Recently, three putative novel species, i.e., *Bifidobacterium aerophilum*, *Bifidobacterium avesanii* and *Bifidobacterium ramosum* have been

proposed by Michelini et al. [12]. However, we have decided not to include these taxa in our analyses because they have not been validated by the International Committee on Systematic Bacteriology. Nevertheless, comparison between 16S partial gene sequences of these putative novel species showed identity values below 95.6 % when compared with Goo31D, Ham19E, Rab10A, Tam1G, Uis4E and Uis1B 16S sequences, clearly supporting that these letter strains are belonging to different bifidobacterial species. The performed ANI analyses that are aimed at assessing genetic similarity levels between the six novel taxa showed strains Goo31D, Ham19E, Rab10A, Uis4E and Uis1B do not belong to any other identified bifidobacterial species sequenced so far, and are thus proposed as *Bifidobacterium anseris* sp. nov. (Goo31D = LMG 30189^T = CCUG 70960^T), *Bifidobacterium criceti* sp. nov. (Ham19E = LMG 30188^T = CCUG 70962^T), *Bifidobacterium italicum* sp. nov. (Rab10A = LMG 30187^T = CCUG 70963^T), *Bifidobacterium margollesii* sp. nov. (Uis1B = LMG 30296^T = CCUG 70959^T) and *Bifidobacterium parmae* sp. nov. (Uis4E = LMG 30295^T = CCUG 70964^T), respectively. Furthermore, Tam1G displays ANI percentage with *B. saguini* DSM 23967 similar to the threshold value used for species recognition [31]. Nevertheless, GGDC analysis highlights an estimated DDH value well below 70 %, so we propose to assign the name *Bifidobacterium imperatoris* sp. nov. (Tam1G = LMG 30297^T = CCUG 70961^T) to this isolate. In addition, in order to increase the number of bifidobacterial isolates and to provide a complete overview of the bifidobacterial population in other mammalian species, further

isolation attempts are ongoing based on metagenomics analyses performed.

Acknowledgments

This work was funded by the EU Joint Programming Initiative – A Healthy Diet for a Healthy Life (JPI HDHL, <http://www.healthydietforhealthylife.eu/>) to DvS (in conjunction with Science Foundation Ireland [SFI], Grant number 15/JP-HDHL/3280) and to MV (in conjunction with MIUR, Italy). We thank GenProbio srl for financial support of the Laboratory of Probiogenomics. Part of this research is conducted using the High Performance Computing (HPC) facility of the University of Parma. LM is supported by Fondazione Cariparma, Parma, Italy. DvS is a member of The APC Microbiome Institute funded by Science Foundation Ireland (SFI), through the Irish Government's National Development Plan (Grant number SFI/12/RC/2273). The authors declare that they have no competing interests.

References

- [1] M. Ventura, C. Canchaya, A. Tauch, G. Chandra, G.F. Fitzgerald, K.F. Chater, D. van Sinderen, Genomics of Actinobacteria: tracing the evolutionary history of an ancient phylum, *Microbiology and molecular biology reviews* : MMBR, 71 (2007) 495-548.
- [2] M. Ventura, F. Turrone, G.A. Lugli, D. van Sinderen, Bifidobacteria and humans: our special friends, from ecological to genomics perspectives, *J Sci Food Agr*, 94 (2014) 163-168.
- [3] D.G. Bergey, W. WB, Parte AC: The Actinobacteria, in: *Bergey's Manual of Systematic Bacteriology*, 2012.
- [4] F. Turrone, C. Milani, S. Duranti, C. Ferrario, G.A. Lugli, L. Mancabelli, D. van Sinderen, M. Ventura, Bifidobacteria and the infant gut: an example of co-evolution and natural selection, *Cellular and molecular life sciences* : CMLS, (2017).
- [5] J.L. Round, S.K. Mazmanian, The gut microbiota shapes intestinal immune responses during health and disease, *Nature reviews. Immunology*, 9 (2009) 313-323.
- [6] F. Turrone, V. Taverniti, P. Ruas-Madiedo, S. Duranti, S. Guglielmetti, G.A. Lugli, L. Gioiosa, P. Palanza, A. Margolles, D. van Sinderen, M. Ventura, *Bifidobacterium bifidum* PRL2010 modulates the host innate immune response, *Applied and environmental microbiology*, 80 (2014) 730-740.
- [7] E. Avershina, K. Lundgard, M. Sekelja, C. Dotterud, O. Storro, T. Oien, R. Johnsen, K. Rudi, Transition from infant- to adult-like gut microbiota, *Environmental microbiology*, 18 (2016) 2226-2236.
- [8] V. Bunesova, E. Vlkova, V. Rada, J. Killer, S. Musilova, Bifidobacteria from the gastrointestinal tract of animals: differences and similarities, *Beneficial microbes*, 5 (2014) 377-388.
- [9] C. Milani, L. Mancabelli, G.A. Lugli, S. Duranti, F. Turrone, C. Ferrario, M. Mangifesta, A. Viappiani, P. Ferretti, V. Gorfer, A. Tett, N. Segata, D. van Sinderen, M. Ventura, Exploring Vertical Transmission of Bifidobacteria from Mother to Child, *Applied and environmental microbiology*, 81 (2015) 7078-7087.
- [10] F. Turrone, E. Foroni, P. Pizzetti, V. Giubellini, A. Ribbera, P. Merusi, P. Cagnasso, B. Bizzarri, G.L. de'Angelis, F. Shanahan, D. van Sinderen, M. Ventura, Exploring the diversity of the bifidobacterial population in the human intestinal tract, *Applied and environmental microbiology*, 75 (2009) 1534-1545.
- [11] S. Duranti, M. Mangifesta, G.A. Lugli, F. Turrone, R. Anzalone, C. Milani, L. Mancabelli, M.C. Ossiprandi, M. Ventura, *Bifidobacterium vansinderenii* sp. nov. isolated from faeces of emperor tamarin (*Saguinus imperator*). *International Journal of Systematic and Evolutionary Microbiology*., In press (2017).
- [12] S. Michelini, M. Modesto, G. Filippini, C. Spiezio, C. Sandri, B. Biavati, A. Pisi, P. Mattarelli, *Bifidobacterium aerophilum* sp. nov., *Bifidobacterium avesanii* sp. nov. and *Bifidobacterium ramosum* sp. nov.: Three novel taxa from the faeces of cotton-top tamarin (*Saguinus oedipus* L.), *Syst Appl Microbiol*, 39 (2016) 229-236.
- [13] S. Michelini, K. Oki, E. Yanokura, Y. Shimakawa, M. Modesto, P. Mattarelli, B. Biavati, K. Watanabe, *Bifidobacterium myosotis* sp. nov., *Bifidobacterium tissieri* sp. nov. and *Bifidobacterium hapali* sp. nov., isolated from faeces of baby common marmosets (*Callithrix jacchus* L.), *International journal of systematic and evolutionary microbiology*, 66 (2016) 255-265.
- [14] J. Praet, I. Meeus, M. Cnockaert, M. Aerts, G. Smaghe, P. Vandamme, *Bifidobacterium commune* sp. nov. isolated from the bumble bee gut, *Antonie van Leeuwenhoek*, 107 (2015) 1307-1313.
- [15] D. Laureys, M. Cnockaert, L. De Vuyst, P. Vandamme, *Bifidobacterium aquikefiri* sp. nov., isolated from water kefir, *International journal of systematic and evolutionary microbiology*, (2016).
- [16] M. Modesto, S. Michelini, I. Stefanini, C. Sandri, C. Spiezio, A. Pisi, G. Filippini, B. Biavati, P. Mattarelli, *Bifidobacterium lemorum* sp. nov., from faeces of the ring-tailed lemur (*Lemur catta*), *International journal of systematic and evolutionary microbiology*, 65 (2015) 1726-1734.
- [17] S. Michelini, M. Modesto, A.M. Pisi, G. Filippini, C. Sandri, C. Spiezio, B. Biavati, B. Sgorbati, P. Mattarelli, *Bifidobacterium eulemuris* sp. nov. isolated from the faeces of the black lemur (*Eulemur macaco*), *International journal of systematic and evolutionary microbiology*, (2016).
- [18] C. Milani, G.A. Lugli, S. Duranti, F. Turrone, F. Bottacini, M. Mangifesta, B. Sanchez, A. Viappiani, L. Mancabelli, B. Taminau, V. Delcenserie, R. Barrangou, A. Margolles, D. van Sinderen, M. Ventura, Genomic Encyclopedia of Type Strains of the Genus *Bifidobacterium*, *Applied and environmental microbiology*, 80 (2014) 6290-6302.

- [19] C. Milani, M. Mangifesta, L. Mancabelli, G.A. Lugli, K. James, S. Duranti, F. Turrone, C. Ferrario, M.C. Ossiprandi, D. van Sinderen, M. Ventura, Unveiling bifidobacterial biogeography across the mammalian branch of the tree of life, *ISME J*, (2017).
- [20] F. Turrone, J.R. Marchesi, E. Foroni, M. Gueimonde, F. Shanahan, A. Margolles, D. van Sinderen, M. Ventura, Microbiomic analysis of the bifidobacterial population in the human distal gut, *The ISME journal*, 3 (2009) 745-751.
- [21] M. Ventura, R. Zink, Rapid identification, differentiation, and proposed new taxonomic classification of *Bifidobacterium lactis*, *Applied and environmental microbiology*, 68 (2002) 6429-6434.
- [22] C. Milani, A. Hevia, E. Foroni, S. Duranti, F. Turrone, G.A. Lugli, B. Sanchez, R. Martin, M. Gueimonde, D. van Sinderen, A. Margolles, M. Ventura, Assessing the fecal microbiota: an optimized ion torrent 16S rRNA gene-based analysis protocol, *PloS one*, 8 (2013) e68739.
- [23] C. Milani, G.A. Lugli, F. Turrone, L. Mancabelli, S. Duranti, A. Viappiani, M. Mangifesta, N. Segata, D. van Sinderen, M. Ventura, Evaluation of bifidobacterial community composition in the human gut by means of a targeted amplicon sequencing (ITS) protocol, *FEMS microbiology ecology*, 90 (2014) 493-503.
- [24] G.A. Lugli, C. Milani, L. Mancabelli, D. van Sinderen, M. Ventura, MEGAnnotator: a user-friendly pipeline for microbial genomes assembly and annotation, *FEMS microbiology letters*, 363 (2016).
- [25] B. Chevreur, T. Pfisterer, B. Drescher, A.J. Driesel, W.E. Muller, T. Wetter, S. Suhai, Using the miraEST assembler for reliable and automated mRNA transcript assembly and SNP detection in sequenced ESTs, *Genome research*, 14 (2004) 1147-1159.
- [26] D. Hyatt, G.L. Chen, P.F. Locascio, M.L. Land, F.W. Larimer, L.J. Hauser, Prodigal: prokaryotic gene recognition and translation initiation site identification, *Bmc Bioinformatics*, 11 (2010) 119.
- [27] Y. Zhao, J. Wu, J. Yang, S. Sun, J. Xiao, J. Yu, PGAP: pan-genomes analysis pipeline, *Bioinformatics*, 28 (2012) 416-418.
- [28] A.J. Enright, S. Van Dongen, C.A. Ouzounis, An efficient algorithm for large-scale detection of protein families, *Nucleic acids research*, 30 (2002) 1575-1584.
- [29] K. Katoh, K. Misawa, K. Kuma, T. Miyata, MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform, *Nucleic acids research*, 30 (2002) 3059-3066.
- [30] M.A. Larkin, G. Blackshields, N.P. Brown, R. Chenna, P.A. McGettigan, H. McWilliam, F. Valentin, I.M. Wallace, A. Wilm, R. Lopez, J.D. Thompson, T.J. Gibson, D.G. Higgins, Clustal W and Clustal X version 2.0, *Bioinformatics*, 23 (2007) 2947-2948.
- [31] M. Richter, R. Rossello-Mora, Shifting the genomic gold standard for the prokaryotic species definition, *Proceedings of the National Academy of Sciences of the United States of America*, 106 (2009) 19126-19131.
- [32] J.J. Campanella, L. Bitincka, J. Smalley, MatGAT: an application that generates similarity/identity matrices using protein or DNA sequences, *Bmc Bioinformatics*, 4 (2003) 29.
- [33] J.P. Meier-Kolthoff, A.F. Auch, H.P. Klenk, M. Goker, Genome sequence-based species delimitation with confidence intervals and improved distance functions, *BMC bioinformatics*, 14 (2013) 60.
- [34] P. Schumann, Peptidoglycan Structure, *Method Microbiol*, 38 (2011) 101-129.
- [35] L.T. Miller, Single derivatization method for routine analysis of bacterial whole-cell fatty acid methyl esters, including hydroxy acids, *Journal of clinical microbiology*, 16 (1982) 584-586.
- [36] L.D. Kuykendall, M.A. Roy, J.J. Oneill, T.E. Devine, Fatty-Acids, Antibiotic-Resistance, and Deoxyribonucleic-Acid Homology Groups of *Bradyrhizobium-Japonicum*, *International journal of systematic bacteriology*, 38 (1988) 358-361.
- [37] M. Ventura, C. Canchaya, A. Del Casale, F. Dellaglio, E. Neviani, G.F. Fitzgerald, D. van Sinderen, Analysis of bifidobacterial evolution using a multilocus approach, *International journal of systematic and evolutionary microbiology*, 56 (2006) 2783-2792.
- [38] G.A. Lugli, C. Milani, F. Turrone, S. Duranti, C. Ferrario, A. Viappiani, L. Mancabelli, M. Mangifesta, B. Taminau, V. Delcenserie, D. van Sinderen, M. Ventura, Investigation of the evolutionary development of the genus *Bifidobacterium* by comparative genomics, *Applied and environmental microbiology*, 80 (2014) 6383-6394.
- [39] K.T. Konstantinidis, J.M. Tiedje, Genomic insights that advance the species definition for prokaryotes, *Proceedings of the National Academy of Sciences of the United States of America*, 102 (2005) 2567-2572.

Table 1. Phenotypic information for strains representing the putative novel species and their closest related species.

Temperature	Tam 1G		<i>B. saguini</i> DSM 23967		Uis 4E		Uis 1B		<i>B. hapali</i> DSM 100202		<i>B. fissieri</i> DSM 100201		<i>B. stellenboschense</i> DSM 23968		Rab10A		Ham19E		Goo31D		<i>B. choerinum</i> LMG 10510	
	Ae	An	Ae	An	Ae	An	Ae	An	Ae	An	Ae	An	Ae	An	Ae	An	Ae	An	Ae	An	Ae	An
20°C	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
25°C	-	++	-	+	-	++	-	-	-	-	-	-	-	+	+++	++	+	++	+	++	-	-
30°C	++	+++	+	+	++	++	++	++	++	++	++	++	++	+	+++	++	++	++	++	++	++	++
35°C	+	+++	++	++	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++
37°C	+	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++
40°C	+++	++	++	++	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++
42°C	++++	+++	+++	+++	++++	+++	+++	+++	+++	+++	+++	+++	+++	+	+++	++	+	+	+	+	+	+++
45°C	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
pH values	24h	48h	24h	48h	24h	48h	24h	48h	24h	48h	24h	48h	24h	48h	24h	48h	24h	48h	24h	48h	24h	48h
3.5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
4	-	-	-	-	+	+	++	++	+	+	+	+	+	+	+	+	+	+	+	+	+	+
4.5	-	-	-	-	+	+	+++	+++	++	++	++	++	++	++	+	+	+	+	+	+	+	+
5	+	++	+	+	++	+++	+++	+++	++	++	++	++	++	++	+	+	+	+	+	+	+	++
5.5	++	+++	++	++	+++	+++	+++	+++	++	++	++	++	++	++	++	++	++	++	++	++	++	++
6	+++	++++	+++	+++	+++	+++	+++	+++	++	++	++	++	++	++	++	++	++	++	++	++	++	++

Ae aerobic; An anaerobic.

Symbols: -, no growth; + poor growth as measured by $0.5 \leq OD_{600} \leq 0.8$; ++ growth as measured by $0.51 \leq OD_{600} \leq 0.8$; +++ good growth as measured by $0.81 \leq OD_{600} \leq 1.9$; ++++ very good growth as measured by $OD_{600} \geq 1.91$.

Table 2. Fermentation profiles and enzymatic activities of novel bifidobacterial species and their closest related species.

Fermentation	Tam 1G		<i>B. saguini</i> DSM 23967		Uis 4E		Uis 1B		<i>B. hapali</i> DSM 100202		<i>B. tissieri</i> DSM 100201		<i>B. stellenboschense</i> DSM 23968		Rab10A		Ham19E		Goo31D		<i>B. choerinum</i> LMG 10510	
	2 4	4 8	2 4	4 8	2 4	4 8	2 4	4 8	2 4	4 8	2 4	4 8	2 4	4 8	2 4	4 8	2 4	4 8	2 4	4 8	2 4	4 8
Arabinose	+	++	+	+	+	+	-	+	+	+	-	+	+	+	-	-	-	-	-	-	-	-
Arabinogalactan	+	+	+	+	+	+	++	++	-	+	+	+	+	+	+	+	+	+	+	+	+	+
Cellobiose	+	+	+	+	++	++	++	++	+	+	+	++	+	+	-	-	-	-	-	-	-	-
Fructo-oligosaccharides	++	++	++	++	++	++	++	++	+	+	+	++	+	+	+	+	+	+	+	+	+	+
Fructose	+	+	++	++	++	++	+	+	++	++	+	++	++	++	-	-	-	-	-	-	-	-
Fucose	+	+	+	+	+	+	+	+	+	+	-	+	+	+	-	-	-	-	-	-	-	-
Galactose	+	++	+	++	+	++	++	++	+	+	+	++	++	++	+	+	+	+	+	+	+	+
Galacto-oligosaccharides	++	++	++	++	++	++	+	+	+	++	+	++	++	++	+	+	+	+	+	+	+	+
Glucose	++	++	++	++	++	++	++	++	++	++	++	++	++	++	+	+	+	+	+	+	+	++
Glycogen	++	++	++	++	++	++	+	+	-	++	-	++	++	++	+	+	+	+	+	+	+	++
Inulin	+	+	+	+	++	++	++	++	+	+	+	++	++	++	+	+	+	+	-	-	+	+
Lactose	++	++	++	++	++	++	++	++	++	++	-	++	++	++	+	+	+	+	+	+	+	++
Maltodextrin	++	++	++	++	+	++	+	+	++	++	+	++	++	++	+	+	+	+	+	+	+	++
Maltose	++	++	++	++	++	++	+	+	+	+	+	++	++	++	+	+	+	+	+	+	+	++
Maltotriose	++	++	++	++	++	++	++	++	+	++	+	++	++	++	+	+	+	+	+	+	+	++
Mannitol	+	++	+	+	+	++	++	++	+	+	+	++	++	++	+	+	+	+	+	+	+	+
Mannose	+	+	+	+	++	++	++	++	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Melibiose	+	++	++	++	+	++	-	-	+	++	+	++	++	++	+	+	+	+	+	+	+	++
N-Acetyl-D-galactosamine	+	+	+	+	++	++	++	++	+	+	+	+	+	+	-	+	+	+	-	-	-	-
N-Acetyl-D-glucosamine	+	+	+	+	++	++	++	++	+	+	+	++	++	++	-	+	-	+	-	-	+	+
Pullulan	+	+	+	+	+	+	+	+	-	+	-	+	+	++	+	+	+	+	+	+	+	++
Raffinose	++	++	++	++	++	++	+	+	++	++	-	++	++	++	+	+	+	+	+	+	+	++
Rhamnose	+	+	+	+	++	++	+	+	+	+	+	+	+	+	-	-	+	+	-	+	-	-
Ribose	+	++	++	++	+	++	++	++	+	+	+	+	++	++	+	+	+	+	+	+	+	+
Sorbitol	+	+	+	+	++	++	++	++	+	+	+	+	++	++	+	+	+	+	-	+	+	+
Starch	+	+	+	+	++	++	++	++	+	+	+	+	++	++	+	+	+	+	+	+	+	+
Sucrose	++	++	+	++	+	+	-	+	++	+	+	+	++	++	+	+	+	+	+	+	+	++
Trehalose	+	+	+	+	++	++	++	++	+	+	+	++	++	++	+	+	+	+	-	-	+	+
Turanose	+	++	+	++	+	++	++	++	+	++	++	++	++	++	+	+	+	+	+	+	+	+
Xylose	+	++	-	-	+	++	++	++	-	-	-	-	-	-	-	-	-	+	-	+	+	+

Enzymatic activity											
Arginine dihydrolase	+	+	-	+	+	+	-	+	+	+	w
α - Galactosidase	-	-	-	-	-	-	w	-	-	-	-
β - Galactosidase	-	-	w	-	-	-	w	-	-	-	-
α - Glucosidase	-	-	-	-	-	-	w	-	-	-	-
β - Glucosidase	-	-	-	-	-	-	-	-	-	-	-
α - Arabinosidase	w	+	+	-	-	-	+	-	-	-	-
β - Arabinosidase	-	-	w	-	-	-	+	-	-	-	-
Glucuronidase	-	-	-	-	-	-	-	-	-	-	-
N-acetyl-β - Glucosaminidase	-	w	-	-	-	-	-	-	-	-	w
Glutamic acid decarboxylase	+	-	+	-	-	-	+	+	-	+	+
Alkaline phosphatase	+	+	+	+	+	+	+	+	+	+	+
Arginine arylamidase	+	+	+	+	+	+	+	+	+	+	+
Leucyl-glycine arylamidase	+	+	+	-	+	+	+	+	w	+	+
Pyroglutamic acid arylamidase	-	-	+	-	+	-	+	-	-	-	-
Glutamyl glutamic acid arylamidase	+	+	+	+	+	-	+	+	+	+	+
Serine arylamidase	w	+	w	-	+	-	+	w	+	-	+

24 twenty-four hours; 48 forty-eight hours.

Symbols: -, no growth; + poor growth as measured by $0.5 \leq OD_{600}$; ++ growth as measured by $0.51 \leq OD_{600} \leq 0.8$; +++ good growth as measured by $0.81 \leq OD_{600} \leq 1.9$; ++++ very good growth as measured by $OD_{600} \geq 1.91$.

Enzymatic activity: +, Positive; -, Negative; w, Weakly Positive

Table 3. General Genetic features

	Rab10A	Ham19E	Goo31D	Tam1G	Uis4E	Uis1B
Biological origin	European rabbit	European hamster	Domestic goose	Emperor tamarin	Pygmy marmoset	Pygmy marmoset
Average Coverage	104.3	108.6	115.4	55.52	139.27	66.93
Contigs	72	44	9	62	44	80
Genome length	2,276,351	2,155,882	2,166,761	2,639,899	2,820,211	2,789,387
Average GC percentage	65.45	62.53	64.3	56.13	65.81	61.91
Predicted ORFs	1,825	1,733	1,681	2,215	2,247	2,281
tRNA	52	53	52	59	62	60
rRNA*	4	4	4	4	6	3
16S identity	98.36% <i>B. choerinum</i>	97.32% <i>B. choerinum</i>	97.65% <i>B. choerinum</i>	99% <i>B. saguini</i>	96.8% <i>B. stellenbosche nse</i>	96.2% <i>B. tissieri</i>
ITS identity	87.2% <i>B. choerinum</i>	68.6% <i>B. animalis</i>	84.7% <i>B. choerinum</i>	89% <i>B. saguini</i>	81.3% <i>B. stellenbosche nse</i>	65.7% <i>B. biavatii</i>
ANI value	91.8% <i>B. choerinum</i>	87.81% <i>B. choerinum</i>	91.47% <i>B. choerinum</i>	94.55% <i>B. saguini</i>	93.45% <i>B. stellenbosche nse</i>	88.04% <i>B. hapali</i>

*Predicted number of rRNA loci

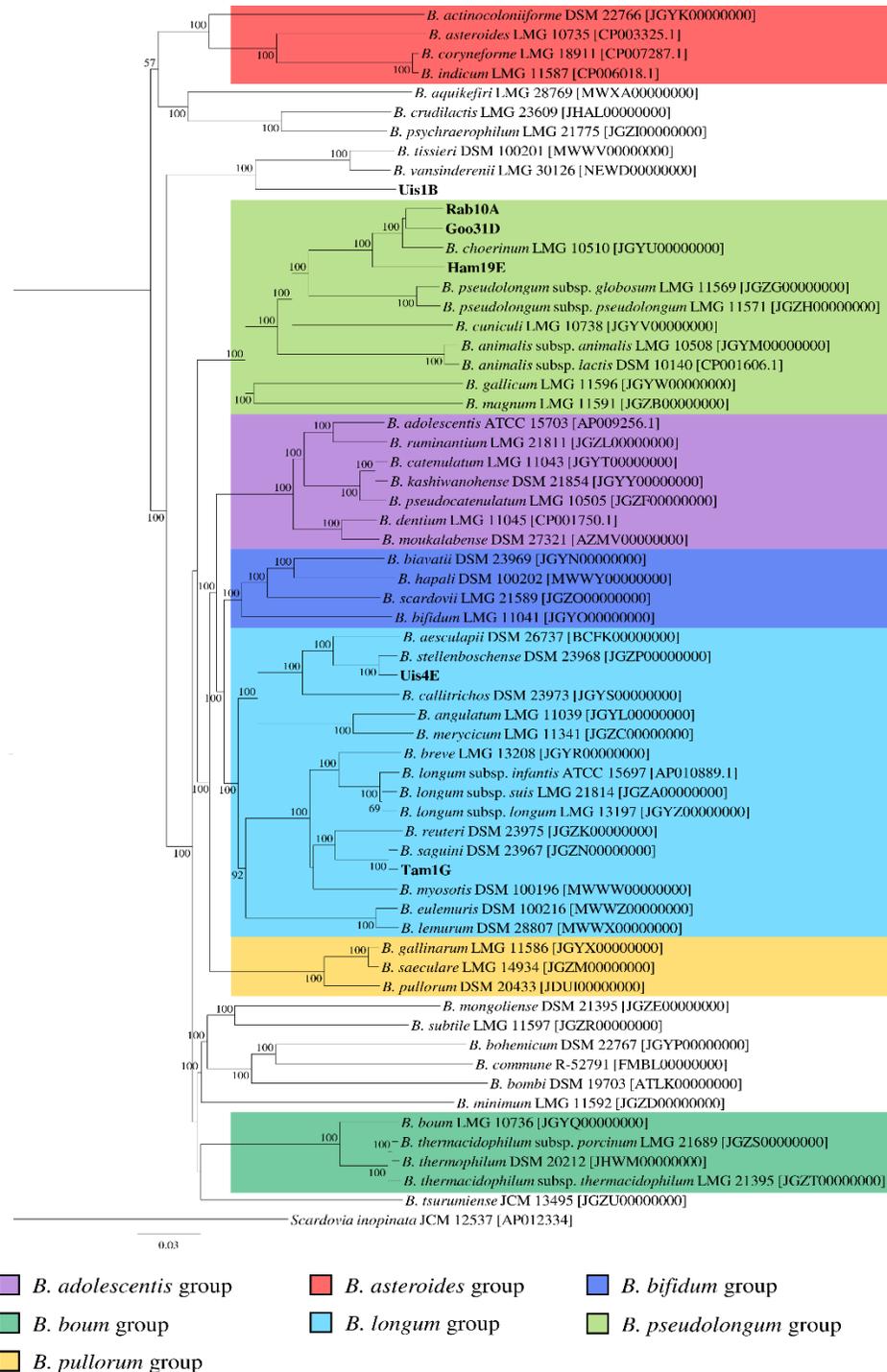


Figure 1. Phylogenetic tree of the *Bifidobacterium* genus based on the concatenation of 233 core amino acid sequence genes of Goo31D, Ham19E, Rab10A, Tam1G, Uis4E and Uis1B and members of the *Bifidobacterium* genus. The core genes-based tree shows the subdivision of the seven phylogenetic groups of the *Bifidobacterium* genus represented with different colors. The phylogenetic tree was built by the neighbor-joining method with corresponding sequences of *Scardovia inopinata* JCM 12537 being employed as outgroup. Bootstrap percentages above 50 are shown at node points, based on 1000 replicates of the phylogenetic tree.

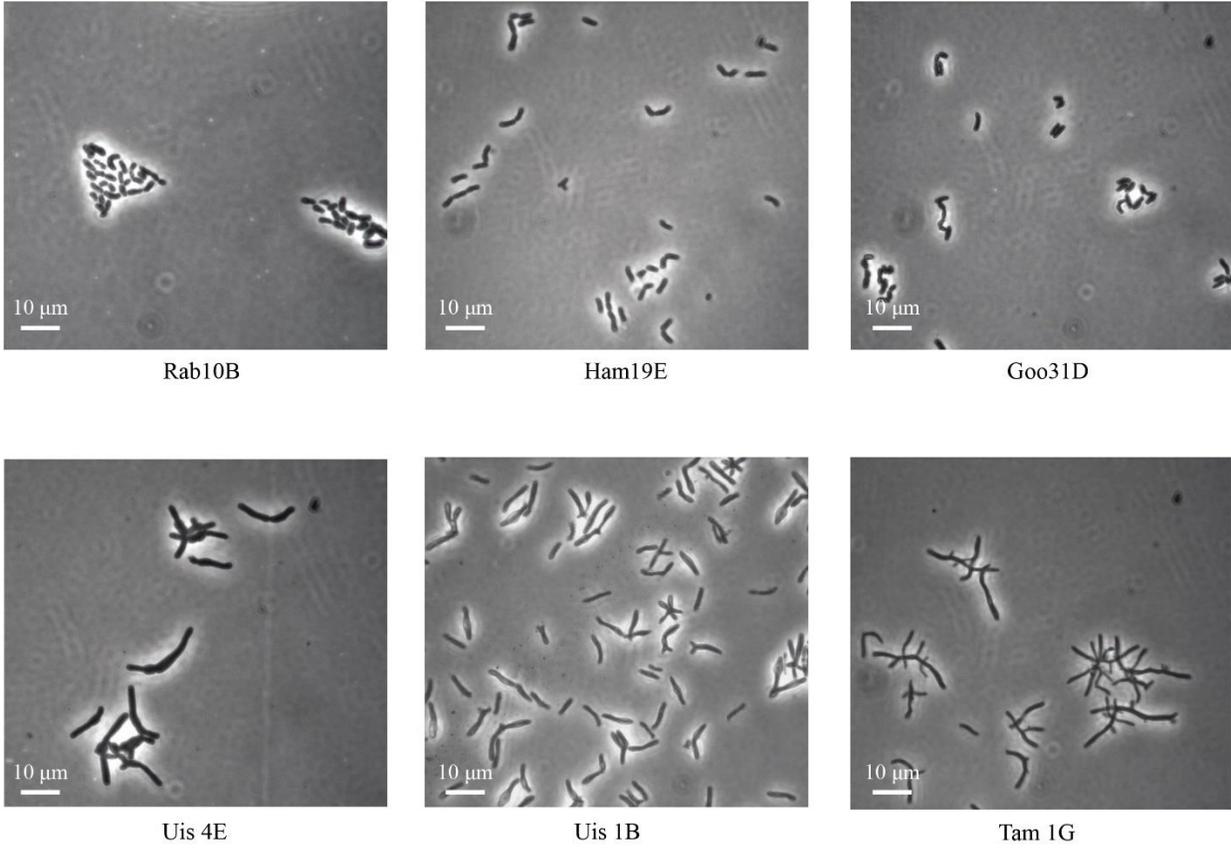


Figure 2. Rab10A, Ham19E, Goo31D, Uis4E, Uis1B and Tam1G cellular morphologies as determined by the use of phase-contrast microscopy. Bar, 10 μm.

Chapter 6

The Sortase-Dependent Fimbriome of the Genus *Bifidobacterium*: Extracellular Structures with Potential To Modulate Microbe-Host Dialogue

Milani C, Mangifesta M, Mancabelli L, Lugli GA, Mancino W,
Viappiani A, Faccini A, van Sinderen D, Ventura M, Turrone F.

This chapter was published in:

Appl Environ Microbiol. 2017 Sep 15;83(19). pii: e01295-17.. Print 2017 Oct 1.

Reprinted with permission from American Society for Microbiology.

The sortase-dependent fimbriome of the genus *Bifidobacterium*: extracellular structures with potential to modulate microbe-host dialogue

Running title: The bifidobacterial fimbriome

Key words: bifidobacteria, gut microbiota, bacterial interactions, genomics, metagenomics

Christian Milani¹, Marta Mangifesta², Leonardo Mancabelli¹, Gabriele Andrea Lugli¹, Walter Mancino¹, Alice Viappiani², Andrea Faccini³, Douwe van Sinderen⁴, Marco Ventura¹ and
Francesca Turroni¹

Laboratory of Probiogenomics, Department of Chemistry, Life Sciences and Environmental Sustainability, University of Parma, Parma, Italy¹; GenProbio srl, Parma, Italy²; Interdepartmental Measure Centre Giuseppe Casnati, University of Parma, Parma, Italy³; APC Microbiome Institute and School of Microbiology, Bioscience Institute, National University of Ireland, Cork, Ireland⁴

Corresponding author: Mailing address for Francesca Turroni Department of Chemistry, Life Sciences and Environmental Sustainability, University of Parma, Parco Area delle Scienze 11/A, 43124 Parma, Italy. Phone: ++39-521-905666. Fax: ++39-521-905604. E-mail: francesca.turroni@unipr.it

Abstract

Bifidobacteria are important gut commensals of mammals, including humans, of any age. However, the molecular mechanisms by which these microorganisms establish themselves in the mammalian gut and persist in this environment are largely unknown. Here, we analyzed the genetic diversity of the predicted arsenal of sortase-dependent pili of known and sequenced members of the *Bifidobacterium* genus, and constructed a bifidobacterial sortase-dependent fimbriome database. Our analyses revealed considerable genetic variability of the sortase-dependent fimbriome among bifidobacterial (sub)species, which appears to have been due to horizontal gene transfer events, and for which we were able to perform an evolutionary mapping. Functional assessment by transcriptome analysis and binding assays involving different substrates, demonstrate how bifidobacterial pili are pivotal in promoting various adhesion abilities to glycans and extracellular matrix proteins, thereby supporting the ecological success of bifidobacteria in the mammalian gut.

Importance

Adhesion of bifidobacterial cells to the mucosa of the large intestine is considered a hallmark for the persistence and colonization of these bacteria in the human gut. In this context, we analyzed the genetic diversity of the predicted arsenal of sortase-dependent pili of known and sequenced

members of the *Bifidobacterium* genus, and constructed a bifidobacterial sortase-dependent fimbriome database. Our analyses revealed considerable genetic variability of the sortase-dependent fimbriome among bifidobacterial (sub)species, which appears to have been due to horizontal gene transfer events. In addition, functional assessment by transcriptome analysis and binding assays involving different substrates, demonstrate how bifidobacterial pili are crucial in promoting various adhesion abilities to glycans and extracellular matrix proteins, thereby supporting the ecological success of bifidobacteria in the mammalian gut. This study represents a complete genomic study regarding the presence of fimbriae in the genus *Bifidobacterium*.

Introduction

Bifidobacteria are microorganisms that are known to colonize the gut of various mammals, including humans, birds and social insects (1-3). In this gut ecosystem, bifidobacteria interact with their host as well as with other members of the microbiota through different strategies (4-7). Bacterial surface appendages, in particular pili or fimbriae, are considered to be important bacterial structures involved in host-microbe interactions (8). Pili were observed in Gram positive and in Gram negative bacteria, and they are found to elicit functions in addition to adherence, such as conjugation, motility, immune-modulation, biofilm formation and electron transfer (9, 10). Structurally, Gram-

positive pili contain multiple copies of one or more different pilin proteins. The pilus shaft is composed of a multimer of the major pilin, and is in turn associated with a smaller number of ancillary pilins (11). Various types of pili have been characterized including sortase-dependent (SD) pili (Types I and II) and Type IV pili (12). In SD pili the pilins that constitute the pilus fiber are covalently linked by sortases (13). The genes encoding an SD pilus structure including the corresponding sortase are generally located in the same locus. Pilus anchoring takes place on the cell wall following Sec-dependent secretion of pilus protein components.

The existence of SD pili in bifidobacteria has only recently been appreciated (14). Bifidobacterial SD pili loci are typically composed of: (i) a *fimA* or *fimP* gene encoding the major pilin protein, (ii) a *fimB* or *fimQ* gene that specifies an ancillary or minor pilin, and that commonly acts as an adhesin at the tip of the pilus (NB. some clusters encode more than one such minor pilin), and (iii) a gene encoding the pilus-specific sortase (14). The deduced amino acid sequences of *fimA/P* and *fimB/Q* contain particular consensus motifs and domains characteristic of a pilin primary structure, including a Sec-dependent secretion signal, the sortase recognition site (Cell Wall Sorting Signal motif), the pilin-like motif (TVXXK) and the E box (14, 15).

Bifidobacteria have also been shown to encode Type-IV pili, which are similar to the tight adherence or Tad pili (16-18), and which, since

they are essential for gut colonisation, are believed to mediate adhesion to host's surfaces (19, 20). The role of pili in modulating adhesion to human gut mucosa, while at the same time having an impact on host-microbe dialogue, has only been studied for a small number of gut-associated bifidobacterial species (19, 21). Furthermore, these extracellular appendages are also believed to mediate aggregation/interaction events involving other members of the gut microbiota (22). The accumulation of genomic data for this bacterial genus (23-25) justifies a re-evaluation of the number, diversity, and distribution of, and the role(s) elicited by the sortase-dependent (SD)-fimbriome of the genus *Bifidobacterium*. Here, we analyze the features of the bifidobacterial SD-fimbriome, (i.e., the collection of predicted SD pili of the genus *Bifidobacterium*) and evaluate how these extracellular structures are crucial in the interaction with glycans that are abundant in the ecological niche of bifidobacteria. Furthermore, comparison of the here identified SD-fimbriome with microbiome datasets allowed us to reconstruct the contribution of bifidobacterial sortase-dependent pili to the overall predicted microbial interactome of the human gut.

Results and discussion

Identification of bifidobacterial pilus loci. We surveyed 158 genome sequences representing all 48 currently recognized (sub)species of the *Bifidobacterium* genus for loci encompassing genes predicted to encode SD pilus structures,

henceforth designated as pilus-encoding loci (PEL). This bifidobacterial genomic datasets included all the genomes belonging to the *Bifidobacterium* genus that were available on NCBI at the time of writing. This analysis led to the identification of 294 PEL that collectively represent the SD-fimbriome of the genus *Bifidobacterium*. The genome of *Bifidobacterium dentium* LMG11045 contains seven PEL, which is the largest number of PEL so far identified in a given genome of a member of the genus *Bifidobacterium*. In contrast, the chromosomes of *Bifidobacterium animalis* subsp. *lactis*, *Bifidobacterium asteroides*, *Bifidobacterium bombi*, *Bifidobacterium callitrichos*, *Bifidobacterium kashiwanohense*, *Bifidobacterium mongoliense*, *Bifidobacterium psychraerophilum*, *Bifidobacterium subtilis*, *Bifidobacterium thermacidophilum* subsp. *thermacidophilum* and *Bifidobacterium longum* subsp. *suis* were shown to contain just a single PEL (see Table S1 in the supplemental material), while 10 strains belonging to *B. actinocoloniforme*, *B. longum* subsp. *longum* and *B. longum* subsp. *infantis* species do not contain any PEL in their genomes (Table S1). Notably, we identified a commonly occurring gene constellation, observed for 269 PEL, that consists of two genes specifying predicted pilus subunits and an associated sortase-encoding gene (Fig. 1). In addition, we identified 19 pilus loci that consist of just a single pilus subunit-encoding gene flanked by a sortase-encoding gene, and six pilus loci consisting of two pilin-

encoding genes without an associated sortase-encoding gene. Within the predicted SD-fimbriome of the genus *Bifidobacterium* (including typical and atypical pilus loci), we identified 156 PEL that are shared by different bifidobacterial (sub)species and that therefore constitute the clusters of orthologous (sortase-dependent) pilus-encoding genes (COPGs) (described below). PEL belonging to a given COPG were defined as displaying 50 % identity on sortase protein sequences. We used sortase-encoding genes because these are highly conserved compared to the pilin subunit-encoding genes (described below). Notably, cross-alignment of all *FimA/P* and *FimB/Q* genes included in the 15 COPG revealed, in addition to high intra-COPG similarity and identity, high homology between pili subunits of different COPG, which is indicative of horizontal gene transfer (HGT) events (see Data set S1 in the supplemental material) (described below). In addition, the bifidobacterial SD-fimbriome includes 138 PEL that are uniquely identified in the chromosome of a single bifidobacterial (sub)species, and thus represent an accessory PEL arsenal. Of these 138 unique pilus loci, 89 are species-specific clusters (i.e., species-specific PEL), and for the remaining 49, five are unique for a specific strain representing strain-specific PEL, while for 44 it was not possible to make such a classification since only one genome sequence of the particular species was available. Altogether, our findings highlight the existence of a very substantial arsenal of SD

pilus-encoding genes within the pan-genome of bifidobacteria. Furthermore, comparative analysis suggests that, in addition to species-specific genetic strategies, there are highly variable, strain-dependent mechanisms employed by these bacteria to interact with their environment.

Classification of the pilus-encoding loci in the *Bifidobacterium* SD-fimbriome. Based on the finding that sortase proteins of Gram-positive bacteria are cell wall-anchored transpeptidases, used by bacteria to assemble pilus subunits, a phylogenetic tree was built by comparing the deduced PEL-associated sortase proteins from all available *Bifidobacterium* members. Notable, sortase homologs in the bifidobacterial genomes outside PEL were not included in our analyses.

The resulting phylogenetic tree was based on 287 sortases encoded by genes present in the 294 identified PEL (see above), identified not only based on the genomes of the reference strains for each (sub)species of the genus *Bifidobacterium*, but also in all analyzed 158 bifidobacterial genomes (Fig. 1). In seven PEL no sortase-encoding gene was identified, suggesting the involvement of a sortase that is encoded by a gene outside the identified PEL, similar to what was previously described for *Streptococcus pyogenes* (26). Such analyses revealed the presence of 15 main inter-species groups of PEL, named COPG, that were shown to possess >50 % sequence identity, and located in the same genomic region (Fig. S1). Furthermore, eight COPG were identified that

are present only in multiple strains of the same species. These clusters were found in *Bifidobacterium adolescentis* (and named *ado1*, *ado2* and *ado3*), *Bifidobacterium bifidum* (designated *bif1* and *bif2*), *Bifidobacterium breve* (*bre1*), *Bifidobacterium longum* (*lon1*) and *Bifidobacterium magnum* (*mag1*).

In a small number of PEL, i.e., 39, we identified a gene encoding a third pilus subunit predicted to encode an adhesin protein displaying a similarity above 98 % with homologous proteins involved in bacterial attachment to host tissue surfaces (12, 27). Interestingly, a phylogenetic tree based on these 39 adhesin proteins described above, showed the same clusters as those retrieved by sortase analyses (Fig. S2).

***In silico* characterization of the pilus-encoding loci of the *Bifidobacterium* SD-fimbriome.** Major and minor subunits of *Bifidobacterium* SD-fimbriome were identified based on similarity (at amino acid level) (28, 29) as well as by detailed *in silico* analysis of domains typically found in the pilin subunits. In this context, *in silico* identification of domains and secondary structures of the deduced protein products of the predicted *fimA/P* and *fimB/Q* homologs revealed the typical consensus motifs and domains characteristic of a pilin primary structure, including a Sec-dependent secretion signal, the sortase recognition site (CWSS motif), the pilin-like motif (TVXXXK) and the E box (14) (see Table S1 and S2 in supplemental material). Interestingly, identification of the latter motifs of bifidobacterial *fimA/P* confirms

their widespread distribution in the major subunits of sortase-dependent pili in other common human gut commensals such as members of the genera *Bacteroides*, *Blautia*, *Collinsella*, *Enterococcus*, *Faecalibacterium*, and *Lactobacillus* (Table S2).

Notably, *in silico* inspection of the predicted secondary structures and domains of the predicted FimA/P major subunits confirmed a similar topology associated with a backbone functionality. Analysis of putative FimB/Q minor subunits encoded by bifidobacteria revealed the presence of domains that seem to drive the adhesion of pili to different substrates (11). These include a putative collagen-binding adhesin (CNA) domain identified in the minor subunit of 11 COPGs, which was previously demonstrated to be a primary determinant of the ability to bind collagen in *Staphylococcus aureus* (30, 31). Notably the product of *fimB/Q* small subunits of COPGs 1, 3 and 5 appear to be evolutionary linked to the minor pili subunit-encoding gene *rrgA* of *Streptococcus pneumoniae*, as indicated by a similarity of >40%. This domain has been shown to be pivotal for immune evasion by *S. pneumoniae* as well as for mediating adhesion to the extracellular matrix proteins of the host (32). Altogether, these findings suggest a common functional role exploited by the FimB/Q small subunits of COPG 1, 3 and 5. Moreover, in *Bifidobacterium saeculare* this CNA domain is also present in the FimB/Q-like member of COPG 2. In contrast, the product of the *fimB/Q* gene of *B. pullorum*

classified as COPG 2 encompasses B repeats units of the collagen-binding surface protein (CNA) of *S. aureus*. This domain is involved in “stalk” functions improving the projection of the binding domain (collagen-binding A region) from the bacterial surface (33).

Interestingly, the small subunit of five COPGs also contain a Carbohydrate Binding Module (CBM), described to sustain the interaction of pili with carbohydrates (34). Notably, these FimB/Q paralogs also encompass a CNA domain, suggesting that these pili promote adhesion of bifidobacterial cells to host tissue, such as the gut mucosa, as well as adhesion to carbohydrates present in the gut, including those present in mucin or introduced with the diet (35, 36). More specifically, in addition to a CNA domain homologous to the product of *rrgC* gene of *S. pneumoniae* (37), the products of the *fimB/Q* homologs that make up COPGs 8, 9, 10, 11 and 15 also include domains predicted to be involved in sugar binding. These encompass homologs of domains 4 and 5 of the S-layer sugar-binding protein SbsC of *Geobacillus stearothermophilus* (38) as well as, except for COPG 9, homologs of the CBM 25 of *Bacillus halodurans*, responsible for starch binding (39). Therefore, analysis of the 15 identified COPGs revealed that a small number only contains a CNA domain, while others possess both a CNA as well as a CBM domain (Fig. 2).

Genetic comparison of COPGs within the genus *Bifidobacterium*. Network analyses of the predicted SD-fimbriome of the genus

Bifidobacterium highlighted that among the identified 15 COPGs, cluster 13 as well as clusters 1 and 8 are placed in the central nodes within the reconstructed bifidobacterial pili-based network (Fig. 3). Other SD-fimbriome clusters that were identified in different bifidobacterial (sub)species included clusters 4 and 12. Notably, these latter clusters are identified in the genomes of bifidobacterial taxa residing in the gut of birds, whereas the largest part of the network consisting of clusters 1, 8 and 13 includes the SD pilus arsenal of bifidobacterial (sub)species present in the human gut, with the exception of *B. moukabalense* and *B. ruminantium* (Fig. 3). Interestingly, the PEL from non-human primates cluster together to form COPG 14, which is not connected with any other COPG (Fig. 3), suggesting a divergent evolution followed by COPG 14 with respect to COPG 13. In contrast, COPGs 6 and 7 were found in the genomes of bifidobacterial species present in the gut of social insects (Fig. 3), indicating that the SD-fimbriome arsenal of bifidobacteria may reflect the ecological niche in which the various bifidobacterial (sub)species are commonly encountered.

Extensive acquisition and loss of COPGs during *Bifidobacterium* evolution. *In silico* analyses of the predicted major and minor pilus subunits, as well as the sortases demonstrated similarities between amino acid sequences of different COPGs (Fig. 2). Remarkably, 30 of the PEL classified in the 15 bifidobacterial COPGs were shown to be flanked by transposon

elements, which suggests acquisition by HGT. Other findings supporting possible HGT transfer of PEL are the deviating G+C% content between the sequence of the major pilus subunit-encoding gene and its corresponding genome (being for example 11 % lower in the case of *B. asteroides* LMG10735 or 8 % higher in the case of *B. catenulatum* LMG11043) (Table S3). A similar situation was found for the gene encoding the minor pilus subunit: the G+C% content of this gene ranged from being 12 % lower (in the case of *B. choerinum* 1516) to being 12 % higher (in case of *B. saguini* DSM23967) than that of the corresponding genome. In addition, the predicted FimA/P and FimB/Q proteins also have a different codon usage bias (Table S3). Furthermore, search for homologs of bifidobacterial *FimA/P* and *FimB/Q* genes in the NCBI nr database revealed that a wide range of Gram positive gut colonizers encode genes with BLASTp e-value < 1e⁻³⁰ such as members of the genera *Fusicatenibacter*, *Ruminococcus*, *Faecalibacterium*, *Blautia*, *Clostridium* and *Streptococcus* (see Fig. S3 and S4 in the supplemental material). The presence of homologs to bifidobacterial *FimA/P* and *FimB/Q* genes in other genera may indicate that the latter may have acted as putative donors of PEL to bifidobacterial genomes. These findings are further corroborated by inspection of the phylogenetic tree based on the SD-fimbriome of bifidobacteria plus the various PEL homologs from other Gram positive microorganisms,

which clearly points out that bifidobacterial FimA/P and FimB/Q proteins do not form a monophyletic group with those of other high G+C Gram positive bacteria (Fig. S3 and Fig. S4). Altogether these phylogenetic inconsistencies, GC-content deviation and distinctive codon usage indicate that (elements of) the bifidobacterial SD-fimbriome, similar to the SD-fimbriome of other Actinobacteria (40), were acquired through HGT events (Table S3).

Transcriptional analysis of bifidobacterial pilus-encoding loci.

In order to investigate if the bifidobacterial PEL are differentially transcribed in response to the environment to which bifidobacterial cells are exposed, the levels of *fimB* homolog-specific mRNA were determined by quantitative real-time PCR (qRT-PCR) assays by *in vitro* experiments. We decided to investigate six bifidobacterial strains, i.e., *B. minimum* LMG11592, *B. choerinum* LMG10510, *B. pseudolongum* subsp. *globosum* GLOB10b, *B. pseudolongum* subsp. *pseudolongum* 1520, and *B. kashiwanohense* DSM21854 and *B. asteroides* LMG10735. These strains encompass a unique PEL in their genome consisting of pilin-encoding genes which encode proteins with either a single CNA or a combined CNA-CBM domain (Fig. 4). The experiments were performed using mRNA samples extracted from cultures of the above mentioned bifidobacterial strains (Fig. 4), which had been exponentially grown in MRS medium containing as the unique carbon source a particular glycan, such as starch or xylan (in the

case of *B. minimum* LMG11592, *B. choerinum* LMG10510, *B. pseudolongum* subsp. *globosum* GLOB10b, *B. pseudolongum* subsp. *pseudolongum* 1520b), maltotriose or xylan (in the case of *B. kashiwanohense* DSM21854), xylose or fructose (in the case of *B. asteroides* LMG10735), found in the ecological niche where these strains are naturally residing (41). The observed transcription level of the putative subunit-encoding pilin genes in response to these growth substrates was shown to be variable between the different bifidobacterial strains and highly variable between different pilus gene clusters harboured by the same organism (Fig. 4).

Identification of pili-like structures on the cell envelope in bifidobacteria.

We decided to investigate the occurrence of pilus-like structures in the genus *Bifidobacterium* by AFM. However, in order to avoid possible misidentification of these extracellular appendages due to the potential multiple presence of different pili-like structures for the same bifidobacterial cell as consequence of the occurrence of different PELs, we decided to assay the cell surface of the six strains possessing a unique PEL in their genomes, which were those strains assayed by qPCR experiments (see above). Prior to AFM visualization, bacterial cells were cultivated in the presence of (partially hydrolyzed) xylan (42), or xylose only for *B. asteroides* LMG10735, as these conditions were shown to induce the expression of pilin subunits (see

above), representing glycans that are expected to be present in their natural ecological niches (43). When we assayed the production of pilus-like structures by bifidobacteria cultivated under these different carbon sources by AFM, we noticed the presence of pilus-like structures decorating the cell surfaces (Fig. 5). Notably, these pilus-like structures occupied different positions of the cell surface of the investigated strains, being located at the poles of the cell or uniformly distributed along the cells, while they also varied in pilus density and length (Fig. 5).

It is known that bifidobacteria possess other pilus-like encoding locus, i.e., the type IV pili, which is highly conserved in the bifidobacterial genomes (4). In order to verify if the pili observed by AFM analyses were those encoded by the PEL above described or by the type IV pili-encoded loci (Tad pili), we assayed the transcription of *tadA* gene, which encodes an ATPase that is believed to be crucial for Tad pilus assembly (18), by qRT-PCR. Interestingly, transcription of the *tadA* gene was not detectable under conditions that allowed transcription of the *fimB* gene. This finding is consistent with what has previously been described for *B. breve* UCC2003 whose *tad* locus is only transcribed when this bacterium is present in the (murine) gut (19).

Binding to human receptor fibronectin. In order to evaluate if the *in silico* identified CNA motifs present in the minor pilin subunit of bifidobacteria are involved in adhesion to Extra Cellular Matrix (ECM) proteins, we performed

fibronectin binding assays, following the protocol described in the materials and methods section, involving two strains, i.e., *B. minimum* LMG11592 and *B. pseudolongum* subsp. *pseudolongum* 1520b, each of which possess a unique PEL in their genome. We used fibronectin since this ECM protein may act as a possible epithelial cell receptor involved in the recognition of bacteria (44). As a negative control, we employed bovine serum albumin (BSA). Notably, both *B. pseudolongum* subsp. *pseudolongum* 1520- and *B. minimum* LMG11592-cells were shown to adhere to fibronectin compared to the BSA control, consisting of 8.15E4 cfu/ml (corresponding to 0.37 % recovery adhesion) and 5.18E6 cfu/ml (corresponding to 7.32 % recovery adhesion), respectively (Fig. 4). Interestingly, our experiments with *B. pseudolongum* subsp. *pseudolongum* 1520 cells showed that adhesion to fibronectin is not influenced by the presence of xylan (7.00E4 cfu/ml, corresponding to 0.23 % recovery adhesion), which may be correlated to the fact that the *fimBQ* homolog of this strain encodes a CNA domain only (and not a CBM domain). In contrast, *B. minimum* LMG11592 cells, whose minor pilin subunit is predicted to possess both CNA and CBM domains, showed a significant reduction (5.11E5 cfu/ml; *p*-value <0.04, correspond to 1.29 % recovery adhesion) in adhesion efficacy to fibronectin when the ECM binding trials were performed in the presence of xylan (Fig. 4). These findings suggest that both carbohydrate residues and

glycoproteins are modulating fimbrial binding of *B. minimum* LMG11592 cells.

Bifidobacterial SD-fimbriome across the infant gut microbiota. The functional contribution of bifidobacteria in terms of metabolism and resistome has recently been investigated within several gut microbiomes (42, 45). However, their contribution to the SD-fimbriome of the mammalian gut has not been explored. In order to investigate the presence of bifidobacterial PEL and COPG within the gut of humans, we employed metagenomic datasets from three different projects, encompassing healthy and preterm infants (46), children who had received an intensive antibiotic treatment (47) and healthy adults (<https://www.broadinstitute.org/hmp/human-microbiome-project>).

The presence of bifidobacterial sequences within these samples has previously been evaluated, highlighting a higher abundance of bifidobacterial genes in healthy children compared to children treated with antibiotics and preterm infants (45, 48). Our analyses demonstrated the presence of both bifidobacterial COPG and PEL across these different metagenomic datasets (Fig. 6). Analyzing the overall bifidobacterial read distribution, we found that the species with the higher occurrence within the metagenomic samples reflects the bifidobacterial species of the PEL and COPG reads identified, e.g., *B. breve*, *B. bifidum*, *B. longum* spp., *B. adolescentis* and *B. catenulatum* (Fig. 6).

Interestingly, the COPG that were identified in the examined microbiomes are represented by COPG 1, COPG 8 and COPG 13, while the unique PEL are scattered between the samples with the only exception being *magI* which was not identified in any of the metagenomic samples, confirming its non-human ecological origin (Fig. 6). These data highlight the presence of COPG and PEL related with bifidobacterial species that typically colonize the gut of humans, such as *B. breve*, *B. bifidum*, *B. longum* spp., *B. adolescentis* and *B. catenulatum*. Furthermore, the COPG 8 that corresponds to the *B. adolescentis* group (49) has been identified in non-infant samples suggesting a role in the colonization of a more complex ecological niche (i.e., adult and children that had received antibiotic treatment), probably due to a more varied diet. In order to evaluate if the identified PEL and COPG were expressed within the gut of infants, we employed the metatranscriptomic dataset of healthy infants retrieved from the project above mentioned (46). Selecting the PEL and COPG that were identified within healthy infants, we found a positive correlation between the presence of reads corresponding to certain PEL in metagenomic and metatranscriptomic data sets (Fig. 6). Thus, the genes encoding sortase-dependent pili of bifidobacteria were shown to be expressed when these microorganisms reside within the gut of humans. These results suggest that SD pilus-like structures of bifidobacteria are involved in the interaction with the host and

perhaps also with other members of the microbiota.

Conclusions

Bifidobacteria have been shown to be dominant microorganisms of the (human) infant gut microbiota (1, 50-54), where they are considered to play a pivotal role in promoting gut health of infants through the priming of the immune system as well as acting on the physiology of the gut. Adhesion of bifidobacterial cells to the mucosa of the large intestine is considered a hallmark for the persistence and colonization of these bacteria in the human gut. SD pili produced by the infant gut isolate *B. bifidum* PRL2010 are involved in the adhesion to human cell lines through extracellular matrix proteins as well as in the aggregation with other members of the human gut microbiota (21, 22). However, current knowledge on the SD-fimbriome of the genus *Bifidobacterium* is rather limited (14, 21, 22). In this study, we provide a detailed genetic survey of SD-fimbriome in the genus *Bifidobacterium*, which shows that these structures are widely distributed in this group of microorganisms, suggesting an important role in facilitating environmental interactions. However, in contrast to other pilus types, like type-IV fimbriae, which are much more conserved and genetically homogeneous among members of the *Bifidobacterium* genus, the SD pili display substantial genetic diversity both in sequence and number of PEL per strain. Notably, many of these extracellular structures

are predicted to be acquired by HGT and show a patch-wise genetic organization. Such genetic features of the reconstructed SD-fimbriome of the genus *Bifidobacterium* suggest variable and exchangeable interaction capabilities with different environments. *In silico* analyses coupled with transcriptome analyses and ECM binding experiments revealed how the different members of the bifidobacterial SD-fimbriome may interact with ECM proteins, such as fibronectin as well as with glycans or specifically only one of these substrates. These findings suggest an intriguing mechanism of genetic adaptation of bifidobacteria to the gut of mammals, birds and insects, allowing their successful colonization and persistence in these ecological niches. More specifically, our data are in line with two possible different ecological activities for those bifidobacteria that colonize the mammalian gut. In this context, those bifidobacteria that due to their SD-fimbriome adhere to diet-derived glycans (e.g., starch, xylan, pectin) and thus exhibit a luminal behavior, which may not support a long persistence in the gut. In contrast, other bifidobacterial species possess a SD-fimbriome that allows adherence to the mucosa layer through the ECM proteins are consequently long-term colonization in the gut. Nevertheless, such different adhesive roles played by the bifidobacterial SD-fimbriome might be further influenced by the SD-fimbriome or the extracellular structures produced by other members of the gut microbiota. Nevertheless,

since many bifidobacteria are practically genetically inaccessible, with just a few exceptions (55, 56), we are currently not in a position to test our predictions under *in vivo* circumstances. Future gene inactivation PEL experiments, when such procedures will become available, will therefore be important to confirm our *in silico* results.

Materials and methods

Identification of Pilus-encoding Loci (PEL) and Clusters of Orthologous Pilus Genes (COPG). SD pilus-encoding loci (PEL), which represent type I and type II pili (12), were identified through manual inspection based on homology analyses for all 140 bifidobacterial genomes (publicly) available at the time of this study using a custom database encompassing all known sortase-dependent pili genes reported in the NCBI RefSeq database. Further 18 novel bifidobacterial strains were selected for PEL inspection, resulting in 81 genes deposited at GenBank (accession numbers from MF043305 to MF043385). Clusters of Orthologous Pilus Genes (COPG) were identified through pan-genome analysis performed using the PGAP pipeline (57) as outlined below. The Open Reading Frame (ORF) contents from all PEL identified in this study were organized in functional clusters using the GF (Gene Family) method involving comparison of each deduced protein to all other proteins using BLAST analysis (cutoff E-value of 1×10^{-5} and 50 % identity across at least 50 % of both protein

sequences), followed by clustering into protein families using MCL (graph-theory-based Markov clustering algorithm) (58). Data regarding protein family distribution among PEL was implied for manual reconstruction of the COPG. Information on distribution of individual COPG in bifidobacterial taxa was used to produce a network representation using Gephi software (59). Phylogenetic trees of the minor and major pilus subunits were constructed using ClustalX for sequence alignment (60) and FigTree for visualization.

Growth conditions. *Bifidobacterium* cells were cultivated in an anaerobic atmosphere (2.99 % H₂, 17.01 % CO₂, and 80 % N₂) in a chamber (Concept 400; Ruskin) at 37°C for 24 h in de Man-Rogosa-Sharpe (MRS; Scharlau Chemie, Barcelona, Spain) medium, supplemented with 0.05 % (wt/vol) L-cysteine hydrochloride. *Bifidobacterium* cells were inoculated in 10 ml of MRS without carbohydrates (Scharlau Chemie, Barcelona, Spain) supplemented with 0.5 % of starch, or maltotriose, or xylan, or xylose or fructose (Sigma Aldrich) as unique carbon sources. Carbohydrates were purchased from Sigma (Milan, Italy) or Carbosynth (Berkshire, UK). Carbohydrates were dissolved in water and then sterilized by filtration using 0.2 micron filter size and then added to autoclaved MRS with the exception of xylan (Sigma, Aldrich) which was autoclaved with MRS. Cell suspensions were mixed and incubated at 37°C for 24 h under anaerobic conditions.

RNA isolation. Total RNA was isolated using a previously described method (61). Briefly, cell pellets/tissue materials were resuspended in 1 ml of QUIAZOL (Qiagen, UK) and placed in a tube containing 0.8 g of glass beads (diameter, 106 μm ; Sigma). Cells were lysed by shaking the mix on a BioSpec homogenizer at 4°C for 2 min (maximum setting). The mixture was then centrifuged at 12,000 rpm for 15 min, and the upper phase containing the RNA-containing sample was recovered. The RNA sample was further purified by phenol extraction and ethanol precipitation according to an established method (62). The quality of the RNA was checked by analysing the integrity of rRNA molecules by Tape Station analysis (Agilent).

Reverse transcription and qRT-PCR analyses. Reverse transcription to cDNA was performed with the iScript™ Select cDNA Synthesis Kit (Bio-Rad Laboratories, Hercules, CA) using the following thermal cycle regime: 5 min at 25°C, 30 min at 42°C, and 5 min at 85°C. The mRNA expression levels of cytokines were analyzed with SYBR Green technology in RT-qPCR using SsoFast EvaGreen Supermix (Bio-Rad Italia, Segrate, Italy) on a Bio-Rad CFX96 system according to the manufacturer's instructions (Bio-Rad Italia, Segrate, Italy). The primers used are indicated in Supplementary Table S4. Quantitative PCR for housekeeping genes was carried out according to the following amplification cycle protocol: initial hold at 95°C for 2 minutes and then 45 cycles at 95°C for 5 s and 60°C for 30 s. The same PCR cycle

conditions were used for all qRT-PCR experiments performed, though adjusting the annealing temperature according to the employed primers. Gene expression was normalized to the housekeeping gene coding for the 16S rRNA as described previously (63, 64). The amount of template cDNA used for each sample was 12.5 ng.

Sample preparation and Atomic Force Microscopy (AFM) imaging. Bacteria from an overnight culture were harvested by centrifugation washed with PBS by resuspension and collection by centrifugation (4,000 rpm). The washed pellet was then resuspended in 200 μl of PBS and kept on ice until AFM imaging. Mica was then rinsed with milliQ water (Millipore) and dried with nitrogen. After this, 20 μl of bacterial suspension was deposited onto mica for 10 minutes. The mica disk was then rinsed with milliQ water and dried under a weak gas flow of nitrogen. Quality of the sample and density of surface-bound bacteria were verified with an optical microscope.

AFM imaging was performed on dried samples with a XE-100 (Park System) in tapping mode. Commercial diving board silicon cantilevers (MikroMasch) were used. Best image quality was obtained with amplitude 1V and low scan rate (0.5-0.7 Hz). Filamentous structures at the periphery of bacteria were visible in images of 256 x 256 pixels, representing a scan size of 10 μm or less. While imaging both height and amplitude signals were collected. Height images were flattened using XEI software.

Quantification of bacterial binding to Extra Cellular Matrix (ECM) proteins. 96-MicroWell™ plates (Maxisorp Nunc, Roskilde, Denmark) were coated with a solution of 500 µg/ml of fibronectin (Sigma) in 100 µl phosphate-buffered saline (PBS) pH 7.4 (i.e., 137 mM NaCl, 2.7 mM KCl, 10 mM Na₂HPO₄, 2 mM KH₂PO₄). Unbound protein was removed by washing the plates twice with PBS, followed by blocking the wells by rinsing with 1 % bovine serum albumin (BSA) in PBS for 30 minutes at 37°C. The blocking buffer was removed, and the wells were washed twice prior to the addition of bacterial cells in a 100-µl final volume. Incubation with bacteria was performed for 1 hour at 37°C. The wells were washed with PBS and the bacterial cells that adhered to the wells were collected by scraping them into PBS with 0.5 % (vol/vol) Triton X-100; serial dilutions were plated onto MRS or GM17 agar plates. The number of adherent bacteria (cfu/ml) was determined by counting the resulting colonies in duplicate. Adherence data are also expressed as the percentage of adherent bacteria recovered from triplicate wells and the means were gathered from two independent experiments. Statistical differences are expressed as the *p*-value determined by a paired Student's *t* test.

Identification of bifidobacterial SD-fimbriome components in gut metagenomic and metatranscriptomic datasets. All identified bifidobacterial PEL were aligned with microbiota sequenced reads belonging to three

projects previously deposited at the NCBI Sequence Read Archive (SRA). These metagenomic data were obtained from shotgun sequencing of fecal samples of healthy and preterm infants from the Metagenome from Infant Gut Samples project (BioProject ID: PRJNA63661), children administered with antibiotics from the Child Gut Microbiome under Antibiotics project (BioProject ID: PRJEB11685) and healthy adults from the Human Microbiome Project (BioProject ID: PRJNA48479). Metagenomic data sets were filtered using the `fastq-mcf` script (<https://expressionanalysis.github.io/ea-utils/>) (minimum mean quality score, 20; window size, 5; quality threshold, 25; and minimum length, 80) to exclusively retrieve high-quality reads. The resulting reads were aligned against the human genome using the Burrows-Wheeler Aligner program (65) (BWAMEM algorithm with trigger reseeded, 1.5; minimum seed length, 19; matching score, 1; mismatch penalty, 4; gap open penalty, 6; and gap extension penalty, 1) and further processed with the SAMtools software package (66) to remove human reads. The mapping against the bifidobacterial PEL and COPG was performed using Bowtie 2 (67) through multiple-hit mapping and “very-sensitive” policy. The mapping was performed using a minimum score threshold function (`-score-min C,-13,0`) in order to limit reads with at least 98 % full-length identity. The software employed to calculate read counts corresponding to bifidobacterial PEL

and COPG was HTSeq (68) (running in union mode). The percentages of bifidobacterial PEL and COPG for each sample were based on the total amount of filtered reads.

Acknowledgments

This work was funded by the EU Joint Programming Initiative – A Healthy Diet for a Healthy Life (JPI HDHL, <http://www.healthydietforhealthylife.eu/>) and the MIUR to MV. We thank GenProbio srl for financial support of the Laboratory of Probiogenomics. DvS is a member of The APC Microbiome Institute funded by Science Foundation Ireland (SFI), through the Irish Government's National Development Plan (Grant number SFI/12/RC/2273). Part of this research is conducted using the High Performance Computing (HPC) facility of the University of Parma. The authors declare that they have no competing interests.

References

1. Turrone F, Peano C, Pass DA, Foroni E, Severgnini M, Claesson MJ, Kerr C, Hourihane J, Murray D, Fuligni F, Gueimonde M, Margolles A, De Bellis G, O'Toole PW, van Sinderen D, Marchesi JR, Ventura M. 2012. Diversity of Bifidobacteria within the Infant Gut Microbiota. *PLoS One* 7:e36957.
2. Turrone F, van Sinderen D, Ventura M. 2009. Bifidobacteria: from ecology to genomics. *Front Biosci* 14:4673-84.
3. O'Callaghan A, van Sinderen D. 2016. Bifidobacteria and Their Role as Members of the Human Gut Microbiota. *Front Microbiol* 7:925.
4. Ventura M, Turrone F, Motherway MO, MacSharry J, van Sinderen D. 2012. Host-microbe interactions that facilitate gut colonization by commensal bifidobacteria. *Trends Microbiol* 20:467-76.
5. Turrone F, Ventura M, Butto LF, Duranti S, O'Toole PW, Motherway MO, van Sinderen D. 2014. Molecular dialogue between the human gut microbiota and the host: a *Lactobacillus* and *Bifidobacterium* perspective. *Cell Mol Life Sci* 71:183-203.
6. Turrone F, Milani C, Duranti S, Mancabelli L, Mangifesta M, Viappiani A, Lugli GA, Ferrario C, Gioiosa L, Ferrarini A, Li J, Palanza P, Delledonne M, van Sinderen D, Ventura M. 2016. Deciphering bifidobacterial-mediated metabolic interactions and their impact on gut microbiota by a multi-omics approach. *ISME J* doi:10.1038/ismej.2015.236.
7. Arbolea S, Watkins C, Stanton C, Ross RP. 2016. Gut Bifidobacteria Populations in Human Health and Aging. *Front Microbiol* 7:1204.
8. Kline KA, Falke S, Dahlberg S, Normark S, Henriques-Normark B. 2009. Bacterial adhesins in host-microbe interactions. *Cell Host Microbe* 5:580-92.
9. Kline KA, Dodson KW, Caparon MG, Hultgren SJ. 2010. A tale of two pili: assembly and function of pili in bacteria. *Trends Microbiol* 18:224-32.
10. Proft T, Baker EN. 2009. Pili in Gram-negative and Gram-positive bacteria - structure, assembly and their role in disease. *Cell Mol Life Sci* 66:613-35.
11. Krishnan V. 2015. Pilins in gram-positive bacteria: A structural perspective. *IUBMB Life* 67:533-43.
12. Danne C, Dramsi S. 2012. Pili of gram-positive bacteria: roles in host colonization. *Res Microbiol* 163:645-58.
13. Spirig T, Weiner EM, Clubb RT. 2011. Sortase enzymes in Gram-positive bacteria. *Mol Microbiol* 82:1044-59.
14. Foroni E, Serafini F, Amidani D, Turrone F, He F, Bottacini F, O'Connell Motherway M, Viappiani A, Zhang Z, Rivetti C, van Sinderen D, Ventura M. 2011. Genetic analysis and morphological identification of pilus-like structures in members of the genus *Bifidobacterium*. *Microb Cell Fact* 10 Suppl 1:S16.
15. Scott JR, Zahner D. 2006. Pili with strong attachments: Gram-positive bacteria do it differently. *Mol Microbiol* 62:320-30.
16. Danelishvili L, Yamazaki Y, Selker J, Bermudez LE. 2010. Secreted *Mycobacterium tuberculosis* Rv3654c and Rv3655c proteins participate in the suppression of macrophage apoptosis. *PLoS One* 5:e10474.
17. Schreiner HC, Sinatra K, Kaplan JB, Furgang D, Kachlany SC, Planet PJ, Perez BA, Figurski DH, Fine DH. 2003. Tight-adherence genes of *Actinobacillus actinomycetemcomitans* are required for virulence in a rat model. *Proc Natl Acad Sci U S A* 100:7295-300.
18. Tomich M, Planet PJ, Figurski DH. 2007. The *tad* locus: postcards from the widespread colonization island. *Nat Rev Microbiol* 5:363-75.
19. O'Connell Motherway M, Zomer A, Leahy SC, Reunanen J, Bottacini F, Claesson MJ, O'Brien F, Flynn K, Casey PG, Munoz JA, Kearney B, Houston AM, O'Mahony C, Higgins DG, Shanahan F, Palva A, de Vos WM, Fitzgerald GF, Ventura M, O'Toole PW, van Sinderen D. 2011.

- Functional genome analysis of *Bifidobacterium breve* UCC2003 reveals type IVb tight adherence (Tad) pili as an essential and conserved host-colonization factor. *Proc Natl Acad Sci U S A* 108:11217-22.
20. Kachlany SC, Planet PJ, DeSalle R, Fine DH, Figurski DH. 2001. Genes for tight adherence of *Actinobacillus actinomycetemcomitans*: from plaque to plague to pond scum. *Trends in Microbiology* 9:429-437.
 21. Turrone F, Serafini F, Foroni E, Duranti S, O'Connell Motherway M, Taverniti V, Mangifesta M, Milani C, Viappiani A, Roversi T, Sanchez B, Santoni A, Gioiosa L, Ferrarini A, Delledonne M, Margolles A, Piazza L, Palanza P, Bolchi A, Guglielmetti S, van Sinderen D, Ventura M. 2013. Role of sortase-dependent pili of *Bifidobacterium bifidum* PRL2010 in modulating bacterium-host interactions. *Proc Natl Acad Sci U S A* 110:11151-6.
 22. Turrone F, Serafini F, Mangifesta M, Arioli S, Mora D, van Sinderen D, Ventura M. 2014. Expression of sortase-dependent pili of *Bifidobacterium bifidum* PRL2010 in response to environmental gut conditions. *FEMS Microbiol Lett* 357:23-33.
 23. Milani C, Lugli GA, Duranti S, Turrone F, Bottacini F, Mangifesta M, Sanchez B, Viappiani A, Mancabelli L, Tamini B, Delcenserie V, Barrangou R, Margolles A, van Sinderen D, Ventura M. 2014. Genomic encyclopedia of type strains of the genus *Bifidobacterium*. *Appl Environ Microbiol* 80:6290-302.
 24. Duranti S, Milani C, Lugli GA, Turrone F, Mancabelli L, Sanchez B, Ferrario C, Viappiani A, Mangifesta M, Mancino W, Gueimonde M, Margolles A, van Sinderen D, Ventura M. 2015. Insights from genomes of representatives of the human gut commensal *Bifidobacterium bifidum*. *Environ Microbiol* doi:10.1111/1462-2920.12743.
 25. Duranti S, Milani C, Lugli GA, Mancabelli L, Turrone F, Ferrario C, Mangifesta M, Viappiani A, Sanchez B, Margolles A, van Sinderen D, Ventura M. 2016. Evaluation of genetic diversity among strains of the human gut commensal *Bifidobacterium adolescentis*. *Sci Rep* 6:23971.
 26. Barnett TC, Scott JR. 2002. Differential recognition of surface proteins in *Streptococcus pyogenes* by two sortase gene homologs. *J Bacteriol* 184:2181-91.
 27. Klemm P, Schembri MA. 2000. Bacterial adhesins: function and structure. *Int J Med Microbiol* 290:27-35.
 28. Ton-That H, Schneewind O. 2003. Assembly of pili on the surface of *Corynebacterium diphtheriae*. *Mol Microbiol* 50:1429-38.
 29. Mishra A, Das A, Cisar JO, Ton-That H. 2007. Sortase-catalyzed assembly of distinct heteromeric fimbriae in *Actinomyces naeslundii*. *J Bacteriol* 189:3156-65.
 30. Gillaspay AF, Lee CY, Sau S, Cheung AL, Smeltzer MS. 1998. Factors affecting the collagen binding capacity of *Staphylococcus aureus*. *Infect Immun* 66:3170-8.
 31. Snodgrass JL, Mohamed N, Ross JM, Sau S, Lee CY, Smeltzer MS. 1999. Functional analysis of the *Staphylococcus aureus* collagen adhesin B domain. *Infect Immun* 67:3952-9.
 32. Liddington R. 2010. Catching pneumonia. *Structure* 18:6-8.
 33. Deivanayagam CC, Rich RL, Carson M, Owens RT, Danthuluri S, Bice T, Hook M, Narayana SV. 2000. Novel fold and assembly of the repetitive B region of the *Staphylococcus aureus* collagen-binding surface protein. *Structure* 8:67-78.
 34. Boraston AB, Bolam DN, Gilbert HJ, Davies GJ. 2004. Carbohydrate-binding modules: fine-tuning polysaccharide recognition. *Biochem J* 382:769-81.
 35. Koropatkin NM, Cameron EA, Martens EC. 2012. How glycan metabolism shapes the human gut microbiota. *Nat Rev Microbiol* 10:323-35.
 36. Podolsky DK. 1985. Oligosaccharide structures of human colonic mucin. *J Biol Chem* 260:8262-71.
 37. Shaik MM, Maccagni A, Tourcier G, Di Guilmi AM, Dessen A. 2014. Structural basis of pilus anchoring by the ancillary pilin RrgC of *Streptococcus pneumoniae*. *J Biol Chem* 289:16988-97.
 38. Pavkov T, Egelseer EM, Tesarz M, Svergun DI, Sleytr UB, Keller W. 2008. The structure and binding behavior of the bacterial cell surface layer protein SbsC. *Structure* 16:1226-37.

39. Boraston AB, Healey M, Klassen J, Ficko-Blean E, Lammerts van Bueren A, Law V. 2006. A structural and functional analysis of alpha-glucan recognition by family 25 and 26 carbohydrate-binding modules reveals a conserved mode of starch recognition. *J Biol Chem* 281:587-98.
40. Filloux A. 2010. A variety of bacterial pili involved in horizontal gene transfer. *J Bacteriol* 192:3243-5.
41. Bottacini F, Milani C, Turrone F, Sanchez B, Foroni E, Duranti S, Serafini F, Viappiani A, Strati F, Ferrarini A, Delledonne M, Henrissat B, Coutinho P, Fitzgerald GF, Margolles A, van Sinderen D, Ventura M. 2012. Bifidobacterium asteroides PRL2011 genome analysis reveals clues for colonization of the insect gut. *PLoS One* 7:e44229.
42. Milani C, Lugli GA, Duranti S, Turrone F, Mancabelli L, Ferrario C, Mangifesta M, Hevia A, Viappiani A, Scholz M, Arioli S, Sanchez B, Lane J, Ward DV, Hickey R, Mora D, Segata N, Margolles A, van Sinderen D, Ventura M. 2015. Bifidobacteria exhibit social behavior through carbohydrate resource sharing in the gut. *Sci Rep* 5:15782.
43. Milani C, Turrone F, Duranti S, Lugli GA, Mancabelli L, Ferrario C, van Sinderen D, Ventura M. 2016. Genomics of the Genus Bifidobacterium Reveals Species-Specific Adaptation to the Glycan-Rich Gut Environment. *Appl Environ Microbiol* 82:980-91.
44. Dorsey CW, Laarakker MC, Humphries AD, Weening EH, Baumler AJ. 2005. Salmonella enterica serotype Typhimurium MisL is an intestinal colonization factor that binds fibronectin. *Mol Microbiol* 57:196-211.
45. Duranti S, Lugli GA, Mancabelli L, Turrone F, Milani C, Mangifesta M, Ferrario C, Anzalone R, Viappiani A, van Sinderen D, Ventura M. 2017. Prevalence of Antibiotic Resistance Genes among Human Gut-Derived Bifidobacteria. *Appl Environ Microbiol* 83.
46. Morrow AL, Lagomarcino AJ, Schibler KR, Taft DH, Yu Z, Wang B, Altaye M, Wagner M, Gevers D, Ward DV, Kennedy MA, Huttenhower C, Newburg DS. 2013. Early microbial and metabolomic signatures predict later onset of necrotizing enterocolitis in preterm infants. *Microbiome* 1:13.
47. Korpela K, Salonen A, Virta LJ, Kekkonen RA, Forslund K, Bork P, de Vos WM. 2016. Intestinal microbiome is related to lifetime antibiotic use in Finnish pre-school children. *Nat Commun* 7:10410.
48. Lugli GA, Milani C, Turrone F, Tremblay D, Ferrario C, Mancabelli L, Duranti S, Ward DV, Ossiprandi MC, Moineau S, van Sinderen D, Ventura M. 2015. Prophages of the genus Bifidobacterium as modulating agents of the infant gut microbiota. *Environ Microbiol* doi:10.1111/1462-2920.13154.
49. Lugli GA, Milani C, Turrone F, Duranti S, Ferrario C, Viappiani A, Mancabelli L, Mangifesta M, Taminiau B, Delcenserie V, van Sinderen D, Ventura M. 2014. Investigation of the evolutionary development of the genus Bifidobacterium by comparative genomics. *Appl Environ Microbiol* 80:6383-94.
50. Nuriel-Ohayon M, Neuman H, Koren O. 2016. Microbial Changes during Pregnancy, Birth, and Infancy. *Front Microbiol* 7:1031.
51. Martin R, Makino H, Cetinyurek Yavuz A, Ben-Amor K, Roelofs M, Ishikawa E, Kubota H, Swinkels S, Sakai T, Oishi K, Kushiro A, Knol J. 2016. Early-Life Events, Including Mode of Delivery and Type of Feeding, Siblings and Gender, Shape the Developing Gut Microbiota. *PLoS One* 11:e0158498.
52. Yassour M, Vatanen T, Siljander H, Hamalainen AM, Harkonen T, Ryhanen SJ, Franzosa EA, Vlamakis H, Huttenhower C, Gevers D, Lander ES, Knip M, Group DS, Xavier RJ. 2016. Natural history of the infant gut microbiome and impact of antibiotic treatment on bacterial strain diversity and stability. *Sci Transl Med* 8:343ra81.
53. Avershina E, Lundgard K, Sekelja M, Dotterud C, Storro O, Oien T, Johnsen R, Rudi K. 2016. Transition from infant- to adult-like gut microbiota. *Environ Microbiol* 18:2226-36.

54. Roger LC, Costabile A, Holland DT, Hoyles L, McCartney AL. 2010. Examination of faecal *Bifidobacterium* populations in breast- and formula-fed infants during the first 18 months of life. *Microbiology* 156:3329-41.
55. Sakaguchi K, He JL, Tani S, Kano Y, Suzuki T. 2012. A targeted gene knockout method using a newly constructed temperature-sensitive plasmid mediated homologous recombination in *Bifidobacterium longum*. *Applied Microbiology and Biotechnology* 95:499-509.
56. O'Connell KJ, Motherway MO, Liedtke A, Fitzgerald GF, Ross RP, Stanton C, Zomer A, van Sinderen D. 2014. Transcription of Two Adjacent Carbohydrate Utilization Gene Clusters in *Bifidobacterium breve* UCC2003 Is Controlled by LacI- and Repressor Open Reading Frame Kinase (ROK)-Type Regulators. *Applied and Environmental Microbiology* 80:3604-3614.
57. Zhao Y, Wu J, Yang J, Sun S, Xiao J, Yu J. 2012. PGAP: pan-genomes analysis pipeline. *Bioinformatics* 28:416-8.
58. Enright AJ, Van Dongen S, Ouzounis CA. 2002. An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res* 30:1575-84.
59. Jacomy M, Venturini T, Heymann S, Bastian M. 2014. ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software. *PLoS One* 9:e98679.
60. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG. 2007. Clustal W and Clustal X version 2.0. *Bioinformatics* 23:2947-8.
61. Turrone F, Foroni E, Montanini B, Viappiani A, Strati F, Duranti S, Ferrarini A, Delledonne M, van Sinderen D, Ventura M. 2011. Global Genome Transcription Profiling of *Bifidobacterium bifidum* PRL2010 under In Vitro Conditions and Identification of Reference Genes for Quantitative Real-Time PCR. *Appl Environ Microbiol* 77:8578-87.
62. Sambrook (ed). 1989. *Molecular Cloning. A manual laboratory.*
63. Turrone F, Strati F, Foroni E, Serafini F, Duranti S, van Sinderen D, Ventura M. 2012. Analysis of predicted carbohydrate transport systems encoded by *Bifidobacterium bifidum* PRL2010. *Appl Environ Microbiol* 78:5002-12.
64. Turrone F, Ozcan E, Milani C, Mancabelli L, Viappiani A, van Sinderen D, Sela DA, Ventura M. 2015. Glycan cross-feeding activities between bifidobacteria under in vitro conditions. *Front Microbiol* 6:1030.
65. Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25:1754-60.
66. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Genome Project Data Processing S. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25:2078-9.
67. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9:357-9.
68. Anders S, Pyl PT, Huber W. 2015. HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics* 31:166-9.

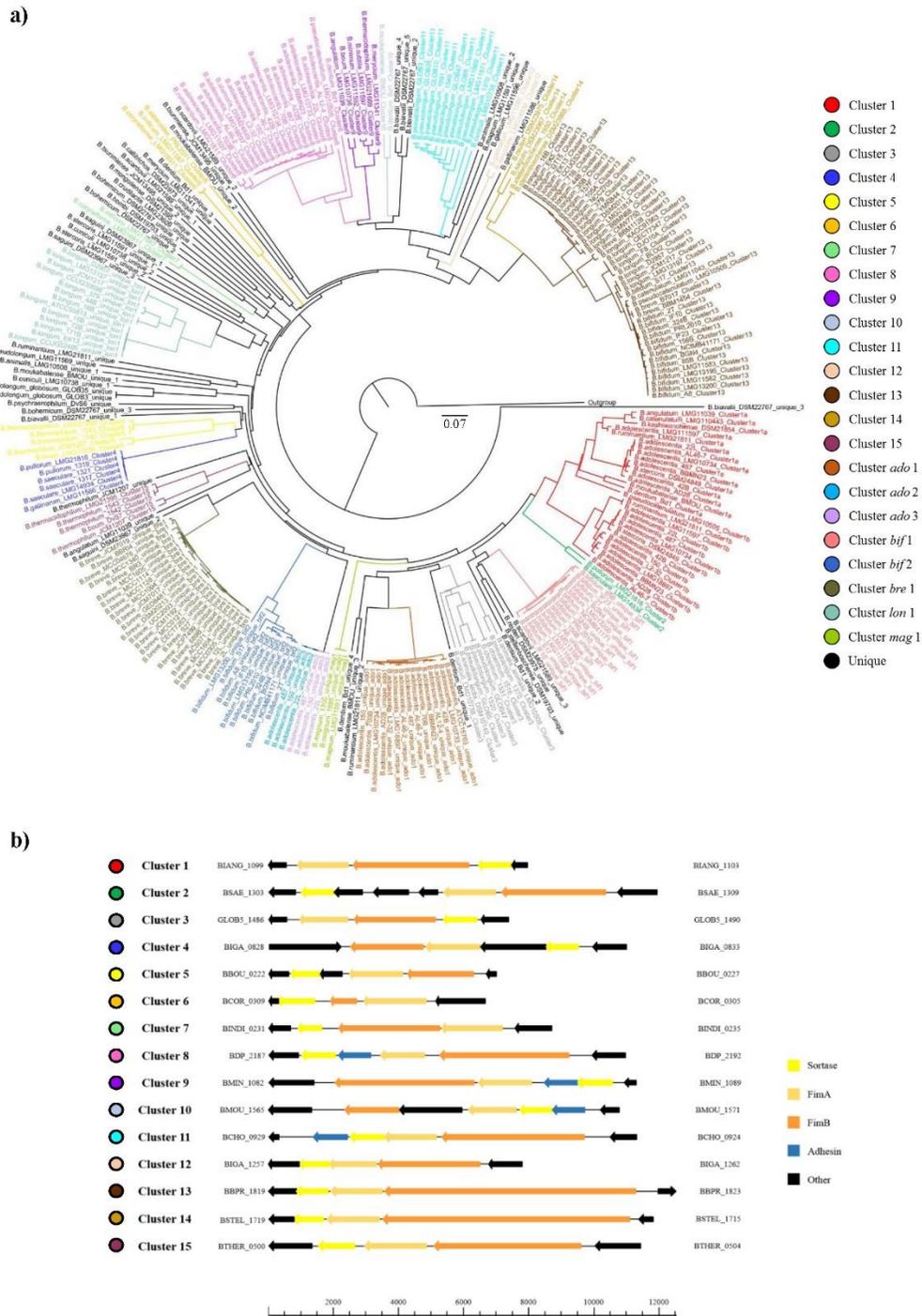


Figure 1. Phylogenetic analyses of SD-fimbriome of bifidobacteria. Panel (a) represents a tree generated from hierarchical clustering based on the identified bifidobacterial sortases. Panel (b) depicts physical maps of a representative strain for the fifteen COPGs. Genes were categorized according to their functions with differently coloured arrows.

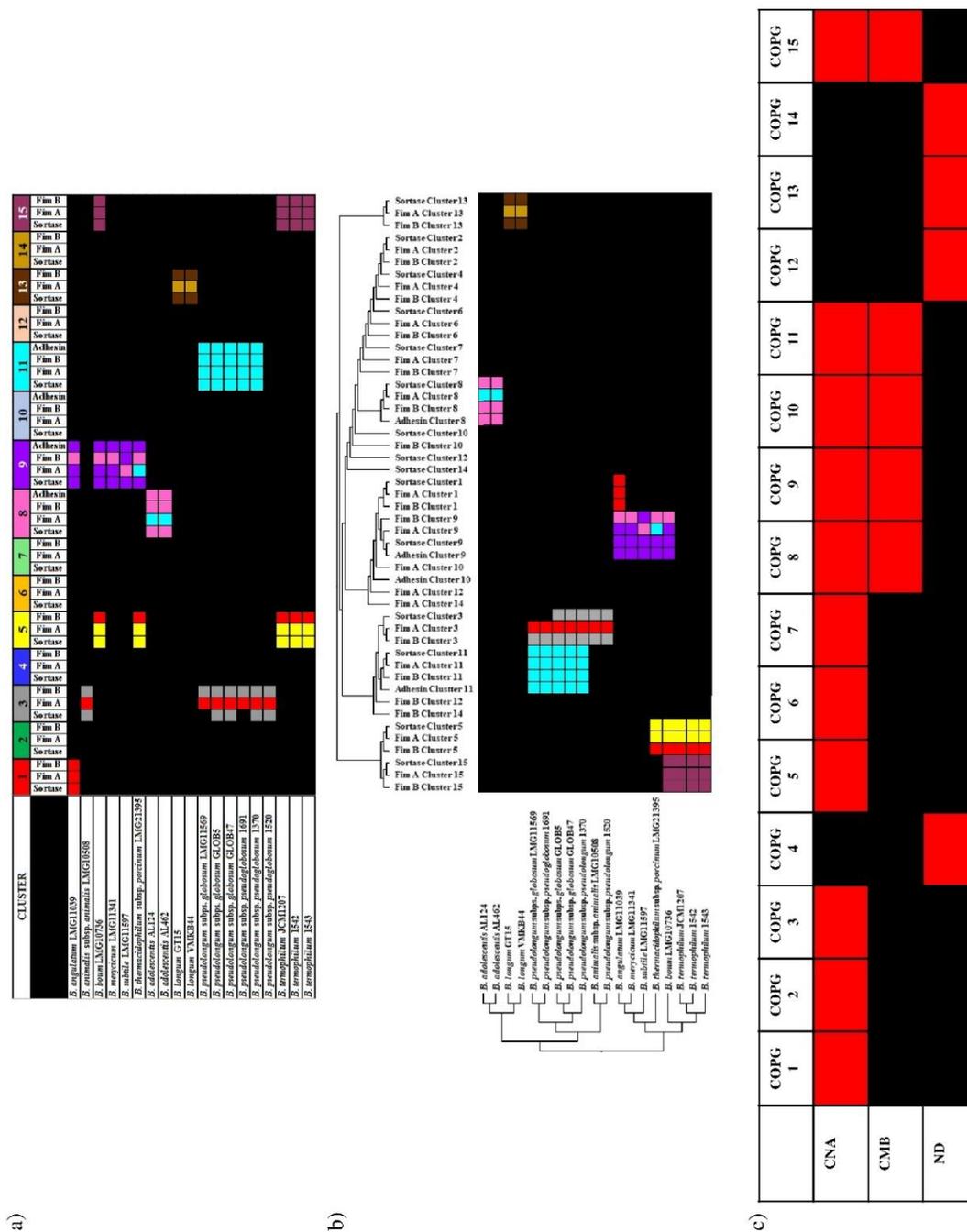


Figure 2. Genetic comparison of COPGs within the genus *Bifidobacterium*. Panels a and b show the similarities between amino acid sequences belonging to the different COPGs reported for each COPG separately or with clustered genes and species, respectively. Column titles and colours represent different COPGs. The same colours in the heatmap represent a similarity >50 % to the COPG. Panel c depicts the presence (red) or absence (black) of CNA and CBM domains in the COPGs. ND, not detected.

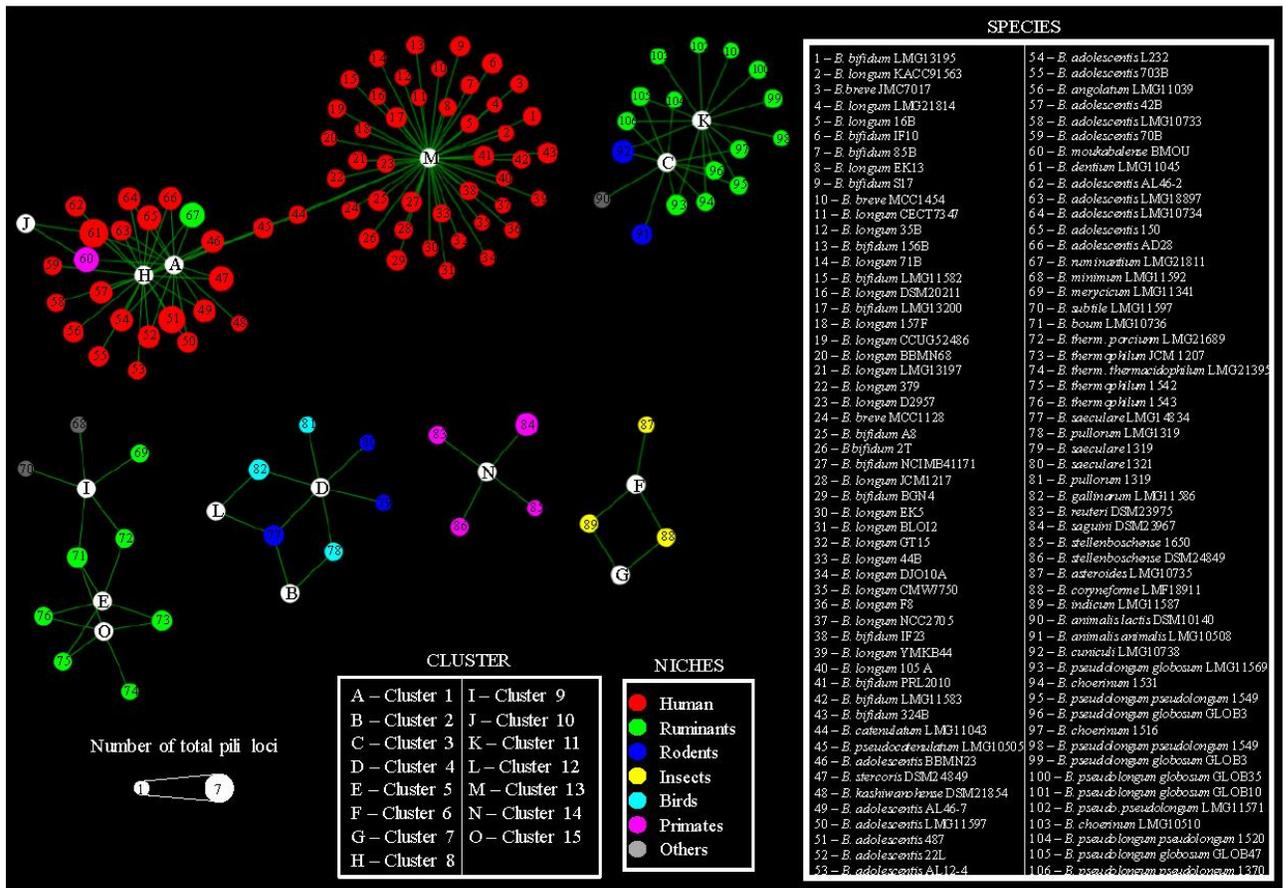


Figure 3. Network analyses of COPGs within the genus *Bifidobacterium*. Network analyses based on presence/absence of COPGs in each *Bifidobacterium* strain are indicated. White circles represent the fifteen identified COPGs, each colour denotes a different ecological niche and the diameter is proportional to the number of total pilus loci (i.e., loci belonging to COPGs as well as unique pilus loci).

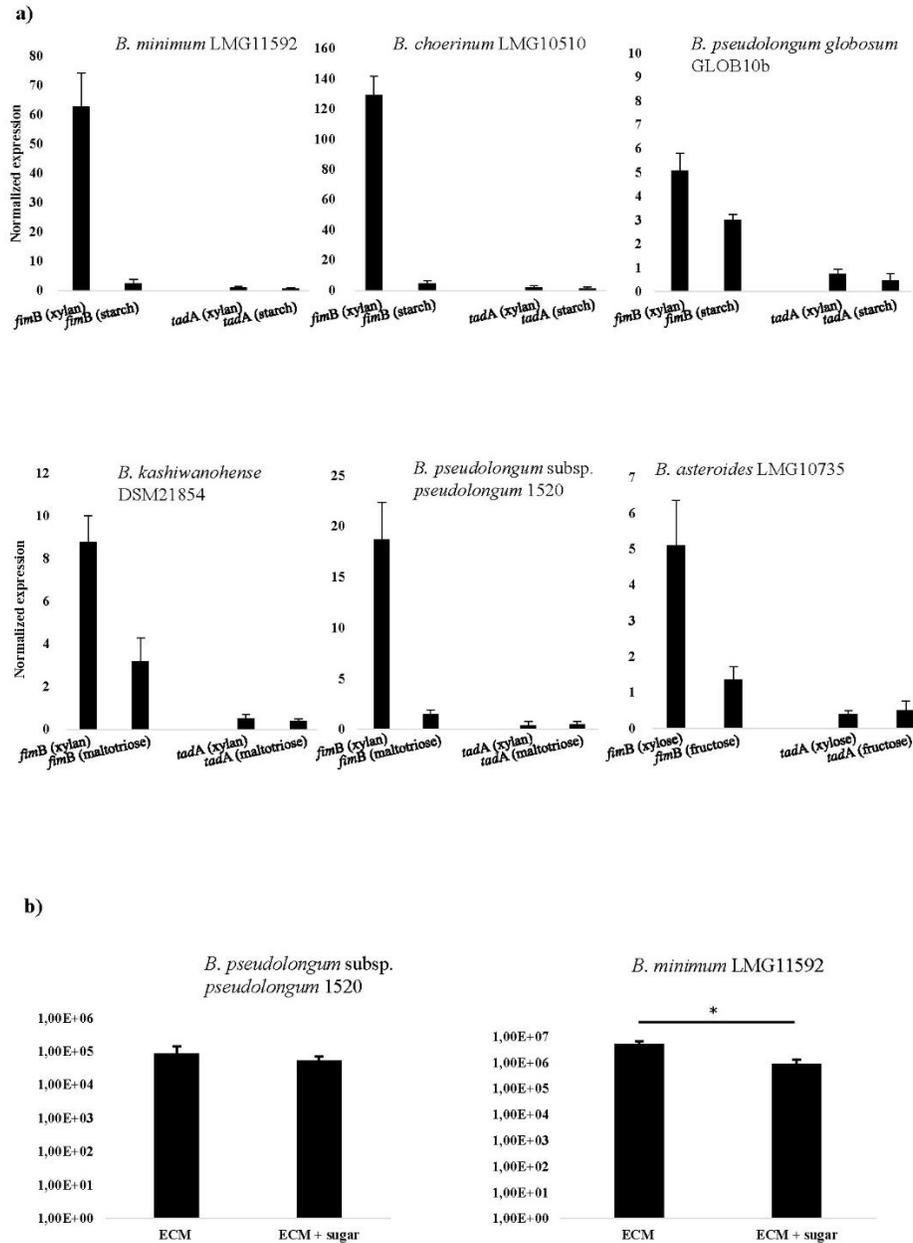


Figure 4. Relative transcription level of bifidobacterial PEL and binding to human receptor fibronectin. Panel a shows transcription levels of *fimB* and *tadA* by qRT-PCR for those bifidobacterial strains when grown on xylan or starch, i.e. *B. minimum* LMG11592, *B. choerinum* LMG10510, *B. pseudolongum subsp. globosum* GLOB10b, *B. pseudolongum subsp. pseudolongum* 1520b (all isolates from the mammalian gut); or grown on xylan or maltotriose for *B. kashiwanohense* DSM21854 (an isolate from infant gut), or grown on fructose or xylose for *B. asteroides* LMG10735, which is an isolate from the bee hindgut. Panel b displays the adhesion of *B. pseudolongum subsp. pseudolongum* 1520b and *B. minimum* LMG11592 to fibronectin alone or to fibronectin with the presence of the sugar xylan. Y-axis represents the cfu/ml.

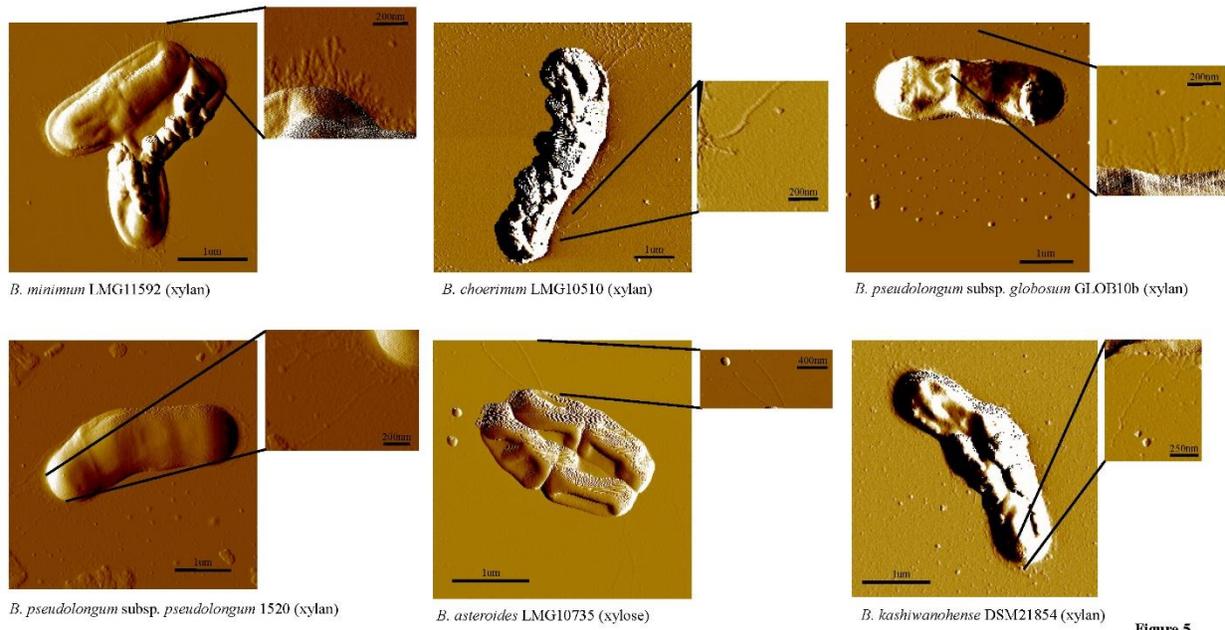


Figure 5

Figure 5. Presence and morphology of pilus-like structures in various bifidobacterial species. Images are captured by Atomic Force Microscope and show pilus-like structures of *B. minimum* LMG11592, *B. choerinum* LMG10510, *B. pseudolongum* subsp. *globosum* GLOB10b, *B. pseudolongum* subsp. *pseudolongum* 1520b, and *B. kashiwanohense* DSM21854 grown on xylan and *B. asteroides* LMG10735 grown on xylose. Scale bar is showed in each image.

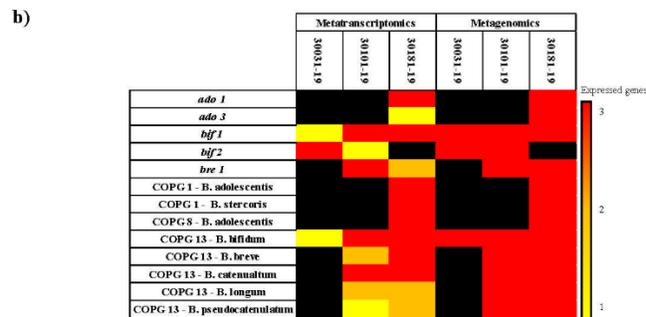
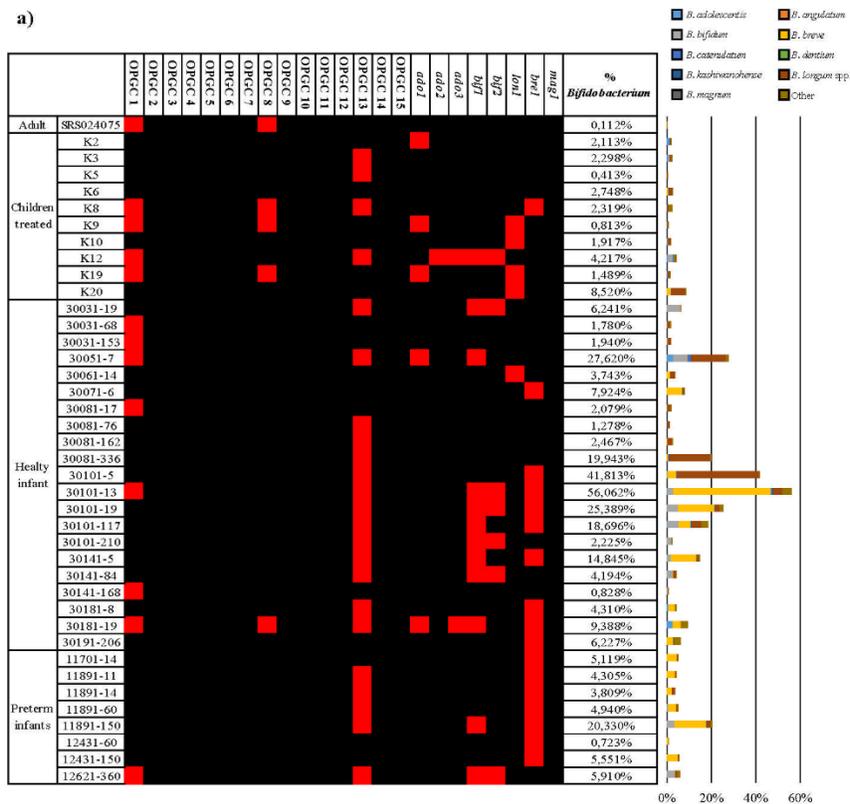


Figure 6. Occurrence of bifidobacterial COPG and PEL in fecal metagenome and metatranscriptome data sets. Panel a displays the presence of bifidobacterial COPG and PEL in the metagenomic samples of adults, healthy infants, preterm infants and children treated with antibiotics. Within the heatmap, red squares indicate the presence of the relative COPG and PEL in each metagenomic sample. Only those samples that possess a high abundance of bifidobacterial reads are reported in the heatmap. The bar plot on the right shows the relative abundance of bifidobacterial species in the analyzed metagenomic samples while the numbers on the top of the heatmap correspond to the related COPG. Panel b displays the presence of specific PEL and COPG in metagenomic samples and their relative expression in the corresponding metatranscriptomic samples. Within the heatmap, colored squares indicate the expression of one, two or three genes of the locus, dyed in yellow, orange and red, respectively.

Chapter 7

Mucosal microbiota of intestinal polyps reveals putative biomarkers of colorectal cancer

Mangifesta M, Mancabelli L, Milani C, Gaiani F, de'Angelis N, de'Angelis GL, van Sinderen D, Ventura M, Turrone F.

This chapter was published in:

Sci Rep. 2018 Sep 18;8(1):13974. doi: 10.1038/s41598-018-32413-2.

Reprinted with permission from Nature Publishing Group.

Mucosal microbiota of intestinal polyps reveals putative biomarkers of colorectal cancer

Running title: colon polyps and microbiota

Key words: colon polyps, intestinal inflammation, gut microbiota, metagenomics, dysbiosis

Marta Mangifesta¹, Leonardo Mancabelli¹, Christian Milani¹, Federica Gaiani², Nicola de' Angelis³,
Gian Luigi de'Angelis², Douwe van Sinderen⁴, Marco Ventura^{1,5} and Francesca Turrone^{1,5}

Laboratory of Probiogenomics, Department of Chemistry, Life Sciences and Environmental Sustainability, University of Parma, Parma, Italy¹; Gastroenterology and Endoscopy Unit, University Hospital of Parma, Parma, Italy²; Department of HPB Surgery and Liver Transplantation, Henri-Mondor Hospital, Université Paris Est-UPEC, Créteil, France³; School of Microbiology & APC Microbiome Institute, University College Cork, Cork, Ireland⁴; Microbiome Research Hub, University of Parma, Parma, Italy⁵;

Corresponding author: Mailing address for Francesca Turrone Department of Chemistry, Life Sciences and Environmental Sustainability, University of Parma, Parco Area delle Scienze 11/A, 43124 Parma, Italy. Phone: ++39-521-905666. Fax: ++39-521-905604. E-mail: francesca.turrone@unipr.it

Abstract

The human intestine retains a complex microbial ecosystem, which performs crucial functions that impact on host health. Several studies have indicated that intestinal dysbiosis may impact on the establishment of life-threatening intestinal diseases such as colorectal cancer. An adenomatous polyp is the result of abnormal tissue growth, which is benign but is considered to be associated with a high risk of developing colorectal cancer, based on its grade of dysplasia. Development of diagnostic tools that are based on surveying the gut microbiota and are aimed at early detection of colorectal cancer represent highly desirable target. For this purpose, we performed a pilot study in which we applied a metataxonomic analysis based on 16S rRNA gene sequencing approach to unveil the composition of microbial communities of intestinal polyps. Moreover, we performed a meta-analysis involving the reconstructed microbiota composition of adenomatous polyps and publicly available metagenomics datasets of colorectal cancer. These analyses allowed the identification of microbial taxa such as *Faecalibacterium*, *Bacteroides* and *Romboutsia*, which appear to be depleted in cancerogenic mucosa as well as in adenomatous polyps, thus representing novel microbial biomarkers associated with early tumor formation. Furthermore, an absolute quantification of *Fusobacterium nucleatum* in polyps further compounded the important role of this microorganism as a valuable putative microbial biomarker for early diagnosis of colorectal cancer

Introduction

Adenomatous polyps or adenomas have long been recognized as precursors of colorectal cancer^{1,2}. Colorectal cancer (CRC) is one of the most commonly diagnosed cancers worldwide and consequently one of the major causes of death in developed countries³. It is known that adenomatous polyps in some cases may evolve to colorectal cancer, even though no currently available scientific evidence unequivocally demonstrate in how many years and what are the precise causes of degeneration. So far there is a growing number of studies that highlight a direct correlation between polyp size, histology and progression of this pathology to CRC⁴. Furthermore, the molecular causes promoting such malignant transformation of polyps are essentially unknown⁵. Several factors have been investigated for polyp involvement in cancer development, such as genetics, epigenetics, diet, life style, obesity, alcohol intake and smoking⁶. In recent decades, it has become evident that gut bacteria and their metabolites may participate in triggering or progression of colorectal cancer through various proposed mechanisms, including the production of reactive oxygen radicals and other genotoxins⁷⁻¹⁰, phenolic compound, and indole production¹¹, as well as conversion of dietary factors into carcinogens and tumor promoters¹², and induction of proinflammatory and procarcinogenic pathways in host epithelial cells¹³⁻¹⁵. These proposed mechanisms have an impact in altering the metabolic environment of the host, which may directly or indirectly influence mutagenesis rates and thus carcinogenesis¹⁶. Various studies have explored the gut microbiota of individuals with

CRC, resulting in the identification of a range of different bacterial groups being associated with carcinogenesis, including *Helicobacter pylori*, *Bacteroides fragilis*, *Fusobacterium nucleatum*, various pathogenic *Escherichia coli*, *Streptococcus gallolyticus* and members of the genus *Clostridium*¹⁷⁻²⁶. Although screening and lifestyle play important roles in early detection and prevention of colorectal cancer, the identification of new diagnostic tools in particular for the early detection of tumors is important. Previous studies have compared the gut microbiota composition of stool samples from healthy people versus patient with polyps^{27,28}. However, the fecal microbiota is only partially representing the autochthonous microbiota that is in direct contact with the intestinal mucosa^{29,30}. Thus, in order to evaluate the role of the indigenous gut microbiota of subjects undergoing colic mucosectomies of polypoid lesions, we performed a pilot study collecting gut biopsies from healthy mucosa and polyps from the same subjects. On these samples, we performed a meta-analysis involving the here reconstructed microbiota composition of adenomatous polyps and publicly available metagenomic datasets of colorectal cancer based on 16S rRNA microbial profiling data³¹⁻³³ in order to search for the occurrence of common microorganisms between polyps and colon cancer. Altogether, these analyses revealed potential novel microbial biomarkers associated with early malignant transformation.

Materials and Methods

Patient recruitment within a pilot study. At the Gastroenterology and Digestive Endoscopy Unit, University Hospital of Parma, Italy, 12 patients

were enrolled, after obtaining informed consent (Table S1). Exclusion criteria included age > of 90, absence of colic polyps in pancolonoscopy, use of antibiotics within the past 3 weeks, inability to sign informed consent. Colonic Mucosa with Polyp (CMP) and Healthy Marginal Tissue (HMT) were obtained from the same subject collected by endoscopy biopsy forceps, thus generating a total of 24 samples (mean age of subjects: 66±12 years old), immediately frozen at -80°C and stored until further processing for DNA extraction. DNA was extracted from each sample using the PowerViral™ Environmental RNA/DNA Isolation Kit Sample (Qiagen) following the manufacturer's instructions (Qiagen Ltd, Strasse, Germany). This study protocol was approved by Ethics Committee of Parma Province and all experiments were performed in accordance with relevant guidelines and regulations. Each patient was characterized by demographic data, level of disease indicators and therapy.

16S rRNA gene amplification. Partial 16S rRNA gene sequences were amplified from extracted DNA using primer pair Probio_Uni (5'-CCTACGGGRSGCAGCAG-3')/ Probio_Rev (5'-ATTACCGCGGCTGCT-3'), which targets the V3 region of the 16S rRNA gene sequence³⁴. Illumina adapter overhang nucleotide sequences were then added to the partial 16S rRNA gene-specific amplicons, which in turn were further processed by employing the 16S Metagenomic Sequencing Library Preparation Protocol (Part #15044223 Rev. B – Illumina; see also below). Amplifications were carried out using a Verity Thermocycler (Applied Biosystems). The integrity of the PCR amplicons was analysed by electrophoresis on a 2200

TapeStation Instrument (Agilent Technologies, USA).

MiSeq sequencing of 16S rRNA Gene-based amplicons. PCR products obtained following the amplification of a section of the 16S rRNA gene were purified by a magnetic purification step involving Agencourt AMPure XP DNA purification beads (Beckman Coulter Genomics GmbH, Bernried, Germany) in order to remove primer dimers. The DNA concentration of the amplified sequence library was estimated employing a fluorimetric Qubit quantification system (Life Technologies). Amplicons were diluted to 4 nM and 5 μ l of each diluted DNA amplicon sample was mixed to prepare the pooled final library. Paired-end sequencing (250 bp x 2) was performed using an Illumina MiSeq sequencer with MiSeq Reagent Kit v3 chemicals-600 cycles (Illumina Inc., California, USA).

16S rRNA-microbial profiling analysis. The fastq files were processed using QIIME³⁵ as previously described³⁴. Paired-end reads were merged, and quality control implementation allowed the retention of sequences with a length between 140 and 400 bp, mean sequence quality score >25 and with truncation of a sequence at the first base if a low quality sequence within a rolling 10 bp window was found. Sequences with mismatched forward and/or reverse primers were omitted. 16S rRNA Operational Taxonomic Units (OTUs) were defined at $\geq 97\%$ sequence homology using uclust³⁶ and OTUs with less than 10 sequences were filtered. The Biological Observation Matrix (BIOM) obtained was analysed by summarize_taxa.py script in order to get the relative abundance of each taxonomic group for all samples. All reads were classified to the

lowest possible taxonomic rank using QIIME³⁵ and a reference dataset from the SILVA database³⁷. The microbial richness of the samples (alpha-diversity) was evaluated with the Chao1 index through the alpha_rarefaction.py script included in the Qiime software suite using default parameters. Similarities between samples (beta-diversity) were calculated by unweighted uniFrac^{35,38}. Hierarchical clusterings were performed with TMeV 4.8.1 (<http://www.tm4.org>) using Pearson correlation.

Statistical analyses. SPSS software (www.ibm.com/software/it/analytics/spss/) were used to complete statistical analysis. All data were presented as means \pm SD. Furthermore, differential abundance of bacteria taxa and alpha-diversity were tested by one-way analysis of variance (ANOVA).

Quantitative PCR. qPCR experiments were based on species-specific primers of *Fusobacterium nucleatum* targeting a previously described prostaglandin transporter-encoding gene³⁹. The genome copy-number and the deduced cell number (since the genes targeted were in single copy per genome) was evaluated by comparing the cycle threshold (Ct) values obtained with those from a standard curve. Standard curves were calculated from serial dilutions of a culture with a known cell number (as determined by viable count assessment) for the bacterial strain versus Ct produced for each target gene. The primer sequences were as follows:
forward primer 5'-CAACCATTACTTTAACTCTACCATGTTCA-3' and reverse primer 5'-GTTGACTTTACAGAAGGAGATTATGTAAA AATC-3'. qPCR was performed using the CFX96 system (BioRad, CA, USA). Each PCR reaction

mix contained the following: 12.5µl 2x SYBR SuperMix Green (BioRad, CA, USA), 1 µl of DNA dilution, each of the forward and reverse primers at 0.5 µM and nuclease-free water was added to obtain a final volume of 20 µl. PCR products were detected with SYBR Green fluorescent dye and amplified according to the following protocol: one cycle of 95°C for 2 minutes, followed by 42 cycles of 95°C for 5 s and 60°C for 30 s. Melting curve: 65°C to 95°C with increments of 0.5°C/s. In each run, negative controls for each primer set were included. Cycle thresholding was calculated using the automated settings for Biorad CFX Manager 3.1 software (BioRad). The entire qPCR experiment was performed a second time using the same samples and methods as outlined above, for the purpose of replication, and very similar results were obtained.

Data Deposition. The 16S rRNA profiling data sequenced in this study were deposited in SRA database under accession number PRJNA415554.

Results and Discussion

Patients and pathological data

A total of twelve patients with polyps were included in this pilot study. From those, histological analyses revealed that four corresponded to adenomatous polyps and eight to hyperplastic polyps. All adenomatous polyps were found in males, while the remaining hyperplastic polyps were collected from three males and five females. Two biopsies were collected from each patient, one corresponding to the polyp and one to healthy mucosa. Patient characteristics and details of endoscopic treatment performed are outlined in Table S1.

Cataloguing of polyp-associated microbiota and comparison to healthy mucosa microbiota.

Twelve Colonic Mucosa with Polyp (CMP) and corresponding Healthy Marginal Tissue (HMT) samples (i.e. total of 24 samples), retrieved from twelve hospitalized subjects, were used to evaluate the microbiota taxonomic profile by means of 16S rRNA gene sequencing analysis as described previously³⁴.

Next generation sequencing of these samples produced a total of 2,059,497 sequencing reads with an average of 85,812 reads per sample (Table S2). Subsequent quality and chimera filtering produced a total of 1,479,134 filtered reads with an average of 61,631 reads, and ranging from 33,783 to 202,018 reads per sample (Table S2).

Rating of rarefaction curves based on the Chao1 biodiversity indexes calculated for the two subgroups of sequenced read pools showed that both curves tend to reach a plateau confirming the high accuracy of the performed 16S rRNA profiling analysis. Furthermore, average rarefaction curves revealed a different trend of bacterial richness of the gut microbiota of healthy samples compared to that found in the polyp sample group, i.e. higher level of HMT vs CMP samples (Fig. S1). However, these analyses displayed no statistical significance, since the ANOVA calculated at the highest rarefaction depths reached by all the samples, i.e. Chao1 average is 1895 ± 480 in HMT and 1678 ± 634 in CMP indicated a p-value > 0.05 (Fig. S1).

16S rRNA profiling of HMT and CMP samples.

Inspection of predicted taxonomic profiles at phylum level for all samples showed that Firmicutes represented the dominant phylum (average of $57.59 \% \pm 10.48 \%$) of the colonic

mucosa samples, out-numbering the Bacteroidetes (average of 9.24 % \pm 8.01%) and Proteobacteria (average of 10.26 % \pm 8.19 %) phyla.

Nevertheless, no major shifts were observed in bacterial populations between healthy mucosa and Polyp-associated mucosa, although we did observe a moderate reduction of Actinobacteria (relative percentage difference of -13.43 %) and Bacteroides (relative percentage difference of -1.02 %) in CMP and a concomitant modest increase of Firmicutes (relative percentage difference of 2.84 %) compared to HMT (Fig. 1a-c).

Furthermore, at genus level, we identified 308 genera of which 225 appear to be present in at least one sample of HMT and CMP groups. In detail, 35 bacterial taxa were present in all 24 samples, thus representing the mucosa core microbiota, i.e. the permanent and common members of the microbial community⁴⁰, which may include specific species important for the maintenance of an efficiently functioning gut ecosystem⁴¹ (Fig. 1b).

Moreover, HMT displayed 29 unique taxa representing the 0.015 % of the total genera abundance, while CMP showed 19 unique representing 0.008 % of the total genera abundance (Fig. 1b).

Taxa associated with HMT and CMP.

In order to further identify differences in microbiota composition between CMP and HMT patients, we focused on 30 taxa, which were shown to be present as the absolute change of $> \pm 0.1$ % or that displayed a significant variation in terms of relative abundance (p-value < 0.05). In this context, these 30 taxa varied in terms of relative abundance, encompassing an increase ranging from 10 % (*Lachnoclostridium*) to 641 % (*Romboutsia*), or a decrease ranging from -16 % (*Ruminococcaceae*

UCG-002) to -74 % (*Ruminococcus* I) in HMT datasets compared to those obtained from CMP samples (Fig. 1c). Interestingly, members of the Actinobacteria phylum such as the *Bifidobacterium* genus are present in higher abundance in HMT as compared to CMP samples. Bifidobacteria species are known to represent typical gut microorganisms with presumed positive effects on human health and recently their absence has been linked to various disease conditions⁴²⁻⁴⁵. In addition, the *Faecalibacterium* genus, which is among bacteria with purported health-promoting activities⁴⁶, as well as *Bacteroides* taxa, which have been found at a lower abundance in IBD patients⁴⁷ are present at a higher level in HMT samples. Moreover, *Romboutsia* taxa is the genus with the highest relative abundance in HMT vs. CMP samples. Notably, *Romboutsia* is a recently described bacterial genus^{48,49} commonly identified in the human gut^{50,51}, and often associated with a healthy status of patients^{50,51}. The drastic reduction of members of this particular genus (-86,51%) in mucosa associated with polyps may represent a potential microbial indicator of a disease condition⁵⁰. In contrast, the higher presence of *Helicobacter* and *Klebsiella*, which are typically associated with carcinogenesis^{52,53} in CMP samples reflect a typical disease condition. Interestingly, among taxa showing a significant variation in terms of relative abundance (p-value < 0.05), Unclassified member (U. m.) of *Peptostreptococcaceae* family occurs with an increased relative abundance of 337.30% in HMT vs. CMP samples, while *Peptoclostridium* is totally absent in all analyzed CMP samples.

In order to integrate and confirm the observed differences in relative abundance between HMT and CMP samples, we defined the trend (increase

or decrease) for each bacterial taxa in healthy subjects versus polyp condition. Considering all patients, we identified that only the genus *Romboutsia* displayed the same trend in at least 80% of HMT samples, while all taxa trend of CMP were < 75 %. These results reinforce the notion that *Romboutsia* may play a key role in maintaining the health status of the host and render this taxon a very valuable candidate as biomarker of intestinal dysbiosis.

Identification of possible cancer biomarkers.

In order to reinforce our taxonomical observation and to enlarge the CMP database, we performed a survey and selection of 16S rRNA-based polyp-associated metagenomic datasets deposited in the publicly NCBI database. However, the published polyp-associated metagenomic datasets are based on fecal samples ^{27,28}. In this context, it is worth mentioning that previous studies on healthy patients or children with rectal bleeding indicated that fecal and mucosal samples belonging to the same individual displayed a very dissimilar microbial composition ^{29,30}. It has been reported that, based on different statistical analyses, fecal and mucosal samples cluster separately, at family and species taxonomic level ^{29,30}. To circumvent the issue of absence of data from polyp-associated biopsies, we decided to use human CRC datasets. In fact, as discussed above, CRC has been considered to represent a later stage of adenomas ^{2,54}. Consequently, we retrieved 293 samples from the four publically available CRC BioProjects ³¹⁻³³ (Table S3). In detail, we selected the datasets based on Illumina or 454 technologies including both healthy and CRC samples obtained from biopsies collected from the human large intestine (Table S3). Notably, to avoid biases caused by different

bioinformatic analyses pipelines ⁵⁵, the sequence reads pools of each study were analyzed involving the same custom script ⁵⁶ based on the QIIME software suite ³⁵ used here for the 24 HMT/CMP samples. These datasets were enriched with the CMP and HMT 16S rRNA microbial profiling data obtained and described in our current study.

Evaluation of the difference between all HHealthy TTissue (HT) and CRC-polyps TTissue (CpT) samples highlighted 38 genera with a p-value < 0.05 calculated through Analysis of variance (ANOVA) (Fig. 2). Furthermore, we investigated the difference between each case-control, HT vs. CRC and HMT vs. CMP, in order to calculate the trend percentage.

Association of ANOVA and trend percentage analyses indicated that CpT samples had a higher abundance of genera belonging to Proteobacteria and Fusobacteria phyla, specifically *Campylobacter* (taxonomic trend prevalence of 100% in CpT) and *Fusobacterium* (taxonomic trend prevalence of 60 % in CpT) compared to healthy samples (Fig. 3). Interestingly, these taxa are commonly found in patients with CRC ⁵⁷, with the *Fusobacterium* genus frequently being associated with tumor progression ⁵⁸. Moreover, CpT samples possessed microbial genera commonly involved in development of intestinal disease or pathologies, such as *Streptococcus* (taxonomic trend prevalence of 80.00 % in CpT) and *Gemella* (taxonomic trend prevalence of 80 % in CpT) ^{59,60}. In contrast, HT samples showed higher abundance of ‘beneficial’ bacterial genera, as well as U. m. of *Lachnospiraceae* family and *Faecalibacterium* (taxonomic trend prevalence of 100 % in HT), *Bacteroides* (taxonomic trend prevalence of 80 % in HT), and *Romboutsia*

(taxonomic trend prevalence of 60 % in HT) compared to CpT.

Furthermore, focusing on genera with a taxonomic trend with a prevalence of 100 % and a significant p-value (< 0.05) calculated between HT and CpT (Fig. 3), it was possible to identify five taxa characteristic of HT and one taxon of CpT subjects. In detail, healthy samples showed high abundance of U. m. of *Lachnospiraceae* family, *Lachnoclostridium*, *Faecalibacterium*, *Ruminiclostridium* and *Subdoligranulum*, while disease-associated CpT samples showed high abundance of *Campylobacter*. These results reflect the assumption that genera belonging to *Lachnospiraceae* and *Ruminococcaceae* families may protect healthy subjects from intestinal disease⁶¹, such as through the production of short-chain fatty acids (SCFAs)⁶². In contrast, the higher abundance of bacteria belonging to *Campylobacter* genus may confirm its correlation with the development of gastrointestinal diseases^{63,64}. Overall, we found a correspondence between the trend that we retrieved in our HMT vs CMP samples, and that of the HT vs CRC samples, revealing a small number of microbial taxa that could be considered as bacterial biomarkers for early detection of tumor or, conversely, reflecting a healthy condition.

Quantification of *Fusobacterium nucleatum* by qPCR. Since several studies associated the presence of *Fusobacterium nucleatum* with a formation and progression of colorectal cancer^{39,64-67} the relative abundances of this species were measured by qPCR in all 24 samples (HTM and CMP biopsies of 12 patients). Notably qPCR analyses displayed that this bacterium has been found in six patients (i.e., 2, 5, 6, 7, 9 and 13) with

a genome copy number ranging from 2.24E1 to 1.25E3 for the HTM samples, and a genome copy number range of 2.08E2 to 3.26E3 for CMP biopsies (Fig. 4). These data clearly indicated that the total load of *Fusobacterium nucleatum* was significantly (p-value <0.04) higher for samples 5, 6, 7 and 9 in the CMP vs. HMT mucosa for each patient. Interestingly, sample 13, showed a total load of *Fusobacterium nucleatum* higher respect all other samples, i.e., a mean value of 1.02E3 for the HMT as compared to a mean value of 2.37E3 for the CMP (p-value < 0.04) (Fig. 4). These preliminary data suggest that this species can, as previously indicated, not only be considered an important bacterial biomarker for the diagnosis of colorectal cancer^{39,66}, but may also serve as a putative valuable microbial biomarker for early diagnosis of malignant displacement of polyps.

Conclusions

Recent findings have shown that intestinal diseases are correlated with microbiota alterations such as in the case of ulcerative colitis^{45,68}. The contribution of gastrointestinal microbiota in the progression and development of colorectal cancer is complex and not fully understood⁶⁷, but previous *in vivo* murine based-analyses highlight the potential role of many individual bacteria to promote carcinogenesis^{28,69}. In the current pilot study, we compared bacterial population in mucosa samples of patients with polyps or CRC, versus healthy mucosa, showing that the *Romboutsia* genus is more abundant in healthy compared to polyp-associated tissue. Remarkably, the relative abundance of *Romboutsia* in CRC tissue was found to be drastically reduced. Thus, we may argue that the absence of this microbial genus is a first

indicator of an alteration of the mucosa. Remarkably, this microbial taxon has already been observed for other gut-associated diseases⁴⁹⁻⁵¹. Interestingly, the *Peptoclostridium* genus was undetectable in all CMP sample tested, while analysis of the HMT samples revealed a relatively high number of U. m. of *Peptostreptococcaceae* family. Our findings were in line with the findings of a meta-analysis in which we compared the CMP samples with CRC samples from publically available metagenomic datasets³¹⁻³³. Our metagenomic analysis revealed that CpT samples had a higher abundance of the Fusobacteria phylum compared to healthy samples. Interestingly, the *Fusobacterium* genus is often involved in tumor progression^{58,70}. At genus level, we identified five taxa characteristic of HT and one taxon typical of CpT subjects. In detail, healthy samples were shown to contain a high abundance of U. m. of *Lachnospiraceae* family, *Lachnoclostridium*, *Faecalibacterium*, *Ruminiclostridium* and *Subdoligranulum*, while CpT samples showed high abundance of *Campylobacter*. The higher abundance of bacteria belonging to the *Campylobacter* genus in CRC mucosa, as well as polyp-associated samples reinforces its correlation with the development of intestinal pathologies^{63,64}. On the other hand, genera belonging to *Lachnospiraceae* and *Ruminococcaceae* families appear to be correlated to beneficial effects on the host as described previously⁶¹. Gut microbiota include a very complex array of microorganisms and its compositional changes have been shown to impact on human health status^{71,72}. The identification of microbial players that are distinctly altered in their abundance in CRC-associated samples may allow early diagnosis of

malignant transformation as well as the identification of protective bacteria missing in polyps or CRC tissue. In conclusion, the data obtained in this pilot study have allowed us to identify possible microbial biomarkers associated with polyps and/or CRC. However, such findings need further validation by analysis of a larger number of samples, in order to further substantiate and fully characterize such potential microbial biomarkers of CRC development, as well as beneficial bacteria, and to identify their functions and interactions with the host.

Funding

This work was funded by the EU Joint Programming Initiative – A Healthy Diet for a Healthy Life (JPI HDHL, <http://www.healthydietforhealthylife.eu/>) and the MIUR to MV. We thank GenProbio srl for financial support of the Laboratory of Probiogenomics. DvS is a member of The APC Microbiome Institute funded by Science Foundation Ireland (SFI), through the Irish Government's National Development Plan (Grant number SFI/12/RC/2273).

Author contributions statement

M.M. performed laboratory experiments, L.M. and C.M. analyzed bioinformatics data, F.G. collected clinical samples and clinical data, N.de' A., G.L.de'A. and D.vS. reviewed the manuscript for important intellectual content, M.V. and F.T. conceived the study, interpreted data and wrote the manuscript.

Acknowledgments

Part of this research is conducted using the High Performance Computing (HPC) facility of the University of Parma. The authors declare that they have no competing interest

References

- 1 Fenoglio, C. M. & Lane, N. The anatomical precursor of colorectal carcinoma. *Cancer* **34**, suppl:819-823 (1974).
- 2 Leslie, A., Carey, F. A., Pratt, N. R. & Steele, R. J. The colorectal adenoma-carcinoma sequence. *Br J Surg* **89**, 845-860, doi:10.1046/j.1365-2168.2002.02120.x (2002).
- 3 Boyle, P. & Ferlay, J. Cancer incidence and mortality in Europe, 2004. *Ann Oncol* **16**, 481-488, doi:10.1093/annonc/mdi098 (2005).
- 4 Lieberman, D., Moravec, M., Holub, J., Michaels, L. & Eisen, G. Polyp size and advanced histology in patients undergoing colonoscopy screening: implications for CT colonography. *Gastroenterology* **135**, 1100-1105, doi:10.1053/j.gastro.2008.06.083 (2008).
- 5 Grady, W. M. & Markowitz, S. D. The molecular pathogenesis of colorectal cancer and its potential application to colorectal cancer screening. *Dig Dis Sci* **60**, 762-772, doi:10.1007/s10620-014-3444-4 (2015).
- 6 Marmol, I., Sanchez-de-Diego, C., Pradilla Dieste, A., Cerrada, E. & Rodriguez Yoldi, M. J. Colorectal Carcinoma: A General Overview and Future Perspectives in Colorectal Cancer. *Int J Mol Sci* **18**, doi:10.3390/ijms18010197 (2017).
- 7 Mandal, P. Potential biomarkers associated with oxidative stress for risk assessment of colorectal cancer. *Naunyn Schmiedebergs Arch Pharmacol* **390**, 557-565, doi:10.1007/s00210-017-1352-9 (2017).
- 8 Huycke, M. M. & Gaskins, H. R. Commensal bacteria, redox stress, and colorectal cancer: mechanisms and models. *Exp Biol Med (Maywood)* **229**, 586-597 (2004).
- 9 Garrett, W. S. Cancer and the microbiota. *Science* **348**, 80-86, doi:10.1126/science.aaa4972 (2015).
- 10 Zitvogel, L., Daillere, R., Roberti, M. P., Routy, B. & Kroemer, G. Anticancer effects of the microbiome and its products. *Nat Rev Microbiol* **15**, 465-478, doi:10.1038/nrmicro.2017.44 (2017).
- 11 Smith, E. A. & Macfarlane, G. T. Formation of Phenolic and Indolic Compounds by Anaerobic Bacteria in the Human Large Intestine. *Microb Ecol* **33**, 180-188 (1997).
- 12 Lax, A. J. Opinion: Bacterial toxins and cancer--a case to answer? *Nat Rev Microbiol* **3**, 343-349, doi:10.1038/nrmicro1130 (2005).
- 13 Gounaris, E. *et al.* Mast cells are an essential hematopoietic component for polyp development. *Proc Natl Acad Sci U S A* **104**, 19977-19982, doi:10.1073/pnas.0704620104 (2007).
- 14 Gounaris, E. *et al.* Live imaging of cysteine-cathepsin activity reveals dynamics of focal inflammation, angiogenesis, and polyp growth. *PLoS One* **3**, e2916, doi:10.1371/journal.pone.0002916 (2008).
- 15 Gounaris, E. *et al.* T-regulatory cells shift from a protective anti-inflammatory to a cancer-promoting proinflammatory phenotype in polyposis. *Cancer Res* **69**, 5490-5497, doi:10.1158/0008-5472.CAN-09-0304 (2009).
- 16 Louis, P., Hold, G. L. & Flint, H. J. The gut microbiota, bacterial metabolites and colorectal cancer. *Nat Rev Microbiol* **12**, 661-672, doi:10.1038/nrmicro3344 (2014).
- 17 Bonnet, M. *et al.* Colonization of the human gut by *E. coli* and colorectal cancer risk. *Clin Cancer Res* **20**, 859-867, doi:10.1158/1078-0432.CCR-13-1343 (2014).
- 18 Collins, D., Hogan, A. M. & Winter, D. C. Microbial and viral pathogens in colorectal cancer. *The Lancet. Oncology* **12**, 504-512, doi:10.1016/S1470-2045(10)70186-8 (2011).
- 19 Swidsinski, A. *et al.* Association between intraepithelial *Escherichia coli* and colorectal cancer. *Gastroenterology* **115**, 281-286 (1998).
- 20 Nam, J. H. *et al.* *Helicobacter pylori* infection is an independent risk factor for colonic adenomatous neoplasms. *Cancer causes & control : CCC* **28**, 107-115, doi:10.1007/s10552-016-0839-x (2017).
- 21 Purcell, R. V. *et al.* Colonization with enterotoxigenic *Bacteroides fragilis* is associated with early-stage colorectal neoplasia. *PLoS One* **12**, e0171602, doi:10.1371/journal.pone.0171602 (2017).
- 22 Scanlan, P. D. *et al.* Culture-independent analysis of the gut microbiota in colorectal cancer and polyposis. *Environ Microbiol* **10**, 789-798, doi:10.1111/j.1462-2920.2007.01503.x (2008).
- 23 Doorackers, E., Lagergren, J., Engstrand, L. & Brusselaers, N. *Helicobacter pylori* eradication treatment and the risk of gastric adenocarcinoma in a Western population. *Gut*, doi:10.1136/gutjnl-2017-315363 (2018).
- 24 Chung, L. *et al.* *Bacteroides fragilis* Toxin Coordinates a Pro-carcinogenic Inflammatory Cascade via Targeting of Colonic Epithelial Cells. *Cell Host Microbe*, doi:10.1016/j.chom.2018.01.007 (2018).

- 25 Eklof, V. *et al.* Cancer-associated fecal microbial markers in colorectal cancer detection. *Int J Cancer* **141**, 2528-2536, doi:10.1002/ijc.31011 (2017).
- 26 Aymeric, L. *et al.* Colorectal cancer specific conditions promote *Streptococcus gallolyticus* gut colonization. *Proc Natl Acad Sci U S A* **115**, E283-E291, doi:10.1073/pnas.1715112115 (2018).
- 27 Hale, V. L. *et al.* Shifts in the Fecal Microbiota Associated with Adenomatous Polyps. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology* **26**, 85-94, doi:10.1158/1055-9965.EPI-16-0337 (2017).
- 28 Peters, B. A. *et al.* The gut microbiota in conventional and serrated precursors of colorectal cancer. *Microbiome* **4**, 69, doi:10.1186/s40168-016-0218-6 (2016).
- 29 Durban, A. *et al.* Assessing gut microbial diversity from feces and rectal mucosa. *Microb Ecol* **61**, 123-133, doi:10.1007/s00248-010-9738-y (2011).
- 30 Ouwehand, A. C., Salminen, S., Arvola, T., Ruuska, T. & Isolauri, E. Microbiota composition of the intestinal mucosa: association with fecal microbiota? *Microbiol Immunol* **48**, 497-500 (2004).
- 31 Geng, J., Fan, H., Tang, X., Zhai, H. & Zhang, Z. Diversified pattern of the human colorectal cancer microbiome. *Gut pathogens* **5**, 2, doi:10.1186/1757-4749-5-2 (2013).
- 32 Burns, M. B., Lynch, J., Starr, T. K., Knights, D. & Blekhnman, R. Virulence genes are a signature of the microbiome in the colorectal tumor microenvironment. *Genome Med* **7**, 55, doi:10.1186/s13073-015-0177-8 (2015).
- 33 Kostic, A. D. *et al.* Genomic analysis identifies association of *Fusobacterium* with colorectal carcinoma. *Genome Res* **22**, 292-298, doi:10.1101/gr.126573.111 (2012).
- 34 Milani, C. *et al.* Assessing the fecal microbiota: an optimized ion torrent 16S rRNA gene-based analysis protocol. *PLoS One* **8**, e68739, doi:10.1371/journal.pone.0068739 (2013).
- 35 Caporaso, J. G. *et al.* QIIME allows analysis of high-throughput community sequencing data. *Nature methods* **7**, 335-336, doi:10.1038/nmeth.f.303 (2010).
- 36 Edgar, R. C. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**, 2460-2461, doi:10.1093/bioinformatics/btq461 (2010).
- 37 Quast, C. *et al.* The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic acids research* **41**, D590-D596, doi:10.1093/nar/gks1219 (2013).
- 38 Lozupone, C. & Knight, R. UniFrac: a new phylogenetic method for comparing microbial communities. *Appl Environ Microbiol* **71**, 8228-8235, doi:10.1128/AEM.71.12.8228-8235.2005 (2005).
- 39 Castellarin, M. *et al.* *Fusobacterium nucleatum* infection is prevalent in human colorectal carcinoma. *Genome Res* **22**, 299-306, doi:10.1101/gr.126516.111 (2012).
- 40 Astudillo-Garcia, C. *et al.* Evaluating the core microbiota in complex communities: A systematic investigation. *Environ Microbiol* **19**, 1450-1462, doi:10.1111/1462-2920.13647 (2017).
- 41 Shetty, S. A., Hugenholtz, F., Lahti, L., Smidt, H. & de Vos, W. M. Intestinal microbiome landscaping: insight in community assemblage and implications for microbial modulation strategies. *FEMS Microbiol Rev* **41**, 182-199, doi:10.1093/femsre/fuw045 (2017).
- 42 Ventura, M. *et al.* Genome-scale analyses of health-promoting bacteria: probiogenomics. *Nat Rev Microbiol* **7**, 61-71, doi:10.1038/nrmicro2047 (2009).
- 43 Hidalgo-Cantabrana, C. *et al.* Bifidobacteria and Their Health-Promoting Effects. *Microbiology spectrum* **5**, doi:10.1128/microbiolspec.BAD-0010-2016 (2017).
- 44 Arboleya, S., Watkins, C., Stanton, C. & Ross, R. P. Gut Bifidobacteria Populations in Human Health and Aging. *Front Microbiol* **7**, 1204, doi:10.3389/fmicb.2016.01204 (2016).
- 45 Duranti, S. *et al.* Elucidating the gut microbiome of ulcerative colitis: bifidobacteria as novel microbial biomarkers. *FEMS Microbiol Ecol* **92**, doi:10.1093/femsec/fiw191 (2016).
- 46 Miquel, S. *et al.* *Faecalibacterium prausnitzii* and human intestinal health. *Current opinion in microbiology* **16**, 255-261, doi:10.1016/j.mib.2013.06.003 (2013).
- 47 Rajilic-Stojanovic, M. *et al.* Global and deep molecular analysis of microbiota signatures in fecal samples from patients with irritable bowel syndrome. *Gastroenterology* **141**, 1792-1801, doi:10.1053/j.gastro.2011.07.043 (2011).
- 48 Gerritsen, J. *et al.* Characterization of *Romboutsia ilealis* gen. nov., sp. nov., isolated from the gastrointestinal tract of a rat, and proposal for the reclassification of five closely related members of the genus *Clostridium* into the genera *Romboutsia* gen. nov., *Intestinibacter* gen. nov., *Terrisporobacter*

- gen. nov. and *Asaccharospora* gen. nov. *Int J Syst Evol Microbiol* **64**, 1600-1616, doi:10.1099/ijs.0.059543-0 (2014).
- 49 Gerritsen, J. *et al.* Genomic and functional analysis of *Romboutsia ilealis* CRIB(T) reveals adaptation to the small intestine. *PeerJ* **5**, e3698, doi:10.7717/peerj.3698 (2017).
- 50 Ricaboni, D., Mailhe, M., Khelaifia, S., Raoult, D. & Million, M. *Romboutsia timonensis*, a new species isolated from human gut. *New microbes and new infections* **12**, 6-7, doi:10.1016/j.nmni.2016.04.001 (2016).
- 51 Milani, C. *et al.* Gut microbiota composition and *Clostridium difficile* infection in hospitalized elderly individuals: a metagenomic study. *Scientific reports* **6**, 25945, doi:10.1038/srep25945 (2016).
- 52 Mera, R. M. *et al.* Dynamics of *Helicobacter pylori* infection as a determinant of progression of gastric precancerous lesions: 16-year follow-up of an eradication trial. *Gut*, doi:10.1136/gutjnl-2016-311685 (2017).
- 53 Stecher, B., Maier, L. & Hardt, W. D. 'Blooming' in the gut: how dysbiosis might contribute to pathogen evolution. *Nat Rev Microbiol* **11**, 277-284, doi:10.1038/nrmicro2989 (2013).
- 54 Dulal, S. & Keku, T. O. Gut microbiome and colorectal adenomas. *Cancer journal* **20**, 225-231, doi:10.1097/PPO.000000000000050 (2014).
- 55 Duvallet, C., Gibbons, S. M., Gurry, T., Irizarry, R. A. & Alm, E. J. Meta-analysis of gut microbiome studies identifies disease-specific and shared responses. *Nat Commun* **8**, 1784, doi:10.1038/s41467-017-01973-8 (2017).
- 56 Mancabelli, L. *et al.* Identification of universal gut microbial biomarkers of common human intestinal diseases by meta-analysis. *FEMS Microbiol Ecol* **93**, doi:10.1093/femsec/fix153 (2017).
- 57 Otto, C. C., Chen, L. H., He, T., Tang, Y. W. & Babady, N. E. Detection of gastrointestinal pathogens in oncology patients by highly multiplexed molecular panels. *Eur J Clin Microbiol Infect Dis* **36**, 1665-1672, doi:10.1007/s10096-017-2981-0 (2017).
- 58 Abed, J. *et al.* Tumor Targeting by *Fusobacterium nucleatum*: A Pilot Study and Future Perspectives. *Frontiers in cellular and infection microbiology* **7**, 295, doi:10.3389/fcimb.2017.00295 (2017).
- 59 Allali, I. *et al.* Gut microbiome compositional and functional differences between tumor and non-tumor adjacent tissues from cohorts from the US and Spain. *Gut Microbes* **6**, 161-172, doi:10.1080/19490976.2015.1039223 (2015).
- 60 Chen, Y. *et al.* Dysbiosis of small intestinal microbiota in liver cirrhosis and its association with etiology. *Scientific reports* **6**, 34055, doi:10.1038/srep34055 (2016).
- 61 Lepage, P. *et al.* Twin study indicates loss of interaction between microbiota and mucosa of patients with ulcerative colitis. *Gastroenterology* **141**, 227-236, doi:10.1053/j.gastro.2011.04.011 (2011).
- 62 Flint, H. J., Scott, K. P., Duncan, S. H., Louis, P. & Forano, E. Microbial degradation of complex carbohydrates in the gut. *Gut Microbes* **3**, 289-306, doi:10.4161/gmic.19897 (2012).
- 63 Akutko, K. & Matusiewicz, K. *Campylobacter concisus* as the etiologic agent of gastrointestinal diseases. *Advances in clinical and experimental medicine : official organ Wroclaw Medical University* **26**, 149-154, doi:10.17219/acem/61428 (2017).
- 64 Warren, R. L. *et al.* Co-occurrence of anaerobic bacteria in colorectal carcinomas. *Microbiome* **1**, 16, doi:10.1186/2049-2618-1-16 (2013).
- 65 McCoy, A. N. *et al.* *Fusobacterium* is associated with colorectal adenomas. *PLoS One* **8**, e53653, doi:10.1371/journal.pone.0053653 (2013).
- 66 Kostic, A. D. *et al.* *Fusobacterium nucleatum* potentiates intestinal tumorigenesis and modulates the tumor-immune microenvironment. *Cell Host Microbe* **14**, 207-215, doi:10.1016/j.chom.2013.07.007 (2013).
- 67 Bullman, S. *et al.* Analysis of *Fusobacterium* persistence and antibiotic response in colorectal cancer. *Science* **358**, 1443-1448, doi:10.1126/science.aal5240 (2017).
- 68 Chandran, P., Satthaporn, S., Robins, A. & Eremin, O. Inflammatory bowel disease: dysfunction of GALT and gut bacterial flora (I). *The surgeon : journal of the Royal Colleges of Surgeons of Edinburgh and Ireland* **1**, 63-75 (2003).
- 69 Wu, S., Morin, P. J., Maouyo, D. & Sears, C. L. *Bacteroides fragilis* enterotoxin induces c-Myc expression and cellular proliferation. *Gastroenterology* **124**, 392-400, doi:10.1053/gast.2003.50047 (2003).
- 70 Yu, T. *et al.* *Fusobacterium nucleatum* Promotes Chemoresistance to Colorectal Cancer by Modulating Autophagy. *Cell* **170**, 548-563 e516, doi:10.1016/j.cell.2017.07.008 (2017).

- 71 Russell, S. L. *et al.* Early life antibiotic-driven changes in microbiota enhance susceptibility to allergic asthma. *EMBO reports* **13**, 440-447, doi:10.1038/embor.2012.32 (2012).
- 72 Arrieta, M. C., Stiemsma, L. T., Amenyogbe, N., Brown, E. M. & Finlay, B. The intestinal microbiome in early life: health and disease. *Frontiers in immunology* **5**, 427, doi:10.3389/fimmu.2014.00427 (2014).

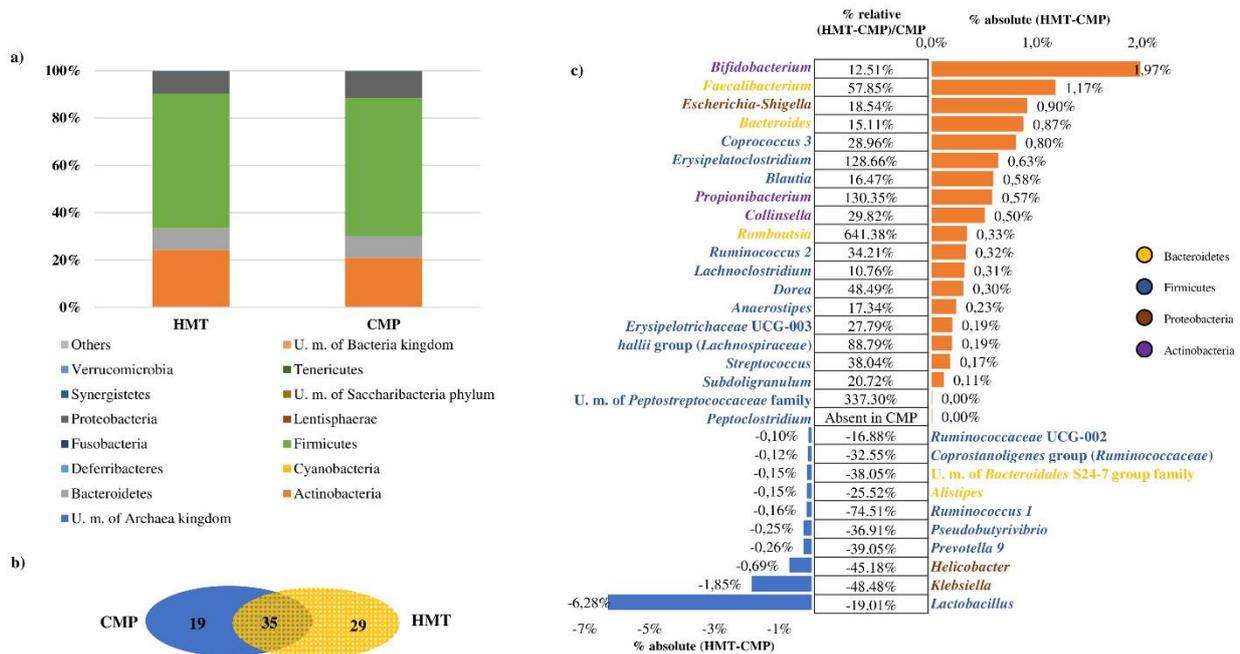


Figure 1. Exploration of the diversity in HMT and CMP groups. Panel a represents a bar plot of the bacterial phyla identified in the 24 analyzed samples. Panel b depicts a Venn diagram illustrating the total, unique and shared number of OTUs predicted for HMT and CMP datasets. Panel C shows the variation in terms of relative abundance of > 0.4%, and showing either an increase of > 10% or decrease of < -16% in CMP data sets as compared to those obtained from HMT samples.

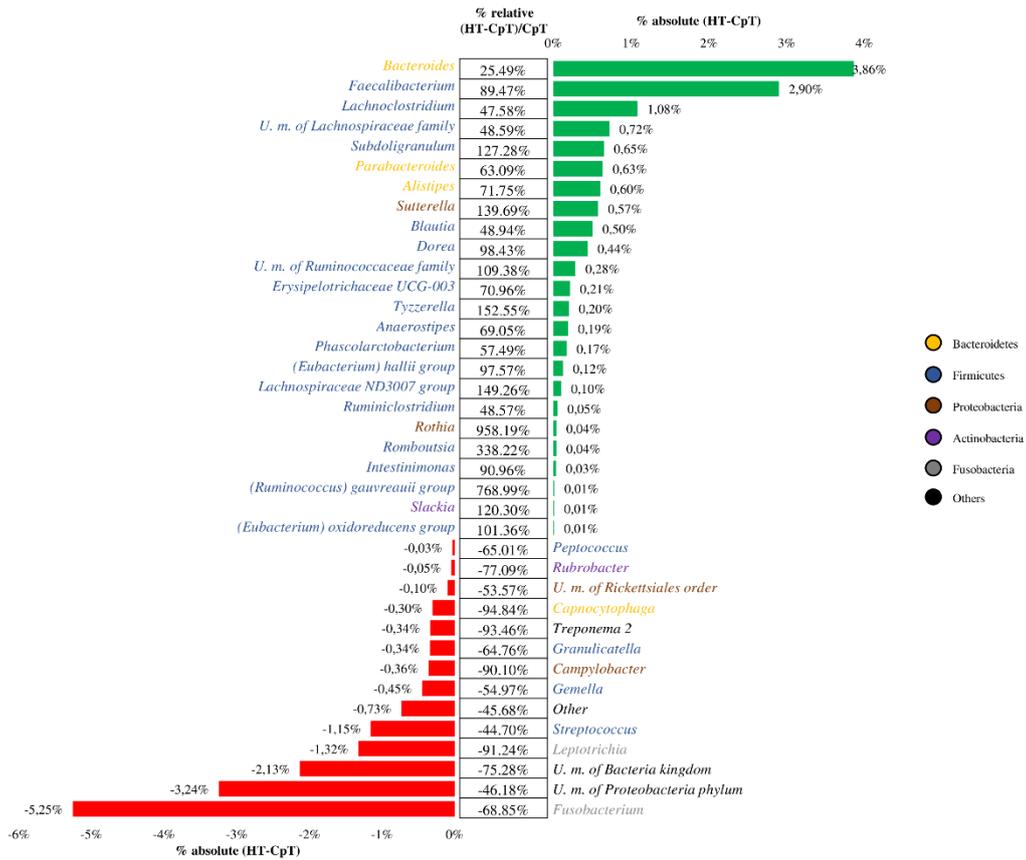


Figure 2. Exploration of the diversity in HT and CpT groups. The image shows the variation in terms of relative abundance of HT and CpT samples. Only taxa with a p-value < 0.05 were shown.

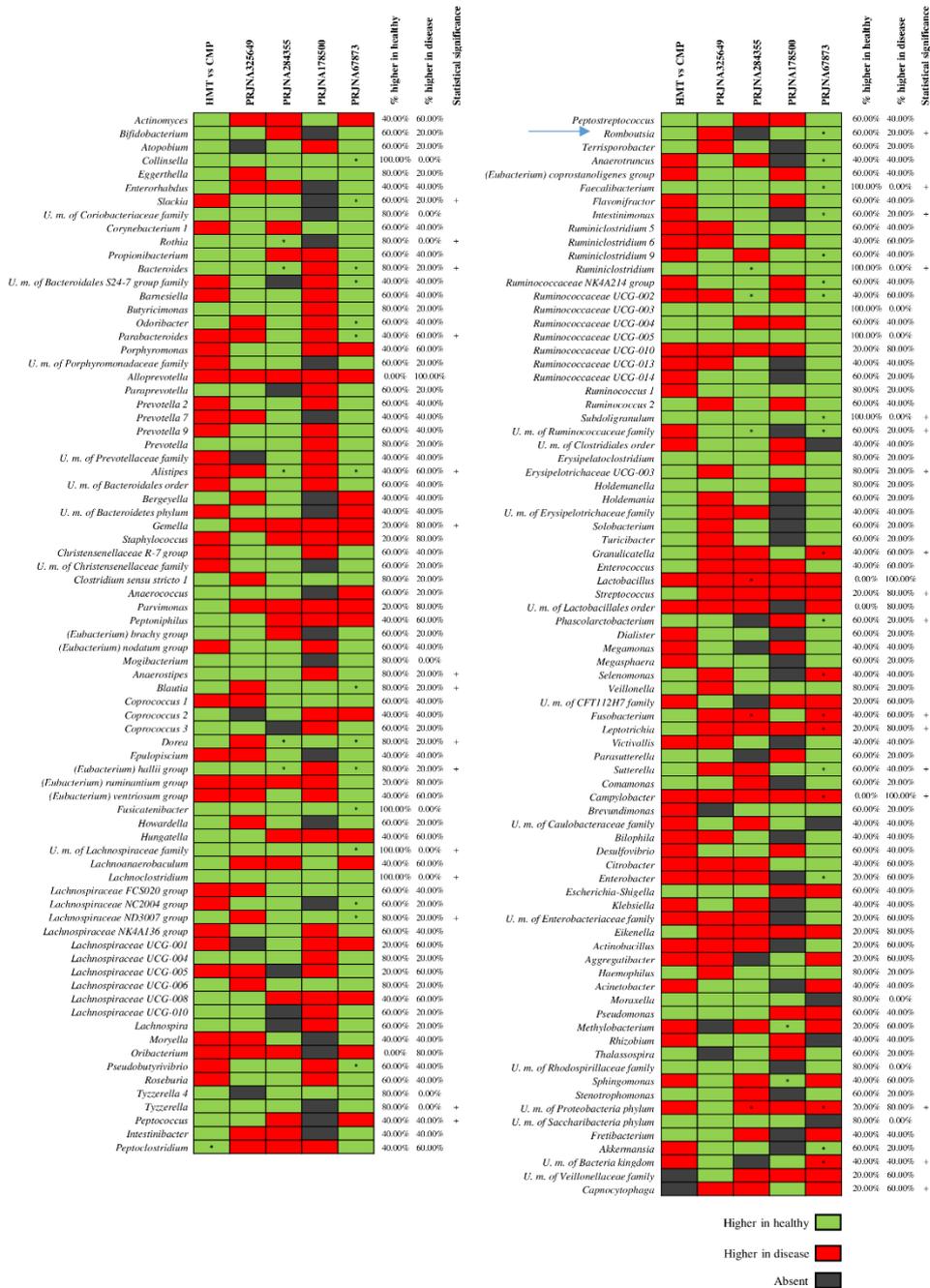


Figure 3. Evaluation of taxa trend in the gut microbiota of HT and CpT samples. Asterisks indicate statistically significant differences between each case-control for all five datasets (P-value<0.05). The heatmap shows the taxonomic trend of each 16S rRNA gene-based metagenomic dataset. Names of identified bacterial taxa are listed on the left, while names of sample groups used are listed at the top. Healthy Marginal Tissue (HMT); Colonial Mucosa with Polyps(CMP); ColoRectal Cancer (CRC); all Healthy Tissue (HT) and CRC-polyps Tissue (CpT). The green color represents high abundance in healthy cases, red color represents high abundance in disease cases and black color represents the absence of the bacterial taxa. Columns show the studies analyzed and rows indicate the bacterial taxonomy at genus level.

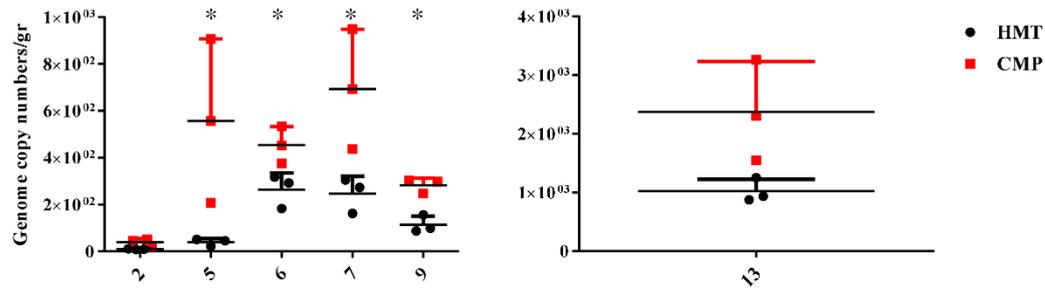


Figure 4. Quantification of *Fusobacterium nucleatum* by qPCR. Results of qPCR are represented in a grouped graph where each sample is symbolized by an individual value plot. The y-axis represents the genome copy number/gr of *Fusobacterium nucleatum* and x-axis shows the patient number. For each patient, HMT biopsies are schematically represented by a black circle and relative CMP portion in red square. Statistical differences were calculated by t-test (*P<0.04).

Chapter 8

Compositional assessment of bacterial communities in probiotic supplements by means of metagenomic techniques

Lugli GA[‡], Mangifesta M[‡], Mancabelli M, Milani C, Turrone F, Viappiani A, van Sinderen D, Ventura M

This chapter was published in:

Int J Food Microbiol. 2019 Jan 21;294:1-9. doi: 10.1016/j.ijfoodmicro.2019.01.011.

Reprinted with permission from Elsevier.

‡ these authors contribute equally

Compositional assessment of bacterial communities in probiotic supplements by means of metagenomic techniques.

Running title: Unveiling genomes of probiotics; a metagenomic approach

Gabriele Andrea Lugli^{1¶}, Marta Mangifesta^{1,2¶}, Leonardo Mancabelli¹, Christian Milani¹, Francesca Turrone^{1,3}, Alice Viappiani², Douwe van Sinderen⁴ and Marco Ventura^{1,3}

Laboratory of Probiogenomics, Department of Chemistry, Life Sciences, and Environmental Sustainability, University of Parma, Parma, Italy¹; GenProbio srl, Parma, Italy²; Microbiome Research Hub, University of Parma, Parma, Italy³; APC Microbiome Institute and School of Microbiology, Bioscience Institute, National University of Ireland, Cork, Ireland⁴

¶ These authors contributed equally

Correspondence. Mailing address for Marco Ventura Laboratory of Probiogenomics, Department of Chemistry, Life Sciences, and Environmental Sustainability, University of Parma, Parco Area delle Scienze 11a, 43124 Parma, Italy. Phone: ++39-521-905666. Fax: ++39-521-905604. E-mail: marco.ventura@unipr.it

Abstract

Health promoting or probiotic bacteria are commonly incorporated into a variety of functional foods and drug formulations, due to their purported ability to confer benefit to host health. Despite the extensive commercial exploitation of probiotic formulations there are still major knowledge gaps regarding the precise molecular mechanism of action and corresponding genetic/genomic properties of probiotic bacteria. In the current study, we describe a metagenomic approach which allows determination of the composition of probiotic supplements through next-generation sequencing analyses based on rRNA-associated sequences to assess bacterial composition of the product combined with a shotgun metagenomics approach directed to decode the genome sequences of the probiotic strains for each product assayed. The here developed approach has been tested for 10 probiotic supplements, revealing inconsistencies between the identified probiotic strains and the declared strains as indicated by the producers. Furthermore, the decoded bacterial genome sequence of *Bifidobacterium animalis* subsp. *lactis* BB-12 from a 1995 frozen dried stock revealed genetic evidence for genome evolution and stability of this microorganism when compared with the re-constructed genome of the identical strain from a probiotic supplement of 2017.

Keywords: Genomics, Metagenomics, Probiotics, ITS, Next Generation Sequencing

Introduction

Probiotic products are defined as live non-pathogenic microorganisms that, when

administered as dietary supplements, confer a health benefit to the host (Passariello et al., 2014). In order for a probiotic microorganism to be provided of the ‘generally recognized as safe’ (GRAS) status, it should not carry any transmissible antibiotic resistance genes, not produce harmful toxins, not evoke any undesirable side effects on the host, nor induce unwanted (e.g. pro-inflammatory) immune responses (Doron and Snyderman, 2015; Sanders et al., 2014). Probiotics are supplemented in various formulations, both conventional pharmaceutical products, which include tablets, capsules and powders, as well as non-conventional foods, e.g., cheeses, yogurts, creams and milk (Govender et al., 2014; Sanders et al., 2014). The currently used bacterial taxa incorporated in such formulations encompass members of the *Bacillus*, *Bifidobacterium*, *Enterococcus*, *Escherichia* and *Lactobacillus* genera (Govender et al., 2014). Such microorganisms had to fulfill strict genetic requirements, such as an unambiguous taxonomical identification and a precise evaluation of their genetic content (Donelli et al., 2013; Fontana et al., 2013; Johnson and Klaenhammer, 2014). Recently, a genetic survey of probiotic products commercialized in the USA highlighted that the determined microbial composition of just four of the 13 products tested were in accordance with the strains declared by the producers (Drago et al., 2010). A similar investigation, aimed at exploring the number of viable bacteria in probiotic products from the European market, unveiled that 10 out of the 24 examined products did not contain the indicated number of viable bacteria, while in four of such products no viable bacteria of the declared species could be retrieved at all (Toscano

et al., 2013). However, the large part of these analyses are based on culture-dependent approaches aimed to isolate and enumerate the microorganisms contained in the probiotic products by the use of selective media (Drago et al., 2010; Goldstein et al., 2014; Toscano et al., 2013). Conversely, metagenomic approaches has been performed to evaluate the bacterial relative abundance within the probiotic products by sequencing of the 16S rRNA sequence using Next-Generation Sequencing (NGS), and thus unveiling bacterial DNA contaminations (Morovic et al., 2016; Patro et al., 2016). Another crucial aspect of probiotic products that has pretty much been ignored is the genetic stability of the incorporated microorganisms. In fact, it is very well known that bacterial genomes evolve through the accumulation of changes (e.g. point mutations, horizontal gene transfer events, reshuffling of their genetic content, gene decay events) (Ventura et al., 2007). Such phenomena may happen at very low frequency under natural conditions but in industrial bacterial strains these events may occur at much higher rates as a consequence of repetitive cultivation under very specific and possible highly selective conditions. In this context, it has been shown that the continuous cultivation of *Bifidobacterium longum* NCC2705 for 1,000 generations in laboratory, lead to a massive reshaping of its chromosome characterized by two large deletions (Lee et al., 2008). In addition, comparative genome analyses of the industrial *Corynebacterium glutamicum* ATCC 13032 strain obtained from two different group revealed intriguing genetic differences represented by a putative prophage island and three insertion sequences (Ventura et al., 2007). In a similar

fashion, the genome instability of *Lactobacillus rhamnosus* GG has been reported, identifying a missing DNA segment from the genome constituting a deletion of 34 and 84 genes from two isolates (Sybesma et al., 2013).

In the current study, we describe a metagenomic approach for the identification of probiotic strains that have been incorporated into probiotic supplements. This methodology, referred to here as Genetic Identity Card (GIC), is able to reveal the microbiota composition of probiotic supplements starting from different probiotic formulations. A combination of NGS analyses that target specific rRNA sequences for compositional analysis and shotgun metagenomics was employed to assess the relative abundance and the overall genetic repertoire of the probiotic strains. Notably, the GIC approach, when applied to 10 probiotic supplements, showed several inconsistencies between identified probiotic strains and strains declared by the producers.

Furthermore, we evaluated the genetic stability of one of the most intensively used probiotic strains, *Bifidobacterium animalis* subsp. *lactis* BB-12, by comparative genomic analyses involving the genome sequences of this strain reconstructed from the shotgun metagenomics data and the chromosome sequences that were decoded from an ancestral stock of the identical strain.

Results and Discussion

1.1 Taxonomical composition of probiotic supplements

The taxonomical composition of 10 probiotic products was investigated through 16S rRNA-based profiling (Fig. 1). Illumina-mediated 16S rRNA microbial profiling produced a total of

750,124 sequencing reads with an average of filtered reads of 60,119 per sample (Table S1). At genus level, the analysis showed that probiotic products B, C, D, E, F, G, I and J reflected the bacterial composition declared by the producer, while probiotic products A and H revealed the presence of additional bacteria. In detail, probiotic products A and H appeared to contain bacteria belonging to the *Lactobacillus* and *Streptococcus* genera, which suggest possible bacterial contamination (Fig. 2).

In order to further validate the possibility of bacterial contamination as identified by the first step of the GIC pipeline based on 16S rRNA sequencing, we performed PCR reactions on products A and H using primers specific for the *Lactobacillus* and *Streptococcus* genera (Moura et al., 2007; Picard et al., 2004). The PCR results of replicates based on three different products packaged individually for each probiotic formulation confirmed the occurrence of *Lactobacillus* and *Streptococcus* (DNA contamination) of the probiotic products A and H, respectively. These results confirmed DNA contamination in probiotic formulation during the production, excluding contaminations during the analysis.

1.2 Prediction of bifidobacterial and fungal composition

In order to characterize the bifidobacterial population present in the probiotic products A, C, D, E, F, G and H, we used the recently developed bifidobacterial ITS profiling protocol (Milani et al., 2014) (Fig. 1). Moreover, the indicated presence of *Saccharomyces cerevisiae* in the product I was verified by amplification and sequencing of fungal ITS as previously described (Bokulich and Mills,

2013) (Fig. 1). Quality filtering of the sequenced ITS produced an average of 7496 and 13,928 high-quality and full-length reads per sample, respectively (Table S1). The bifidobacterial analysis showed that all probiotic products had the microbial composition as declared by the producer except for probiotic product H that contained, in addition to the expected *Bifidobacterium longum* subsp. *longum* taxon, *Bifidobacterium animalis* subsp. *lactis* (Fig. 2). Fungal analysis of probiotic product I revealed the presence of *Saccharomyces cerevisiae*, thereby validating the fungal composition as outlined on the package leaflet.

1.3 Culturomics efforts directed to reconstruct bacterial contents of probiotic products

In order to confirm the obtained *in silico* metagenomic results, we performed a census of the bacterial populations of all assayed probiotics products by the use of a culturomics approach in which we attempted to isolate the bacterial strains from the assessed probiotic products on selective media (Table 1). These analyses were coupled with a taxonomical identification of the isolates through amplification of their 16S rRNA gene sequences followed by amplicon sequencing.

Such tests confirmed the bacterial composition identified with the metagenomics approaches described above with just a few deviations (see below) from what would have been expected based on the associated product declarations. In fact, the culture-dependent approach allowed the identification of isolates that belong to *Lactobacillus* genus in product A, and *Streptococcus* and *Lactobacillus* genera in the commercial probiotic formulation H. Furthermore, we isolated bacterial colonies belonging to *B.*

animalis subsp. *lactis* and *Lb. paracasei* species in products H and J, respectively. These findings therefore confirmed the results obtained by metagenomic analyses, clearly indicating the occurrence of bacterial strains in product A, H and J that were not indicated in the product information.

1.4 Reconstruction of the probiotic genomes

In order to investigate the genomic repertoire of the strains identified in the metagenomic profiling and culturomics efforts, we subjected each probiotic product to shotgun metagenomics analysis (Fig. 1). The sequencing output ranged from 1.49 to 6.36 million of paired-end reads, based on the quantity of DNA loaded in the sequencer estimated by means of the number of probiotic strains declared by the producer of each sample, for a total of 33.89 million. Thus, the generated data from the last step of the GIC pipeline allowed us to perform a metagenomic assembly. Consequently, each reconstructed genome portion from each metagenomic assembly was taxonomically classified at species level. Accordingly, genome reconstruction from the metagenomics data allowed us to identify chromosomal sequences of 23 bacterial/yeast strains, including *B. animalis* subsp. *lactis*, *Bifidobacterium bifidum*, *B. longum* subsp. *longum*, *Enterococcus faecium*, *Lactobacillus acidophilus*, *Lactobacillus delbrueckii* subsp. *bulgaricus*, *Lactobacillus helveticus*, *Lactobacillus paracasei*, *Lactobacillus plantarum*, *Lactobacillus rhamnosus*, *Streptococcus thermophilus* and a yeast strain belonging to the *S. cerevisiae* species (Fig. 3).

The performed genome reconstruction of probiotic products A, B, C, E, F, G and I disclosed the declared bacterial composition in the product

information (Fig. 3). However, the contamination identified in probiotic supplement A through the 16S rRNA profiling and culturomics analyses was not detected in the metagenome analysis, probably due to the low abundance of these strains within the samples. In fact, the number of reads needed for the reconstruction of a bacterial genome must be sufficiently high in order to obtain a minimum genome coverage for the read assembly. Thus, the success in genome reconstruction is correlated with genome length and abundance of the strain in the product. The same issue also affected metagenomic data of probiotic products D and J, where we were unsuccessful in identifying chromosomal sequences of *Lb. rhamnosus* and *Lactobacillus fermentum*, respectively. Notably, both of these two strains were acknowledged to be present by the producers (Fig. 3). Therefore, these results suggest that the latter strains are present at very low abundance with respect to other bacterial strains that were present in sample D and J, or were essentially absent. To validate this notion, qPCR analysis was performed on serial dilutions starting from 1 g of probiotic powder from products D and J in order to identify these apparently missing microbes, as well as on products A, D, G and H so as to compare the abundance of a number of genomes that we were able to re-construct, i.e., *B. animalis* subsp. *lactis*, *B. bifidum* and *S. thermophilus*. The normalized results show the presence of *Lb. rhamnosus* in product D at an abundance that was less than 10⁵ colonies forming unit (CFU), while *Lb. fermentum* was not detected confirming its presumed absence in product J. Furthermore, qPCR analysis of the *Streptococcus* strain present in product H, for which we could not reconstruct the genome, displayed a similar CFU

value to that of *Lb. rhamnosus* in sample D (7.87×10^4 CFU). In contrast, *B. animalis* subsp. *lactis*, *B. bifidum* and *S. thermophilus* present in products A, D and G exhibited higher CFUs within the products, i.e., 1.5×10^6 , 5.89×10^8 and 3.19×10^5 CFU. Thus, as demonstrated by qPCR analyses, where the relative abundance of a strain does not exceed 10^5 CFU, genome reconstruction could not be accomplished.

The GIC pipeline also allowed us to identify various bacterial contaminations or undeclared microbes in probiotic products. In this context, metagenomics analysis of probiotic supplements H and J allowed the reconstruction of *B. animalis* subsp. *lactis* and *Lb. paracasei* genomes, respectively (Fig. 3). Thus, a high level of undeclared or contaminant bacteria was present in these two probiotic supplements, suggesting that proper controls in the supply chain of production should be performed to insure the probiotic strain content of these products.

Furthermore, genome reconstruction of strains present in probiotic products as obtained through the GIC pipeline will permit comparative genome analyses with sequences from other strains belonging to the same species. This application may disclose unique genes of the probiotic strains with respect to publicly available genomes of the same species, perhaps unveiling probiotic characteristics of such microorganisms.

1.5 Genome stability and evolution

During evolution, bacterial genomes are subjected to genetic events shaping genome sequences, i.e., gene duplication, gene loss, chromosomal rearrangements and horizontal gene transfer (Magadam et al., 2013; Soucy et al., 2015). These events, which may occur at different rates, are

responsible for bacterial speciation and are driven by the competitive environment in which these bacterial cells reside (Koonin, 2015). Among bacteria, industrial strains used as started cultures as well as probiotic bacteria appear to be subjected to rapid genome evolution processes, due to the continuous cultivation of these microorganisms (Lee et al., 2008; Machielsen et al., 2010; Sybesma et al., 2013).

Public database contain complete genomes of *B. animalis* subsp. *lactis* BB-12 (present in product C), *B. longum* subsp. *longum* CECT7347 (present in product H) and *Lb. acidophilus* La-14 (occurring in products I and J). The reconstructed genomes as obtained by the GIC pipeline were aligned with previously sequenced genomes of these commercially exploited probiotic strains (Fig. S1). The genome comparisons resulted in the identification of several single nucleotide polymorphisms (SNPs) characterizing the genome evolution of these strains, ranging from five to 418 between sample J and C, respectively (Table 2). Notably, SNPs placed at the very extreme border of the contigs were excluded from our analyses in order to report consistent polymorphisms between sequences. Thus, the evaluation of the SNPs that affect coding sequences highlighted that the published *Lb. acidophilus* La-14 genome sequence when compared to the re-constructed genomes of samples I and J showed just five and one SNPs, respectively (Table 2). Similarly, *B. animalis* subsp. *lactis* BB-12 and *B. longum* subsp. *longum* CECT7347 genomes, when compared with the re-constructed genomes of samples C and H, identified 191 and 174 SNPs, respectively (Table 2). These data disclosed a high frequency of SNPs occurrence in the *Bifidobacterium* species in

respect to *Lb. acidophilus* genomes. Nonetheless, the high frequency in SNPs occurrence may be correlated with the quality of the genome sequences deposited. In fact, while *Lb. acidophilus* was sequenced at a 149-fold coverage (Stahl and Barrangou, 2013), *B. animalis* subsp. *lactis* BB-12 and *B. longum* subsp. *longum* CECT7347 genomes were sequenced at 11-fold coverage with Sanger methodology and 32-fold coverage using a 454 Roche sequencer, respectively (Chenoll et al., 2013; Garrigues et al., 2010).

Further insights related to evolution of probiotic strains was provided by the decoding of genome sequences of a freeze dried stock of *B. animalis* subsp. *lactis* BB-12 (strain named as BB-12-1995), which was produced in 1995 and had since then been maintained in a freezer. The genome sequence of this “ancestral” BB-12 strain was compared with the previous sequenced *B. animalis* subsp. *lactis* BB-12 strain (strain named as BB-12-2010), whose genome was decoded in 2010. In the same fashion, the chromosome sequences belonging to *B. animalis* subsp. *lactis*, which were reconstructed from the metagenomic data of probiotic C (strain named as BB-12-2017), thus constituting a ‘modern’ version of *B. animalis* subsp. *lactis* BB-12 strain was compared to BB-12-1995. Such analysis revealed that strain BB-12-1995 has been moderately modified during this time span by accumulating several point mutations within the genome sequence without showing any deletion (Fig. 4). The genome of BB-12-1995 was assembled at 249-fold coverage, resulting in 161 SNPs when compared with BB-12-2017 genome, of which 148 SNPs reside in open reading frames (ORFs) (Fig. 4 and Table S2). SNPs identified at the very edge of contigs were discarded, obtaining

an average of SNP frequency of 99.4 % between the identified nucleotide variations (Table S2). Notably, the 148 SNPs that affect coding regions, involved 89 ORFs of which 80 ends in nonsynonymous substitutions that change the codon that encodes for a different amino acid (Table S2). Therefore, this analysis clearly demonstrates that the genome of *B. animalis* subsp. *lactis* BB-12 has been subjected to modification in the course of the evolution of the last 22 years.

2. Conclusions

Health promoting or probiotic bacteria are commonly incorporated into a variety of probiotic supplements and dairy products, due to their purported ability to confer benefit to host health (Govender et al., 2014; Sanders et al., 2014). Metagenomic based analyses performed on commercialized probiotic products reveal inconsistency in the bacterial composition, highlighting the need for a consistent strategy to perform quality checks during the production of the probiotic supplements (Morovic et al., 2016; Patro et al., 2016). Thus, we have established a metagenomic approach named Genetic Identity Card, which characterizes the microbial content of probiotic products by means of a metagenomics approach based on rRNA-associated sequences and shotgun metagenomics. Validation of this analysis was performed using 10 different supplements, revealing inconsistency of the determined bacterial content of four probiotic formulations on the one hand and the microbial content as declared by the producers on the other. Genome comparison with previously sequenced strains used in the supplements allowed us to investigate genome stability of these strains. Remarkably, comparative

genomic analyses based on the genome of the probiotic *B. animalis* subsp. *lactis* BB-12 strain collected in 1995 and the current version on the market revealed 162 SNPs that may have been introduced and selected during cultivation of this commercial strain.

Thus, the GIC pipeline is a collection of metagenomic methodologies aimed to identify and characterize the bacterial composition as well as the genetic features, of probiotic products including both single- and multiple strain formulations, without the need of a direct isolation of strains from these probiotic supplements. In the close future, the GIC approach may be implemented by the enumeration of the bacterial cells using a flow cytometric approach that will allow a normalization of the reads for each bacterial strain determined by metagenomics analyses (Vandeputte et al., 2017).

4. Materials and Methods

4.1 Probiotics included in this study

We randomly selected 10 powder-based probiotic supplements in order to analyze products containing either a single-strain or multi-strains, which were retrieved from supermarkets and processed through the GIC pipeline before the expiration date indicated in the products' information. We recoded the commercial name of each probiotic product to keep the anonymity of the product. A portion of the lyophilizate, consisting of 20 mg of product, was obtained from each probiotic. DNA was extracted from each probiotic using DNeasy PowerFood Microbial Kit (Qiagen Ltd, Strasse, Germany) following the manufacturer's instructions (Qiagen Ltd, Strasse, Germany).

4.2 Isolation of probiotic from probiotic supplements

One milliliter of each sample was mixed with nine mL sterile Phosphate-buffered saline (PBS) to make an initial dilution. Serial dilutions were made for each sample and one mL of the appropriate dilution was mixed with different media agar plates. For the isolation of probiotics, different media were assessed, including de Man, Rogosa and Sharpe (MRS) agar (Difco Laboratories, USA) supplemented with 0.05 % (wt/vol) L-cysteine hydrochloride and 50 µg/ml mupirocin (Delchimica, Italy), M17 agar (Oxoid Ltd., UK) with modification (2 % lactose after autoclaving) and YPD (Sigma-Aldrich, St Louis MO). The M17 agar plates were incubated at 37 °C for 48 h in aerobic condition, MRS agar plates were incubated in anaerobic atmosphere (2.99 % H₂, 17.01 % CO₂, and 80% N₂) in a chamber (Concept 400; Ruskin) at 37 °C for 48 h. YPD agar plates were incubated at 30°C in aerobic condition. Colonies with distinct morphological differences (based on color, shape, size, rough or smooth surface) were selected and then purified using another agar plate of the same culture medium. DNA was extracted from each isolate through rapid mechanic cell lysis as described previously (Turrone et al., 2009). Isolates from each sample were stored at -80°C in the presence of glycerol (30 %, v/v).

4.3 Genus-specific PCR and taxonomic identification of probiotic isolates

For all bacterial species tested we used a specific couple of primers. All PCR amplifications were performed from 2 µL of a genomic DNA preparation at 10 ng/µL which was transferred directly to a 12.5 µL PCR mixture containing Platinum PCR SuperMix 1X (Invitrogen, USA)

and 100 pM of each oligo. PCR reactions were performed on a Verity Thermocycler (Applied Biosystems, USA). Electrophoretic profiles were visualized by SYBR Safe DNA gel stain (Invitrogen). For *Streptococcus*-specific PCR we used primers Str1 (5'-GTACAGTTGCTTCAGGACGTATC-3') and Str2 (5'-ACGTTTCGATTTTCATCACGTTG-3') (Picard et al., 2004). For *Lactobacillus*-specific and *Bifidobacterium*-genus PCR we used the couples of primers Lab 0677F (5'-CTCCATGTGTAGCGGTG-3'), Lact71R (5'-TCAAACCTAAACAAAGTTTC-3') and BIF-specific (5'-GGTGTGAAAGTCCATCGCT-3'), 23S_bif (5'-GTCTGCCAAGGCATCCACCA-3') (Moura et al., 2007; Turroni et al., 2009) respectively. For *S. cerevisiae* detection, we used SC-FW (5'-GGACTCTGGACATGCAAGAT-3') and SC-RV (5'-ATACCCTTCTTAACACCTGGC-3') (Salinas et al., 2009).

The identification of each selected isolate was then performed by PCR amplification of a portion of the 16S rRNA gene using primers P0 (5'-GAAGAGTTTGATCCTGGCTCAG-3') and P6 (5'-CTACGGCTACCTTGTACGA-3'). Each 25 µl PCR reaction contained approximately 30 ng of genomic DNA, Platinum PCR SuperMix 1X (Invitrogen, USA) and 100 pM of each oligo. PCR reactions were performed on a Verity Thermocycler (Applied Biosystems, USA). Electrophoretic profiles were visualized by SYBR Safe DNA gel stain (Invitrogen). PCR product purification was performed using the NucleoSpin Gel and PCR Clean-up kit (Macherey-Nagel) following manufacturer instructions. Each 16S rRNA gene generated from individual colonies

originating from each probiotic sample was sequenced and it was then subjected to a BLAST search against the GenBank database.

4.4 Evaluation of bacterial cell density by qPCR

Real-Time PCR reactions were performed on MicroAmp optical plates sealed with MicroAmp optical caps (Applied Biosystems, Foster City, 2 CA) and amplifications were carried out on a CFX96 system (BioRad, CA, USA). Primers used in this study are Ban_Lac_fw (5'-CGAAGAGGACCATGACTTCC-3') and Ban_Lac_rv (5'-CGGTGAGCTCCACATTGA-3') to determine *B. animalis* subsp. *lactis*, Bbif_0282Fw (5'-GCGAACAATGATGGCACCTA-3') and Bbif_0282Rv (5'-GTCGAACACCACGACGATGT-3') to enumerate *B. bifidum* (Turroni et al., 2016), Rha I (5'-CAGACTGAAAGTCTGACGG-3') and Rha II (5'-GCGATGCGAATTTCTATTATT-3') to determinate *Lb. rhamnosus* (Byun and Yoon, 2003) and Str1 (5'-GTACAGTTGCTTCAGGACGTATC-3') and Str2 (5'-ACGTTTCGATTTTCATCACGTTG-3') for *Streptococcus* genus-specific identification (Picard et al., 2004). Each PCR reaction mix contained 7.5 µl 2x SYBR SuperMix Green (BioRad, CA, USA), 5 µl of DNA dilution, forward and reverse primers at 10 pM, and nuclease-free water was added to obtain a final volume of 15 µl. PCR amplicons were detected with SYBR Green fluorescent dye and amplified according to the following protocol: one cycle of 95°C for 2 minutes, 40 cycles of 95°C for 5 s and 60°C for 30 s. Amplicon specificity was judged based on the dissociation curve of PCR end products by

increasing the temperature at a rate of 1°C every 30 s from 60 to 95 °C. Negative controls composed by water were included for each primer set in each run.

4.5 16S rRNA gene amplification

Partial 16S rRNA gene sequences were amplified from extracted DNA using primer pair Probio_Uni and/Probio_Rev, which targets the V3 region of the 16S rRNA gene sequence (Milani et al., 2013). Illumina adapter overhang nucleotide sequences were then added to the partial 16S rRNA gene-specific amplicons, which in turn were further processed by employing the 16S Metagenomic Sequencing Library Preparation Protocol (Part no. 15044223 Rev. B—Illumina; see also below). Amplifications were carried out using a Verity Thermocycler (Applied Biosystems). The integrity of the PCR amplicons was analyzed by electrophoresis on a 2200 TapeStation Instrument (Agilent Technologies, USA).

4.6 MiSeq sequencing of 16S rRNA gene-based amplicons

PCR products obtained following amplification of part of the 16S rRNA gene sequences were purified by a magnetic purification step involving Agencourt AMPure XP DNA purification beads (Beckman Coulter Genomics GmbH, Bernried, Germany) in order to remove primer dimers. The DNA concentration of the amplified sequence library was estimated through fluorimetric Qubit quantification system (Life Technologies). Amplicons were diluted to 4 nM and 5 µl of each diluted DNA amplicon sample was mixed to prepare the pooled final library. Sequencing was performed using an Illumina MiSeq sequencer with MiSeq Reagent Kit v3 chemicals.

4.7 Analysis of 16S rRNA microbial profiling datasets

The fastq files were processed using QIIME (Caporaso et al., 2010) as previously described (Milani et al., 2013). Paired-end reads were merged, and quality control implementation allowed the retention of sequences with a length between 140 and 400 bp, mean sequence quality score >25 and with truncation of a sequence at the first base if a low quality within a rolling 10-bp window was found. Sequences with mismatched forward and/or reverse primers were omitted. 16S rRNA operational taxonomic units (OTUs) were defined at ≥ 99 % sequence homology using uclust (Edgar, 2010). All reads were classified to the lowest possible taxonomic rank using QIIME (Caporaso et al., 2010) and a reference dataset from the SILVA database v. 128 (Quast et al., 2013). The classification at species level was confirmed through manual alignment of the OTUs using online `blastn` tool (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>).

4.8 ITS microbial and fungal profiling analysis

Profiling of known bifidobacterial species was performed using the primer pair Probio_bif_uni/Probio_bif_rev, an improved bifidobacterial ITS database encompassing all publicly available bifidobacterial genomes and a custom bioinformatics script, as described previously (Milani et al., 2014). Moreover, prediction of the fungal composition at species level was completed using the primer pair BITS/B58S3 and based on the UNITE database (Koljalg et al., 2005).

4.9 Shotgun metagenomics

DNA library preparation was performed using the Nextera XT DNA sample preparation kit (Illumina, San Diego, CA) according to the manufacturer's

instructions. In brief, 1 ng input DNA from each sample was used for library preparation. The isolated DNA underwent fragmentation, adapter ligation and amplification. The ready-to-go libraries were pooled equimolarly, denaturated and diluted to a sequencing concentration of 1.8 pM. Sequencing was performed on NextSeq 550 instrument (Illumina, San Diego, CA), according to the manufacturer's instructions, using the 2x150 bp High Output sequencing kit, and spike-in of 1 % PhiX control library.

4.10 Genome reconstruction of the probiotic strains

The generated paired fastq files were used as input for SPAdes assembler v3.9 (Bankevich et al., 2012), for the novo metagenomic assemblies using default parameters coupled with k-mer sizes of 21, 33, 55 and 77, and enabling the metagenomic flag option (-meta). ORFs of each assembled contig were predicted with Prodigal (Hyatt et al., 2010) and annotated by means of the software MEGAnnotator (Lugli et al., 2016). The annotated contigs were taxonomically classified based on the gene hit obtained through the NCBI RefSeq databases using the script Contig Classifier (CoCla) (Lugli et al., 2017). Species-specific contigs were subdivided in different files, obtaining the reconstructed genomes of each bacterial strain included in each probiotic sample.

4.11 Genome analyses and sequence comparisons

Genome sequence alignments were performed using MAUVE software (Darling et al., 2010),

while the SNPs count was obtained through read mapping by means of Burrows-Wheeler Aligner v0.7.15 (Li and Durbin, 2009).

4.12 Availability of data and materials

The 16S rRNA- and ITS-profiling data sequenced in this study were deposited in the Sequence Read Archive (SRA) database under accession number PRJNA422025. Shotgun metagenomics data are accessible through SRA study accession number PRJNA422026.

Acknowledgements

This work was funded by the EU Joint Programming Initiative – A Healthy Diet for a Healthy Life (JPI HDHL, <http://www.healthydietforhealthylife.eu/>) to DvS (in conjunction with Science Foundation Ireland [SFI], Grant number 15/JP-HDHL/3280) and to MV (in conjunction with MIUR, Italy). We thank GenProbio srl for financial support of the Laboratory of Probiogenomics. DvS is a member of The APC Microbiome Institute funded by Science Foundation Ireland (SFI), through the Irish Government's National Development Plan (Grant number SFI/12/RC/2273). Part of this research is conducted using the High Performance Computing (HPC) facility of the University of Parma.

Competing interests

The authors declare that they have no competing interests.

References

- Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M., Nikolenko, S.I., Pham, S., Prjibelski, A.D., Pyshkin, A.V., Sirotkin, A.V., Vyahhi, N., Tesler, G., Alekseyev, M.A., Pevzner, P.A., 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 19, 455-477.
- Bokulich, N.A., Mills, D.A., 2013. Improved selection of internal transcribed spacer-specific primers enables quantitative, ultra-high-throughput profiling of fungal communities. *Appl Environ Microbiol* 79, 2519-2526.
- Byun, J.R., Yoon, Y.H., 2003. Identification and Comparison of the Nucleotide Sequence of 16S-23S rRNA Gene Intergenic Small SR(Spacer Region) of *Lactobacillus rhamnosus* ATCC 53103 with Those of *L. casei*, *L. acidophilus* and *L. helveticu*. *Asian-Aust. J. Anim. Sci.* Vol 16, 1816-1821.
- Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., Fierer, N., Pena, A.G., Goodrich, J.K., Gordon, J.I., Huttley, G.A., Kelley, S.T., Knights, D., Koenig, J.E., Ley, R.E., Lozupone, C.A., McDonald, D., Muegge, B.D., Pirrung, M., Reeder, J., Sevinsky, J.R., Turnbaugh, P.J., Walters, W.A., Widmann, J., Yatsunenko, T., Zaneveld, J., Knight, R., 2010. QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* 7, 335-336.
- Chenoll, E., Codoñer, F., Silva, A., Ibáñez, A., Martínez-Blanch, J., Bollati-Fogolín, M., Crispo, M., Ramírez, S., Sanz, Y., Ramón, D., Genovés, S., 2013. Genomic Sequence and Pre-Clinical Safety Assessment of *Bifidobacterium longum* CECT 7347, a Probiotic able to Reduce the Toxicity and Inflammatory Potential of Gliadin-Derived Peptides. *Journal of Probiotics & Health*.
- Darling, A.E., Mau, B., Perna, N.T., 2010. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PLoS One* 5, e11147.
- Donelli, G., Vuotto, C., Mastromarino, P., 2013. Phenotyping and genotyping are both essential to identify and classify a probiotic microorganism. *Microb Ecol Health Dis* 24.
- Doron, S., Snyderman, D.R., 2015. Risk and safety of probiotics. *Clin Infect Dis* 60 Suppl 2, S129-134.
- Drago, L., Rodighiero, V., Celeste, T., Rovetto, L., De Vecchi, E., 2010. Microbiological evaluation of commercial probiotic products available in the USA in 2009. *J Chemother* 22, 373-377.
- Edgar, R.C., 2010. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26, 2460-2461.
- Fontana, L., Bermudez-Brito, M., Plaza-Diaz, J., Munoz-Quezada, S., Gil, A., 2013. Sources, isolation, characterisation and evaluation of probiotics. *Br J Nutr* 109 Suppl 2, S35-50.
- Garrigues, C., Johansen, E., Pedersen, M.B., 2010. Complete genome sequence of *Bifidobacterium animalis* subsp. *lactis* BB-12, a widely consumed probiotic strain. *J Bacteriol* 192, 2467-2468.
- Goldstein, E.J., Citron, D.M., Claros, M.C., Tyrrell, K.L., 2014. Bacterial counts from five over-the-counter probiotics: are you getting what you paid for? *Anaerobe* 25, 1-4.
- Govender, M., Choonara, Y.E., Kumar, P., du Toit, L.C., van Vuuren, S., Pillay, V., 2014. A review of the advancements in probiotic delivery: Conventional vs. non-conventional formulations for intestinal flora supplementation. *AAPS PharmSciTech* 15, 29-43.
- Hyatt, D., Chen, G.L., Locascio, P.F., Land, M.L., Larimer, F.W., Hauser, L.J., 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 11, 119.
- Johnson, B.R., Klaenhammer, T.R., 2014. Impact of genomics on the field of probiotic research: historical perspectives to modern paradigms. *Antonie Van Leeuwenhoek* 106, 141-156.
- Koljalg, U., Larsson, K.H., Abarenkov, K., Nilsson, R.H., Alexander, I.J., Eberhardt, U., Erland, S., Hoiland, K., Kjøller, R., Larsson, E., Pennanen, T., Sen, R., Taylor, A.F., Tedersoo, L., Vralstad, T., Ursing, B.M., 2005. UNITE: a database providing web-based methods for the molecular identification of ectomycorrhizal fungi. *New Phytol* 166, 1063-1068.
- Koonin, E.V., 2015. The Turbulent Network Dynamics of Microbial Evolution and the Statistical Tree of Life. *J Mol Evol* 80, 244-250.
- Lee, J.H., Karamychev, V.N., Kozyavkin, S.A., Mills, D., Pavlov, A.R., Pavlova, N.V., Polouchine, N.N., Richardson, P.M., Shakhova, V.V., Slesarev, A.I., Weimer, B., O'Sullivan, D.J., 2008. Comparative genomic analysis of the gut bacterium *Bifidobacterium longum* reveals loci susceptible to deletion during pure culture growth. *BMC Genomics* 9, 247.
- Li, H., Durbin, R., 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754-1760.
- Lugli, G.A., Milani, C., Mancabelli, L., Turroni, F., Ferrario, C., Duranti, S., van Sinderen, D., Ventura, M., 2017. Ancient bacteria of the Otzi's microbiome: a genomic tale from the Copper Age. *Microbiome* 5, 5.

Lugli, G.A., Milani, C., Mancabelli, L., van Sinderen, D., Ventura, M., 2016. MEGAnnotator: a user-friendly pipeline for microbial genomes assembly and annotation. *FEMS Microbiol Lett* 363.

Machielsen, R., van Alen-Boerriqter, I.J., Koole, L.A., Bongers, R.S., Kleerebezem, M., Van Hylckama Vlieg, J.E., 2010. Indigenous and environmental modulation of frequencies of mutation in *Lactobacillus plantarum*. *Appl Environ Microbiol* 76, 1587-1595.

Magadum, S., Banerjee, U., Murugan, P., Gangapur, D., Ravikesavan, R., 2013. Gene duplication as a major force in evolution. *J Genet* 92, 155-161.

Milani, C., Hevia, A., Foroni, E., Duranti, S., Turrone, F., Lugli, G.A., Sanchez, B., Martin, R., Gueimonde, M., van Sinderen, D., Margolles, A., Ventura, M., 2013. Assessing the fecal microbiota: an optimized ion torrent 16S rRNA gene-based analysis protocol. *PLoS One* 8, e68739.

Milani, C., Lugli, G.A., Turrone, F., Mancabelli, L., Duranti, S., Viappiani, A., Mangifesta, M., Segata, N., van Sinderen, D., Ventura, M., 2014. Evaluation of bifidobacterial community composition in the human gut by means of a targeted amplicon sequencing (ITS) protocol. *FEMS Microbiol Ecol* 90, 493-503.

Morovic, W., Hibberd, A.A., Zabel, B., Barrangou, R., Stahl, B., 2016. Genotyping by PCR and High-Throughput Sequencing of Commercial Probiotic Products Reveals Composition Biases. *Front Microbiol* 7, 1747.

Moura, P., Simoes, F., Girio, F., Loureiro-Dias, M.C., Esteves, M.P., 2007. PCR monitoring of *Lactobacillus* and *Bifidobacterium* dynamics in fermentations by piglet intestinal microbiota. *J Basic Microbiol* 47, 148-157.

Passariello, A., Agricole, P., Malfertheiner, P., 2014. A critical appraisal of probiotics (as drugs or food supplements) in gastrointestinal diseases. *Curr Med Res Opin* 30, 1055-1064.

Patro, J.N., Ramachandran, P., Barnaba, T., Mammel, M.K., Lewis, J.L., Elkins, C.A., 2016. Culture-Independent Metagenomic Surveillance of Commercially Available Probiotics with High-Throughput Next-Generation Sequencing. *mSphere* 1.

Picard, F.J., Ke, D., Boudreau, D.K., Boissinot, M., Huletsky, A., Richard, D., Ouellette, M., Roy, P.H., Bergeron, M.G., 2004. Use of tuf sequences for genus-specific PCR detection and phylogenetic analysis of 28 streptococcal species. *J Clin Microbiol* 42, 3686-3695.

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., Glockner, F.O., 2013. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* 41, D590-596.

Salinas, F., Garrido, D., Ganga, A., Veliz, G., Martinez, C., 2009. Taqman real-time PCR for the detection and enumeration of *Saccharomyces cerevisiae* in wine. *Food Microbiol* 26, 328-332.

Sanders, M.E., Klaenhammer, T.R., Ouweland, A.C., Pot, B., Johansen, E., Heimbach, J.T., Marco, M.L., Tennila, J., Ross, R.P., Franz, C., Page, N., Pridmore, R.D., Leyer, G., Salminen, S., Charbonneau, D., Call, E., Lenoir-Wijnkoop, I., 2014. Effects of genetic, processing, or product formulation changes on efficacy and safety of probiotics. *Ann N Y Acad Sci* 1309, 1-18.

Soucy, S.M., Huang, J., Gogarten, J.P., 2015. Horizontal gene transfer: building the web of life. *Nat Rev Genet* 16, 472-482.

Stahl, B., Barrangou, R., 2013. Complete Genome Sequence of Probiotic Strain *Lactobacillus acidophilus* La-14. *Genome Announc* 1.

Sybesma, W., Molenaar, D., van, I.W., Venema, K., Kort, R., 2013. Genome instability in *Lactobacillus rhamnosus* GG. *Appl Environ Microbiol* 79, 2233-2239.

Toscano, M., de Vecchi, E., Rodighiero, V., Drago, L., 2013. Microbiological and genetic identification of some probiotics proposed for medical use in 2011. *J Chemother* 25, 156-161.

Turrone, F., Foroni, E., Pizzetti, P., Giubellini, V., Ribbera, A., Merusi, P., Cagnasso, P., Bizzarri, B., de'Angelis, G.L., Shanahan, F., van Sinderen, D., Ventura, M., 2009. Exploring the diversity of the bifidobacterial population in the human intestinal tract. *Appl Environ Microbiol* 75, 1534-1545.

Turrone, F., Milani, C., Duranti, S., Mancabelli, L., Mangifesta, M., Viappiani, A., Lugli, G.A., Ferrario, C., Gioiosa, L., Ferrarini, A., Li, J., Palanza, P., Delledonne, M., van Sinderen, D., Ventura, M., 2016. Deciphering bifidobacterial-mediated metabolic interactions and their impact on gut microbiota by a multi-omics approach. *ISME J* 10, 1656-1668.

Vandeputte, D., Kathagen, G., D'Hoe, K., Vieira-Silva, S., Valles-Colomer, M., Sabino, J., Wang, J., Tito, R.Y., De Commer, L., Darzi, Y., Vermeire, S., Falony, G., Raes, J., 2017. Quantitative microbiome profiling links gut community variation to microbial load. *Nature* 551, 507-511.

Ventura, M., Canchaya, C., Tauch, A., Chandra, G., Fitzgerald, G.F., Chater, K.F., van Sinderen, D., 2007. Genomics of Actinobacteria: tracing the evolutionary history of an ancient phylum. *Microbiol Mol Biol Rev* 71, 495-548.

Table 1. Culturomics analyses

Probiotics	Medium	Anaerobic condition	Aerobic condition	Identified 16S ribosomal RNA sequences
A	MRS agar	37°C for 48h	-	<i>B. animalis</i> subsp. <i>lactis</i>
	Brain Heart Infusion	-	37°C for 48h	<i>E. faecium</i>
	MRS agar	-	37°C for 48h	<i>Lactobacillus</i> spp.
B	MRS agar	-	37°C for 48h	<i>Lb. paracasei</i>
C	MRS agar	37°C for 48h	-	<i>B. animalis</i> subsp. <i>lactis</i>
	MRS agar	37°C for 48h	-	<i>Lb. acidophilus</i>
	Brain Heart Infusion	-	37°C for 48h	<i>Streptococcus</i> spp.
D	MRS agar	37°C for 48h	-	<i>B. bifidum</i>
	MRS agar	37°C for 48h	-	<i>Lb. acidophilus</i>
E	MRS agar	37°C for 48h	-	<i>B. bifidum</i>
F	MRS agar	37°C for 48h	-	<i>B. longum</i>
	MRS agar	-	37°C for 48h	<i>Lb. helveticus</i>
G	MRS agar	37°C for 48h	-	<i>B. bifidum</i>
	MRS agar	37°C for 48h	-	<i>Lb. acidophilus</i>
	MRS agar	37°C for 48h	-	<i>Lb. rhamnosus</i>
	Brain Heart Infusion	-	37°C for 48h	<i>S. thermophilus</i>
	MRS agar	37°C for 48h	-	<i>Lb. delbruekii</i>
H	MRS agar	-	37°C for 48h	<i>Lactobacillus</i> spp.
	MRS agar	37°C for 48h	-	<i>B. animalis</i> subsp. <i>lactis</i>
	MRS agar	37°C for 48h	-	<i>B. longum</i> subsp. <i>longum</i>
I	Brain Heart Infusion	-	37°C for 48h	<i>E. faecium</i>
	YPD agar	-	30°C for 48h	<i>S. cerevisiae</i>
	MRS agar	37°C for 48h	-	<i>Lb. acidophilus</i>
J	MRS agar	37°C for 48h	-	<i>Lb. acidophilus</i>
	MRS agar	-	37°C for 48h	<i>Lb. paracasei</i>
	MRS agar	37°C for 48h	-	<i>Lb. plantarum</i>
	MRS agar	37°C for 48h	-	<i>Lb. rhamnosus</i>

Table 2. SNPs between re-constructed and reference genomes

Probiotics	Re-constructed genomes	Reference strain	SNPs	Intergenic SNPs	Coding SNPs
C	<i>B. animalis</i> subsp. <i>lactis</i>	BB-12	418	227	191
H	<i>B. longum</i> subsp. <i>longum</i>	CECT7347	184	10	174
I	<i>Lb. acidophilus</i>	La-14	9	4	5
J	<i>Lb. acidophilus</i>	La-14	5	4	1

0

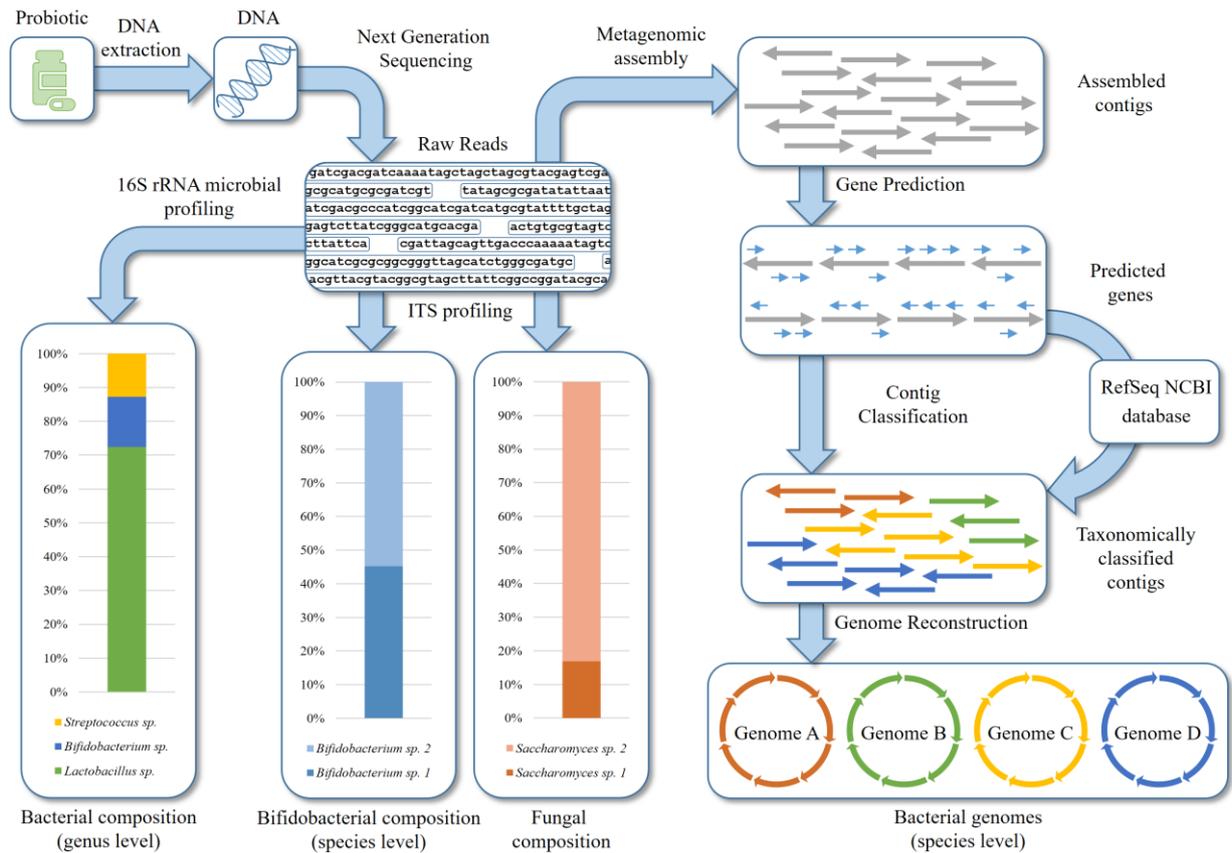


Figure 1. Schematic representation of the metagenomic approach named Genetic Identity Card. The methodology involves three analyses based on NGS, i.e., from left to right, 16S rRNA microbial profiling, ITS profiling and assembly of shotgun metagenomic data.

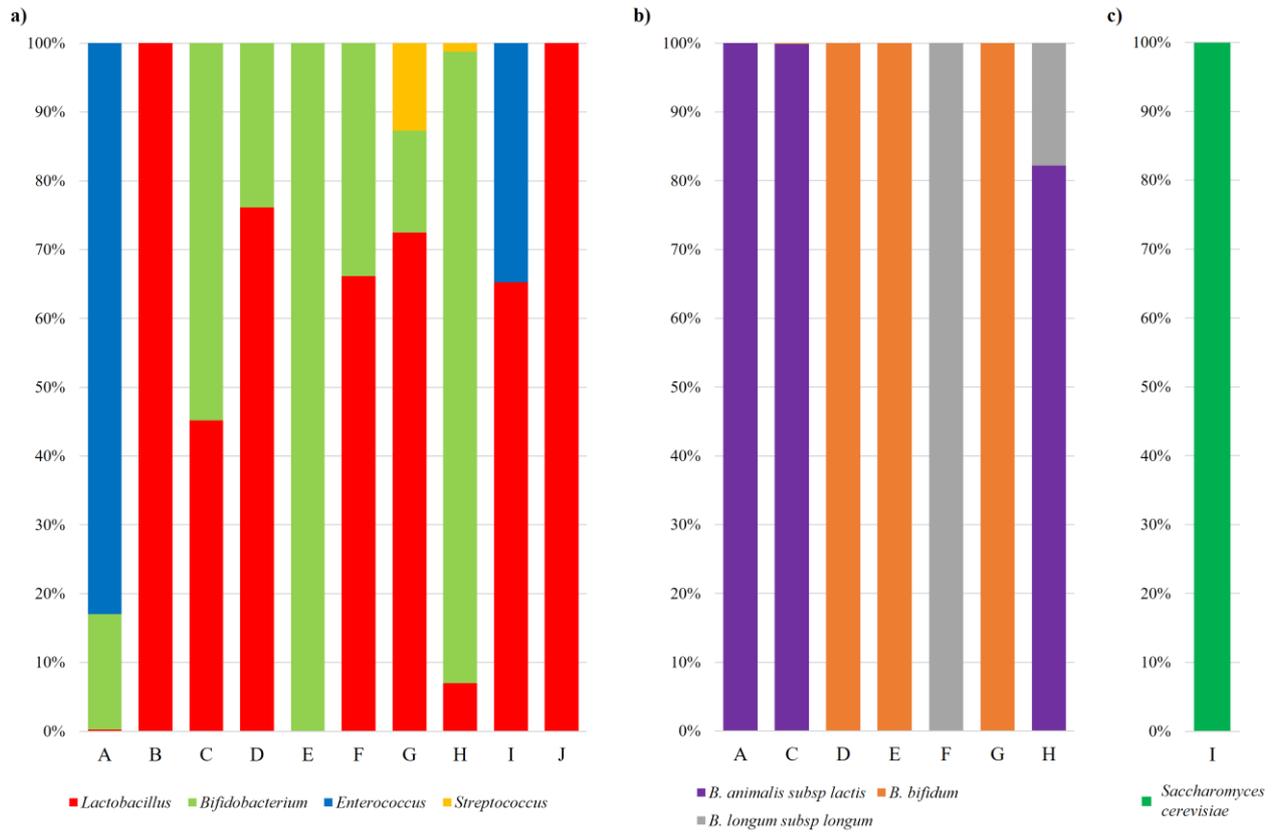


Figure 2. The 16S rRNA gene and ITS profiling of probiotic products. Panel a shows a bar plot of each probiotic sample representing the OTUs obtained through 16S rRNA gene profiling. Panel b displays a bar plot for each probiotic sample that include bifidobacterial strains in the formulations representing the results achieved through ITS bifidobacterial profiling. Panel c exhibits the fungal ITS profiling of sample I.

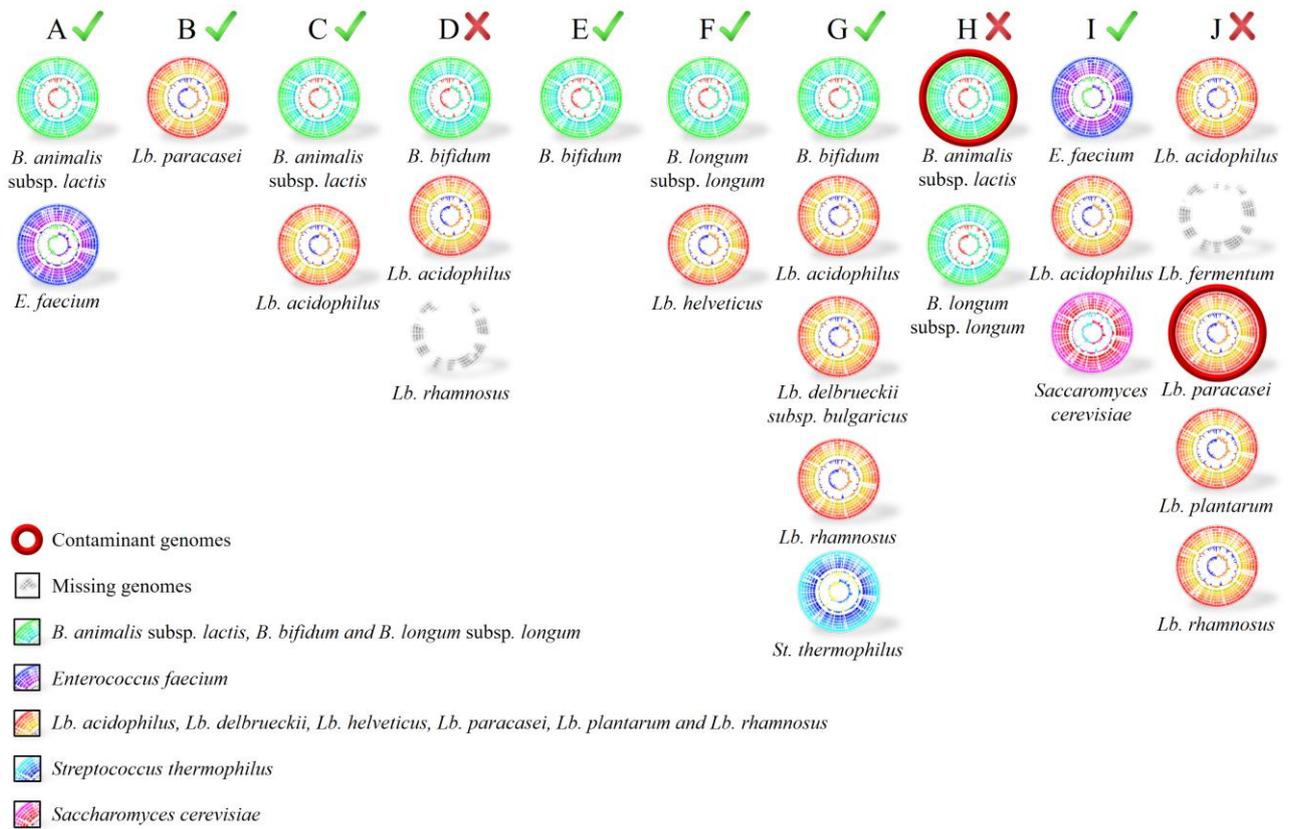


Figure 3. Re-constructed genomes from probiotic products by means of the GIC metagenomic pipeline. Each re-constructed genome is represented by a genetic map arranged in vertical position by means of the 10 probiotic supplements. Apparently missing genomes were illustrated as fading genomes, while undeclared bacteria genomes are highlighted with red circles. Green letter v indicates probiotic products that showed the re-construction of the strains declared by the producer, while red letter x reveal products with undeclared or contaminant strains.

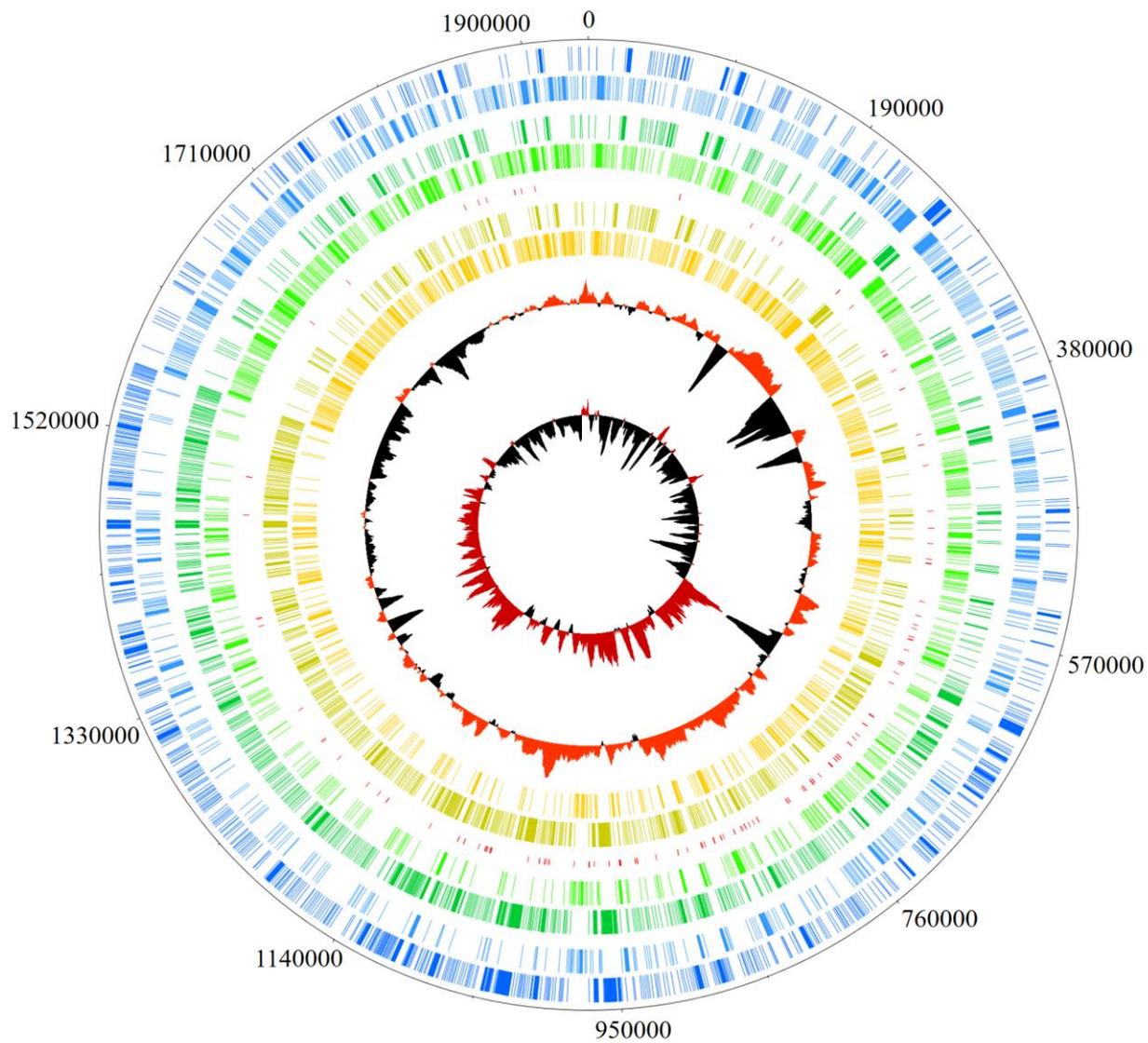


Figure 4. Genome evolution of strain *B. animalis* subsp. *lactis* BB-12. Circular genome atlas of *B. animalis* subsp. *lactis* BB-12-1995, BB-12-2010 and BB-12-2017 with SNP positions between genomes. Starting from the external circle, BB-12-2010 gene locations (blue circles), BB-12-1995 gene positions (green circles), SNPs spot between BB-12-1995 and BB-12-2017 (red lines), BB-12-2017 gene locations (yellow circles), BB-12 G+C% deviation (black/orange) and BB-12 GC skew (G-C/G+C) (black/red) are reported.

Chapter 9

Phylotype-Level Profiling of Lactobacilli in Highly Complex Environments by Means of an Internal Transcribed Spacer-Based Metagenomic Approach

Milani C[‡], Duranti S[‡], Mangifesta M[‡], Lugli GA, Turrone F, Mancabelli L,
Viappiani A, Anzalone R, Alessandri G, Ossiprandi MC, van Sinderen D,
Ventura M

This chapter was published in:

Appl Environ Microbiol. 2018 Jul 2;84(14). pii: e00706-18. doi: 10.1128/AEM.00706-18

Reprinted with permission from American Society for Microbiology.

[‡]these authors contribute equally

Phylotype-level profiling of lactobacilli in highly complex environments by means of an ITS-based metagenomic approach.

Key words: *Lactobacillus*, microbiota, ITS, Next Generation Sequencing, Illumina.

Christian Milani^{1*}, Sabrina Duranti^{1*}, Marta Mangifesta^{1*}, Gabriele Andrea Lugli¹, Francesca Turroni^{1,2}, Leonardo Mancabelli¹, Alice Viappiani³, Rosaria Anzalone¹, Giulia Alessandri¹, Maria Cristina Ossiprandi⁴, Douwe van Sinderen⁵ and Marco Ventura^{1,2}

*These authors contributed equally.

Laboratory of Probiogenomics, Department of Life Sciences, University of Parma, Italy¹;
Microbiome Research Hub, University of Parma, Parma, Italy² GenProbio srl, Parma, Italy³;
Department of Veterinary Medical Science, University of Parma, Parma, Italy⁴; APC Microbiome
Ireland and School of Microbiology, Biosciences building, University College Cork, Ireland⁵

Corresponding author. Mailing address for Marco Ventura Department of Life Sciences, University of Parma, Parco Area delle Scienze 11a, 43124 Parma, Italy. Phone: +39-521-905666. Fax: +39-521-905604. E-mail: marco.ventura@unipr.it

Abstract

The genus *Lactobacillus* is a widespread taxon, members of which are highly relevant to functional and fermented foods, while they are also commonly present in host-associated gut and vaginal microbiota. Substantial efforts have been undertaken to disclose the genetic repertoire of all members of the genus *Lactobacillus*, yet their species-level profiling in complex matrices is still undeveloped due to the poor phylotype resolution of profiling approaches based on the 16S rRNA gene. To overcome this limitation, an ITS-based profiling method was developed to accurately profile lactobacilli at species-level. This approach encompasses a genus-specific primer pair combined with a database of ITS sequences retrieved from all available *Lactobacillus* genomes and a script for the Qiime software suite that performs all required steps to reconstruct a species-level profile. This methodology was applied to several environments, i.e., human gut and vagina, cecum of free range chickens, as well as whey and fresh cheese. Interestingly, data collected confirmed a relevant role of lactobacilli present in functional and fermented foods in defining the population harbored by the human gut, while, unsurprisingly perhaps, the cecum of free range chickens was observed to be dominated by lactobacilli characterized in birds living in natural environments. Moreover, vaginal swabs confirmed the existence of previously-hypothesized community state types, while analysis of whey and fresh cheese revealed a dominant presence of single *Lactobacillus* species used as starters for cheese production. Furthermore, application of this ITS profiling method to a mock *Lactobacillus*

community allowed a minimal resolution level of <0.006 ng/ μ l.

ImportanceThe genus *Lactobacillus* is a large and ubiquitous taxon of high scientific and commercial relevance. Despite the fact that the genetic repertoire of lactobacilli species has been extensively characterized, the ecology of this genus has been explored by metataxonomic techniques that are accurate down to the genus or phylogenetic group level only. Thus, the distribution of lactobacilli in environmental or processed food samples is relatively unexplored. The profiling protocol described here relies on the use of the Internally Transcribed Spacer to perform an accurate classification in a target population of lactobacilli with <0.006 ng/ μ l sensitivity. This approach was used to analyze five sample types collected from both human and animal host-associated microbiota as well as from the cheese production chain. Availability of a tool for species-level profiling of lactobacilli may be highly useful for both academic research and a wide range of industrial applications.

Introduction

The genus *Lactobacillus* is a widespread and diverse taxon encompassing more than 170 species and 17 subspecies, which are classified as Gram-positive, non-spore-forming and catalase-negative facultative anaerobes (1, 2). Moreover, based on their metabolic capability to produce lactic acid as the main metabolic end product of carbohydrate fermentation, lactobacilli are classified as members of the Lactic Acid Bacteria (LAB). Notably, 16S rRNA gene-based phylogenetic analyses revealed the existence of 22 distinct phylogenetic groups of

Lactobacillus species (24 including pediococci) (2-4).

Regarding their ecological distribution, lactobacilli are found in a wide range of environments, including plants, water, soil, silage and different body sites of humans and other animals as members of host-associated microbiomes, such as those colonizing the oral cavity, the vagina and the gastrointestinal tract (GIT) (4, 5). Moreover, 37 species of this genus have been granted the Qualified Presumption of Safety (QPS) status by the European Food Safety Authority (EFSA) (6). Thus, they are extensively used in the food industry, in particular in fermented foods due to their high performance in lactic acid fermentation coupled with high tolerance for low pH, preservative and organoleptic properties, and production of exopolysaccharides that contribute to the texture of foods (2, 7). In this context, members of the genus *Lactobacillus* have in recent years gained significant scientific and commercial interest as health-promoting microorganisms, as evidenced by the fact that 22 species encompass strains patented as probiotics in Europe (8).

The high commercial and scientific relevance of lactobacilli coupled to the recent introduction of next-generation sequencing technologies has recently led to genome decoding of all (then) known *Lactobacillus* species (3, 7). The retrieved genomic data has been exploited for comparative genomic analyses, and has allowed identification of many shared or distinct genetic features of this genus. Furthermore, this genomic information has permitted the reconstruction of their metabolic potential, has shed light on host-microbe interactions, such as adhesion to the mucus layer and modulation of the immune system of the host,

and has revealed particular microbe-microbe interactions with other commensals or (opportunistic) pathogens (1, 7, 8).

Despite the large body of data concerning the physiology and genetics of lactobacilli, knowledge about the ecology and distribution in environmental or host-associated niches of individual species relies mainly on culture-dependent studies. This is partly due to the resolution limit of currently used metagenomic approaches. Although microbial profiling based on partial 16S rRNA gene is able to discriminate between phylogenetic groups of lactobacilli due to the high phylogenetic diversity of this genus, it cannot provide an accurate species-level resolution. Moreover, the majority of the current exiting studies of lactobacilli populations based on 16S rRNA gene profiling do not even perform phylogenetic group-level analyses. To offer a more refined taxonomic view of lactobacilli in a given environment or sample, we developed a profiling approach based on amplification of the internally transcribed spacer (ITS) sequence. Notably, due to their high variability, ITS sequences have previously been exploited in a wide range of studies encompassing the identification of unique species-specific restriction patterns of lactobacilli, as well as the identification and characterization of *Leuconostoc* strains and for the genotyping of *Streptococcus pneumoniae* strains (9-11). The developed methodology in the current work is able to determine the composition of lactobacilli-containing communities down to the species level. The method was validated through the analysis of a sample artificially constituted by DNA of 14 lactobacilli taxa at known concentration. Furthermore, we applied this methodology for the

precise investigation of bacterial communities harbored by human-, animal- and food-associated matrices that were previously explored down to the genus-level only.

Results and discussion

Analysis of ITS variability within the *Lactobacillus* genus.

Genomes of 1523 strains assigned to the genus *Lactobacillus* and corresponding to 176 species were retrieved from the NCBI genome database, and then processed using the MEGAnnotator software (12) for prediction of rRNA genes in order to ensure the same high-quality standard for all sequences of ribosomal loci included in this study (Table S1). Notably, the genomic sequences of 892 *Lactobacillus* strains, representing the 58.6 % of the total strain pool analyzed, did not harbor complete rRNA loci, i.e. encompassing complete 5S, 16S and 23S rRNA genes. In contrast, at least one complete ribosomal rRNA genes locus was identified for 631 of the 1523 analyzed strains, corresponding to 70 species and a custom script was then used to extract a total of 1788 Internally Transcribed Spacer (ITS) sequences. Assembly of draft genomes generally generates the collapse of reads that correspond to rRNA genes into a single rRNA locus. However, availability of multiple draft sequences of a given *Lactobacillus* taxon, complemented with analysis of 217 complete genomes of *Lactobacillus* species, allowed us to retrieve an average of 25.5 ITS sequences per species. Interestingly, 137 of the 1788 retrieved ITS sequences include stretches of >3 undefined (N) nucleotides, thus highlighting a high rate of assembly-related issues and/or low-quality regions in genomes deposited at the NCBI genome

database. Comparative analysis of the 1651 complete ITS sequences without multiple contiguous nucleotide ambiguities revealed that 92.5 % of the ITS sequences range between 200 and 500 bp.

As previously observed for bifidobacteria (13, 14), alignment of ITS sequences from *Lactobacillus* genomes shows a high level of diversity, probably due to a high mutation frequency, and corresponding to a high evolutionary rate, as reflected by multiple substitutions at a given nucleotide position and indicative of mutational saturation of such ITS sequences. While this particularly high mutation frequency prevents phylogeny inference (15), it is suitable for metagenomic amplicon-based profiling below the genus level, as previously validated for members of the genus *Bifidobacterium* (13).

Design of a PCR primer pair for ITS-profiling of the *Lactobacillus* genus.

Many profiling approaches have been developed to accurately reconstruct the taxonomic composition of complex bacterial communities. These include methods based on low-coverage sequencing of full-length 16S rRNA genes and the use of technologies providing long reads, i.e. Sanger and PacBio. Nevertheless, despite the fact that full-length sequencing of the 16S rRNA gene allows high accuracy in taxonomic assignment, the low sequencing coverage permits the detection only of dominant taxa and prevents profiling of bacteria present at low relative abundance in a given population (10). Furthermore, the use of alternative marker genes has also been proposed, though their use remains limited due to difficulties in the definition of universal primers as well as in the lack of a complete reference database. The advent of

next-generation sequencing, characterized by high coverage and short reads, facilitated the amplification and sequencing of partial 16S rRNA genes, i.e. 16S rRNA gene profiling. This metagenomic method has in recent years been used as the gold standard for taxonomic characterization of environmental and host-associated microbiomes. While this methodology covers all bacterial biodiversity, it is generally only accurate for the reconstruction of taxonomic profiles down to genus level (16) or down to phylogenetic groups in case of genera with a high level of phylogenetic diversity, e.g. the genus *Lactobacillus* (3, 4) since it relies on sequencing of a small region of the whole 16S rRNA gene through next-generation sequencing. To overcome this limitation and to obtain species-level resolution, the use of the ITS sequence as an alternative molecular marker has been proposed (13). In order to develop a universal primer pair suitable for profiling of all members of the *Lactobacillus* genus, we aligned the 16S and 23S rRNA genes flanking the 1651 complete ITS sequences without stretches of undefined nucleotides that were retrieved from lactobacilli genomes deposited at the NCBI database. Manual inspection of the alignments allowed the identification of ‘universal’ primers located at the 5’-end of the 16S rRNA gene and at the 3’-end of the 23S rRNA gene, i.e., Probio-lac_Uni (CGTAACAAGGTAGCCGTAGG) and Probio-lac_Rev (GTYVCGTCCTTCWTCGSC), respectively (Figure 1). Sequence conservation amongst the aligned 16S and 23S rRNA genes is reported in Figure 1 through a WebLogo representation. These primers generate an amplicon of an average length of 380 bp covering the complete ITS region and suitable for 2 X 250

bp paired-end Illumina sequencing followed by single-end bioinformatic analysis of both paired reads (see below). Analysis of single-end reads provided reliable assignment to species level even in cases where a tRNA gene was located within the ITS region (see below). Notably, the final sequence of the primers was defined after multiple iterative alignments to the Silva SSU and LSU databases (17) using the Silva TestProbe v. 3.0 tool (<https://www.arb-silva.de/search/testprobe/>). The latter approach led to introduction of specific IUPAC bases in order to maximize alignment of the primers to all currently available 16S and 23S rRNA gene sequences of lactobacilli corresponding to all known species of this genus, while minimizing alignment to non-lactobacilli ribosomal RNA genes. The usefulness of the Probio-lac_Uni/Probio-lac_Rev primer pair was *in vitro* validated through successful amplicon generation in the case of 31 lactobacilli species belonging to the 23 phylogenetic groups identified previously in the genus *Lactobacillus* (3, 4) (Figure S1). In contrast, no amplification was observed when the Probio-lac_Uni/Probio-lac_Rev primer pair was used to amplify DNA extracted from nine non-*Lactobacillus* taxa (Figure S1). Interestingly, for all tested lactobacilli we observed two PCR fragments, each with a molecular size ranging from 300 to 350 bp, and 500 to 550 bp, corresponding to the ITS region with and without a tRNA gene (see below for details), respectively (Figure S1) (Figure 1). Such ITS patterns confirmed those displayed in previous studies targeting the amplification of the ITS region of lactobacilli (18). Notably, for few taxa we observed a faint amplification fragment of 500 to 550 bp, which might suggest a lower copy

number of ITS regions encompassing tRNA genes in the same genome.

The Probio-lac_Uni/Probio-lac_Rev primer pair was employed for *in silico* PCR amplification of the 631 genomes of the genus *Lactobacillus* encoding at least one rRNA genes locus. This approach facilitated the development of a database encompassing 1651 complete ITS sequences without multiple ambiguous nucleotides, and flanked by partial 16S and 23S rRNA sequences, together constituting the *Lactobacillus* ITS Amplicon database (LITSA database).

Cross-alignment of all retrieved LITSA sequences using MatGAT software (19) was performed in order to evaluate the level of identity between predicted amplicons (Table S2) and to evaluate possible limits imposed by actual lactobacilli taxonomy to the proposed ITS profiling methodology. Notably, this analysis highlighted cases in which comparison of multiple LITSA sequences from the same strain showed low identity. In-depth investigation revealed that 46 of the 62 lactobacilli species included in the LITSA database contain at least one ITS sequence that harbors two tRNA genes (for Alanine and Isoleucine) (Figure 1). Notably, despite the fact that this prediction is limited due to the small number of complete genomes available, the presence of tRNA genes in one or multiple rRNA loci appears to be a common feature of genomes from members of the *Lactobacillus* genus.

Furthermore, cross-alignment analysis also revealed that the majority of the 62 *Lactobacillus* species, for which a complete LITSA sequence was available, can be discriminated (Table S2), with the exception of putatively misclassified strains and/or species (see below). In this context, despite the fact

that lactobacilli are known to possess a very high level of phylogenetic diversity (3, 4), strains corresponding to 18 species showed an average LITSA sequence identity of >97 % with at least one other *Lactobacillus* species, thus showing a very close phylogenetic relationship between such taxa (Table 1). Amongst lactobacilli, *Lactobacillus casei* and *Lactobacillus paracasei* strains possess an average LITSA sequence identity of 99 %, while the amplicon sequences of *Lactobacillus pentosus*, *Lactobacillus plantarum* and *Lactobacillus paraplantarum* strains show up to 100 % identity (Table S2). An in-depth analysis of each strain revealed that 23 of the 25 strains classified as *L. casei* share an average LITSA sequence identity \leq 96.1 % with the type strain *L. casei* ATCC 393, while the average identity with the type strain *L. paracasei* ATCC 394 is \geq 99.7 % (Table S2). In contrast, the putative lactobacilli species *Lactobacillus* sp. FMNP02 shares 99.7 % identity with *L. casei* ATCC 393 (Table S2), thus representing a possible misclassification of the latter strain.

In our attempts to obtain insights into the phylogeny of *L. pentosus*, *L. plantarum* and *L. paraplantarum*, we observed an average LITSA sequence identity of 98.9 % between *L. pentosus* and *L. plantarum* strains (Table S2). Moreover, the two strains of *L. paraplantarum*, for which we were able to predict an rRNA gene locus, show an average LITSA identity of 99.5 % with *L. plantarum* strains (Table S2), thus indicating that such taxa may belong to the same species and therefore cannot be discriminated using metataxonomic techniques. Nevertheless, evaluation of the average nucleotide identity is needed to confirm this hypothesis. Furthermore, we

could not retrieve an *in silico* Probio-lac_Uni/Probio-lac_Rev-corresponding amplicon for the type strains of *L. pentosus* and *L. paraplantarum* due to absence of a complete ITS region in the deposited genomes, and we were therefore unable to evaluate their amplicon identity with the LITSA sequences of *L. plantarum* strains. Notably, these observations suggest that major issues in the classification of the genus *Lactobacillus* still exist, resulting in the unfeasibility of distinguish a number of species through ITS profiling. Thus, as has been proposed previously, it is desirable that a re-evaluation of the taxonomy of lactobacilli is undertaken based on a phylogenomic approach (20, 21), as was also corroborated by recent studies (3, 4, 7).

Development of a bioinformatic tool for ITS-profiling of the *Lactobacillus* genus.

The length of the amplicon produced by the Probio-lac_Uni/Probio-lac_Rev primer pair may exceed 600 bp, particularly when tRNA-encoding sequences are present in the ITS sequence. Thus, sequencing produced non-overlapping paired-end reads even with the maximum length obtainable using Next-Generation Illumina sequencing, i.e. 2 X 250 bp paired-end reads, using the MiSeq Reagents Kit v3 600 cycles chemistry. Nevertheless, each forward and reverse read covers 42 and 60 nucleotides corresponding to the 16S rRNA gene 3'-end and the 23S rRNA gene 5'-end, respectively, which are followed by 190-208 bp of hyper-variable ITS sequence suitable for profiling at species-level (Figure 1). Thus, we developed a package for QIIME software suite v1.9.1 (22) that encompasses the LITSA database and a bash script for analysis of both forward and reverse reads of the *Lactobacillus* ITS profiling data

(probiogenomics.unipr.it/pbi). Notably, the LITSA database will be updated regularly to include additional ITS sequences as new lactobacilli genome sequences become available, thus increasing the number of lactobacilli species that can be profiled. The script performs quality-filtering, *de novo* OTU clustering at 100 % identity and taxonomic classification of OTU reference sequences through RDP classifier with a confidence level of 0.80. Notably, these cut-off values permit discrimination of closely related taxa. Due to the average size of the amplicon, the paired-end reads are not joined prior classification. Instead, the script analyzes both the forward and the reverse reads altogether and provides an average profile.

Notably, the different number of rRNA loci predicted in the genomes of *Lactobacillus* species may generate biases in the retrieved profiles. Thus, we evaluated the average number of ITS regions present in the 217 available complete *Lactobacillus* genomes. This analysis provided data for normalization of 45 of the 62 species of lactobacilli for which a LITSA sequence could be retrieved. Moreover, the average number of rRNA genes loci of the remaining 17 species with only draft genomes was set at 5.6, i.e., the average obtained for all the species with at least a complete genome. Notably, the *Lactobacillus* ITS profiling analysis script includes a normalization step based on the number of rRNA genes loci predicted for all the 62 *Lactobacillus* species for which a LITSA sequence could be retrieved. The output produced by the script is summarized in the “output” folder, which contains the predicted taxonomic profile based on the LITSA database (both non-normalized and normalized for the number of rRNA loci) and the

OTU table in tabular text format that reports the reference sequence and associated taxonomy. All *Lactobacillus* ITS profiles reported in this manuscript correspond to the average between forward and reverse read profiles after normalization for the number of predicted rRNA genes loci.

Assessing detection sensitivity and accuracy using the *Lactobacillus* ITS profiling protocol

In order to provide an evaluation of the sensitivity and accuracy of the Probio-lac_Uni/Probio-lac_Rev primer pair, 14 *Lactobacillus* type strains were employed to artificially compose a mock community (Table S3). The DNA extracted from each taxon grown in pure culture was added to the mix at known amount, ranging from 0.006 ng to 50 ng of DNA, corresponding to 0.006 % to 50 % of the total DNA pool (Figure 2). Sequencing of the mock sample was performed using an Illumina MiSeq with 2X250 bp chemistry, producing 45,146 quality-filtered paired-end reads. Interestingly, *Lactobacillus* ITS profiling of this dataset successfully profiled all *Lactobacillus* species included in this sample, except *Lactobacillus vaginalis* and *Lactobacillus pontis*, for which we could not retrieve a LITSA sequence from analysis of available genome sequences (Figure 2). Thus, even though the Probio-lac_Uni/Probio-lac_Rev primer pair produces an amplicon for these species, the latter cannot be taxonomically classified due to absence of *L. vaginalis* and *L. pontis* in the present version of the LITSA database. This is a temporary limitation and the LITSA database will be updated regularly (probiogenomics.unipr.it/pbi) to include LITSA sequences of newly sequenced genomes in order to cover all the lactobacilli species that currently cannot be profiled. Moreover,

comparison of the retrieved profile with the expected composition revealed a strong correlation for each taxon with few discrepancies (Figure 2). The causes of such differences between expected and observed relative abundance may be imputed to the lack of sufficient information in the LITSA database, at this time, regarding the average number of rRNA loci per genome used for normalization of the ITS profiling data.

Furthermore, since PCR amplicon size has been identified as a source of bias in ITS-based profiling studies of fungi (23), we evaluated the presence of possible biases introduced by amplification of lactobacilli ITS sequences of different length due to the presence or absence of tRNA genes (see above). The 14 *Lactobacillus* species that constitute the mock community (Table S3) were subjected to manual characterization of corresponding rRNA loci. Notably, the ten species for which a complete genome was available, confirmed what had been observed for the *in vitro* PCR, i.e. presence of longer ITS sequences that encompass two tRNA genes (Figure 1; Figure S1). Interestingly, the different intensities observed in the PCR fragments, i.e. 300-350 bp and 500-550 bp (Figure S1), did not influence expected relative abundance of the mock community (Figure 2). Notably, detection of *Lactobacillus rhamnosus* whose concentration in the mock community is 0.006 ng/ μ l indicates that the limit of detection of the lactobacilli ITS profiling is <0.006 ng/ μ l, corresponding to $1.85 \cdot 10^3$ cells/ μ l.

Validation of the *Lactobacillus* ITS profiling protocol through analysis of samples from multiple environments.

Lactobacillus is a highly diverse microbial genus, members of which are found in a wide range of environments (5). To perform a comprehensive testing of the performances of the *Lactobacillus* ITS profiling protocol, we analyzed a total of 25 samples encompassing five human faecal samples, five human vaginal swab samples, five free range chicken cecal samples, five whey samples and five parmesan cheese samples (Table S4). Sequencing was performed with an Illumina MiSeq instrument using 2x250 bp chemistry, producing an average of 15,529 forward and 15,293 reverse quality-filtered reads per sample (Table S4).

Interestingly, analysis of the human faecal samples revealed the presence of human gut colonizers, such as *Lactobacillus rhamnosus*, along with a range of lactobacilli used in functional or fermented foods that are typically part of the human diet, such as *L. plantarum*, *Lactobacillus helveticus*, *Lactobacillus delbrueckii* and *Lactobacillus sakei* (Figure 3).

Moreover, the obtained profiles of the five human vaginal swab samples confirmed the proposed existence of community state types (CSTs) of the vaginal microbiota dominated by specific *Lactobacillus* taxa (24, 25). In fact, HV1 is dominated by *Lactobacillus gasseri*, while *Lactobacillus iners* and *Lactobacillus crispatus* are the most abundant lactobacilli taxa in HV2/HV5 and HV3/HV4, respectively (Figure 3). Furthermore, in all five reconstructed human vaginal profiles, *L. helveticus* is the second most abundant *Lactobacillus* species, as observed in the aforementioned CSTs (24, 25) (Figure 3). Thus, based on the classification proposed by DiGiulio et al. (24), HV1 can be classified as a CST 2, while

HV2/HV5 falls within the CST 3, whereas HV3/HV4 can be attributed to CST 1.

To demonstrate the relevance of an efficient methodology for precise cataloguing of the *Lactobacillus* species for which a complete LITSA sequence is available in different environments, we analyzed five free range chickens cecal samples. The retrieved profiles revealed a high relative abundance (ranging from a total of 53.1 % to 96.8 %) of *Lactobacillus* species previously characterized in poultry or other birds, such as *Lactobacillus salivarius*, *Lactobacillus reuteri*, *Lactobacillus ingluviei*, *Lactobacillus amylovorus*, *Lactobacillus agilis*, *Lactobacillus aviarius* and *Lactobacillus johnsonii* (26-33) (Figure 3). Notably, samples FRC1, FRC2 and FRC3 showed a similar profile with high abundance of *L. salivarius*, *L. ingluviei* and *L. amylovorus*, reflecting the fact that they were kept in the same hen house (Figure 3). Accordingly, samples FRC4 and FRC5, collected in two additional hen houses, showed different profiles characterized by high abundance of *L. aviarius* and *L. johnsonii*, respectively (Figure 3).

For milk and milk-related products, profiling of five whey and five fresh parmesan cheeses (at 1 day of ripening) samples revealed, as expected, similar profiles dominated by *L. helveticus* and *L. delbrueckii* (Figure 3), which represent two lactobacilli species typically used as starter cultures for the production of cheese (34). These data indicate that the *Lactobacillus* ITS profiling approach also represents a valuable tool for monitoring the population of lactobacilli across the cheese production chain.

Results obtained from ITS-profiling were also compared to profiles reconstructed through

analysis of OTUs generated at 99 % identity from 16S rRNA profiling data (Figure 3) (Table S5). Notably, only OTUs classified as lactobacilli have been included in the representation, thus the relative abundance of unclassified lactobacilli reported in the bar plot do not include additional OTUs that could not be attributed to this genus. Moreover, lactobacilli species whose relative abundance is below 5 % in each sample were collapsed under “Others <5 %” in the bar plot representation. Interestingly, the ITS profiling approach provided a more accurate species-level reconstruction of the lactobacilli populations when used to analyze human faecal and vaginal samples as well as free range chicken faecal samples. Moreover, it confirmed and partially improved the simple lactobacilli community of whey and fresh Parmesan cheese samples observed through 16S rRNA gene profiling. In fact, differences in the profiles obtained through 16S rRNA gene and ITS profiling can be observed in all cases (Figure 3) (Table S5). Such differences are caused by the limited number of *Lactobacillus* species that could be discriminated based on partial 16S rRNA gene sequence respect to ITS sequence (Figure 3) (Table S5). Altogether, these results confirm the performance of the *Lactobacillus* ITS profiling protocol observed from analysis of the artificial sample and validate their use, complementary to 16S rRNA gene profiling, for analysis of a wide range of complex environmental and host-associated matrices.

Conclusions

We developed a newly designed method for characterization of the *Lactobacillus* population in complex environments based on the use of the

internally transcribed spacer (ITS), which represents a hypervariable region located between the 16S and the 23S rRNA genes that allows high-accuracy species-level profiling. The accuracy and sensitivity of this method allowed profiling of complex communities of lactobacilli with a successful identification of taxa with abundance of 1.85×10^3 cells/ μ l, which is even lower to what was previously identified for a similar approach developed for the profiling of bifidobacterial communities (13). Notably, despite the fact that the current LITSA database allows the precise profiling of just 62 species, the ITS-profiling approach represents a new metagenomic tool for species-level profiling of complex lactobacilli communities that complements phylogenetic group assignments that can be obtained from 16S rRNA gene profiling data. Moreover, the database will be regularly updated to represent additional lactobacilli species as genomes encompassing complete LITSA sequences are becoming available. When the ITS lactobacilli profiling method was applied to different biological samples, encompassing the stool of human as well as birds, vaginal swabs and cheese, it allowed the reconstruction of the cataloguing of lactobacilli communities residing in these environments. Altogether, these results highlight that ITS-mediated profiling of populations of lactobacilli could be useful not only for academic purposes, but also for industrial applications such as tracing the microbial composition of probiotic products based on lactobacilli as well as of starter cultures in food manufacture.

Material and methods

Sample collection

In the frame work of a more extensive bacterial cataloguing project, this study enrolled stool, vaginal swab, fresh parmesan cheese (one day of ripening), whey and cecal (from free range chickens) samples.

Five fresh stool samples obtained from human healthy volunteers and five cecal samples retrieved from free range chickens were immediately frozen upon collection at -80°C until processing for DNA extraction. The DNA extraction was performed using the QIAamp DNA Stool Mini Kit following the manufacturer's instructions (Qiagen, Manchester, UK). Additionally, five vaginal swab samples were collected in sterile tubes containing 1 ml of DNA-RNA shield from ZYMO Research until bacterial DNA extraction using ZymoBIOMICS™ DNA Miniprep Kit (ZYMO Research). Furthermore, 10 ml samples of whey and 2-4 gr of fresh parmesan cheese were collected in sterile tubes and the DNA was extracted using the DNeasy Mastitis Mini Kit (Qiagen Ltd, Strasse, Germany) following the manufacturer's instructions (Qiagen Ltd). Notably, whey samples and cheese samples at one day of ripening were collected from the same Parmesan cheese producer in Parma, Italy.

Ethical statement. This study was carried out in accordance with the recommendations of the ethical committee of the University of Parma and was approved by the “Comitato di Etica Università degli Studi di Parma”, Italy. All animal procedures were performed according to national guidelines (Decreto legislativo 26/2014).

Bacterial growth conditions and DNA extraction. Type strains of several lactobacilli taxa (Table S3) were growth in Man-Rogosa-Sharpe (MRS) medium (Scharlau Chemie) supplemented

with 0.05 % (w/v) L-cysteine hydrochloride and incubated in an anaerobic atmosphere (2.99 % H₂, 17.01 % CO₂ and 80 % N₂) in a chamber (Concept 400; Ruskin) at 37°C for 24 h. In addition, nine non-lactobacilli microorganisms were used in this study. These included *Bifidobacterium bifidum* LMG11041, which was cultivated in MRS broth as *Lactobacillus* strains; *Collinsella intestinalis* DSM 13280, *Escherichia coli* LMG 2092, and *Klebsiella pneumoniae* CECT 143, which were grown in de MRS broth (Difco, Detroit, MI) supplemented with 0.05% (w/v) l-cysteine (MRSC; Sigma, St. Louis, MO). *Prevotella copri* DSM 18205 and *Blautia coccooides* DSM 935 were cultivated in a combination of Reinforced Clostridial Broth (Merck, Darmstadt, Germany) and Brain-Heart Infusion (Difco), supplemented with 5% (v/v) heat-inactivated fetal bovine serum (LabClinics, Barcelona, Spain) respectively. For *Bacteroides thetaiotaomicron* DSMZ 2079, the latter medium was supplemented with 0.005 % hemin (Sigma) and 0.005 % Vitamin K1 (Sigma). *Faecalibacterium prausnitzii* DSM 17677 was grown in Wilkins-Chalgren Anaerobe broth (Merck), following the recommendations included in the DSMZ medium 339. Finally, an active culture of *Methanobrevibacter smithii* DSM 861 grown in *Methanobacterium* medium (DSMZ 119) was directly supplied by DSMZ.

Bacterial DNA was extracted using GenElute™ Bacterial Genomic DNA kits (SIGMA-ALDRICH) following the manufacturer's instructions. Taxonomic identity of the microorganisms was validated by sequencing the V3 variable region of the 16S rRNA gene using primer pair Probio_Uni and/Probio_Rev (14).

***Lactobacillus* mock community**

The cultures of fourteen different *Lactobacillus* strains were grown separately on Man-Rogosa-Sharpe (MRS) medium (Scharlau Chemie) supplemented with 0.05 % (w/v) L-cysteine hydrochloride and incubated in an anaerobic atmosphere (2.99 % H₂, 17.01 % CO₂ and 80 % N₂) in a chamber (Concept 400; Ruskin) at 37°C until they reached late log phase. The bacteria were enumerated by counting colonies on solid medium and the optical density at 600 nm was determined. The final bacterial cell concentration was approximately 10⁷ cfu/ml. Chromosomal DNA of each strains was extracted as previously described and subsequently mixed.

Specifically, the mock community consists of a pool of known concentration of fourteen different *Lactobacillus* strains to obtain the final quantity of DNA indicated in Table S3. Furthermore, the mix was prepared by combining equal volumes (20 µL) of DNA. The DNA from the mix was diluted to produce a final DNA concentration of 2 ng/µL, and 4 µL of these dilutions were used in each PCR reaction. For the PCR reaction, the primer pair Probio-lac_Uni/Probio-lac_Rev was used and the generated amplicons were sequenced using Illumina MiSeq as described below.

***Lactobacillus* ITS-specific primer design and gene amplification**

The bioinformatics platforms MEGAnnotator (10) and METAnnotatorX (unpublished data) were used to perform 16S and 23S rRNA genes prediction in all 1523 sequenced lactobacilli genomes deposited at the NCBI Genomes database. Primers Probio-lac_Uni (CGTAACAAGGTAGCCGTAGG)/Probio-lac_Rev (GTYVCGTCCTTCWTCGSC) were manually designed based on the alignment of all

16S and 23S rRNA sequences to generate an amplicon encompassing the 3'-end of the 16S rRNA gene, the ITS region and the 5'-end of the 23S rRNA gene. Specificity test was performed using the Silva TestProbe v. 3.0 tool (<https://www.arb-silva.de/search/testprobe/>) that allows alignment of primers sequences to the Silva SSU and LSU databases (15). A custom bioinformatics script was then used to create a database of all the Probio-lac_Uni/Probio-lac_Rev-generated lactobacilli ITS amplicon sequences (LITSA database), encompassing the Internally Transcribed Spacer (ITS) region and partial 16S and 23S rRNA genes.

The PCR conditions used for *Lactobacillus* ITS profiling using the Probio-lac_Uni/Probio-lac_Rev primer pair were 5 min at 95 °C, 30 cycles of 30 s at 95 °C, 30 s at 58 °C, and 40 s at 72 °C, followed by 10 min at 72 °C. Amplification was carried out using a Verity Thermocycler (Applied Biosystems). The integrity of the PCR amplicons was analyzed by gel electrophoresis. An additional specificity test was performed by PCR using the DNA extracted from all known *Lactobacillus* species as well as *B. bifidum* ATCC11041, *C. intestinalis* DSM 13280, *E. coli* LMG 2092, *K. pneumoniae* CECT 143, *P. copri* DSM 18205, *Bl. coccooides* DSM 935, *Bc. thetaiotaomicron* DSMZ 2079, *F. prausnitzii* DSM 17677 and *M. smithii* DSM 861.

WebLogo representation of primer sequence conservation among the retrieved 16S and 23S rRNA genes flanking complete ITS sequences was obtained through the WebLogo website (<http://weblogo.berkeley.edu/>) (35).

Illumina MiSeq sequencing of ITS gene-based amplicons

Illumina adapter overhang nucleotide sequence was added to the PCR amplicons obtained following amplification of the ITS region, as previously described (13). The library of ITS amplicons was prepared following the 16S Metagenomic Sequencing Library Preparation Protocol (Part No. 15044223 Rev. B-Illumina). Sequencing was performed using an Illumina MiSeq sequencer with MiSeq Reagent Kit v3 chemicals.

ITS-based microbiota analysis

Fastq files obtained from metagenomic sequencing of each sample were analyzed using a custom script for QIIME software suite (22) and the LITSA database available at (<http://probiogenomics.unipr.it/pbi/index.html>).

Input data were processed in the following steps: filtering of the reads based on length > 100 nt (primers included) to avoid primer dimers, overall quality > 25 and the presence of forward and reverse primers in the forward and reverse reads, respectively, creation of OTUs constituted by identical sequences using prefix_suffix method and removal of OTUs represented by < 10 sequences. Taxonomy assignment was performed using RDP method (RDP classifier with a confidence level of 0.80) and the LITSA database constituted by ITS sequences retrieved from the 1523 *Lactobacillus* genomes available at the NCBI Genome database. This script is easily modifiable to obtain a profiling based on a different sequence, though will depend on the availability of a corresponding database.

Evaluation of the sensitivity of the Probio-lac_Uni/Probio-lac_Rev primer pair

The artificial sample used for the evaluation of the detection sensitivity and accuracy of the Probio-lac_Uni/Probio-lac_Rev primer set was generated

using known DNA amounts, ranging from 50 to 0.006 ng, of 14 different *Lactobacillus* taxa (Table S3).

Microbiota identification by 16S rRNA gene-amplification, -sequencing and data analysis.

Partial 16S rRNA gene sequences were amplified from extracted DNA using primer pair Probio_Uni / Probio_Rev, which target the V3 region of the 16S rRNA gene sequence (16). 16S rRNA gene amplification and amplicon checks were carried out as previously described (16). 16S rRNA gene sequencing was performed using a MiSeq (Illumina) at the DNA sequencing facility of GenProbio srl (www.genprobio.com) according to a previously reported protocol (16). Following sequencing, the .fastq files were processed using a custom script based on the QIIME software suite (22). Paired-end read pairs were assembled to reconstruct complete Probio_Uni / Probio_Rev amplicons. Quality control retained sequences with a length between 140 and 400 bp and mean sequence quality score >20 while sequences with homopolymers >7 bp and mismatched primers were omitted. 16S rRNA gene Operational Taxonomic Units (OTUs) were defined at $\geq 99\%$ sequence homology using uclust (36) and OTUs with less than 10 sequences were filtered. All reads were classified to the lowest possible taxonomic rank using QIIME (37) and a reference dataset from the SILVA database (Quast et al., 2013).

Nucleotide sequence accession numbers

The raw ITS and 16S rRNA gene sequences reported in this article have been deposited in the NCBI Short Read Archive (SRA) under the accession number PRJNA434072.

Acknowledgements

This work was funded by the EU Joint Programming Initiative – A Healthy Diet for a Healthy Life (JPI HDHL, <http://www.healthydietforhealthylife.eu/>) to DvS (in conjunction with Science Foundation Ireland [SFI], Grant number 15/JP-HDHL/3280) and to MV (in conjunction with MIUR, Italy). We thank GenProbio srl for financial support of the Laboratory of Probiogenomics. This research benefits from the HPC (High Performance

Computing) facility of the University of Parma, Italy. DvS is a member of The APC Microbiome Ireland supported by Science Foundation Ireland (SFI), through the Irish Government's National Development Plan (Grant number SFI/12/RC/2273). Furthermore, DvS is a visiting professor of the University of Parma supported by Tech In Parma. The authors declare that they have no competing interests.

References

1. Goldstein EJ, Tyrrell KL, Citron DM. 2015. *Lactobacillus* species: taxonomic complexity and controversial susceptibilities. *Clin Infect Dis* 60 Suppl 2:S98-107.
2. Salvetti E, Torriani S, Felis GE. 2012. The Genus *Lactobacillus*: A Taxonomic Update. *Probiotics Antimicrob Proteins* 4:217-26.
3. Zheng J, Ruan L, Sun M, Ganzle M. 2015. A Genomic View of *Lactobacilli* and *Pediococci* Demonstrates that Phylogeny Matches Ecology and Physiology. *Appl Environ Microbiol* 81:7233-43.
4. Duar RM, Lin XB, Zheng J, Martino ME, Grenier T, Perez-Munoz ME, Leulier F, Ganzle M, Walter J. 2017. Lifestyles in transition: evolution and natural history of the genus *Lactobacillus*. *FEMS Microbiol Rev* 41:S27-S48.
5. Walter J. 2008. Ecological role of *Lactobacilli* in the gastrointestinal tract: implications for fundamental and biomedical research. *Appl Environ Microbiol* 74:4985-96.
6. Ricci A, Allende A, Bolton D, Chemaly M, Davies R, Girones R, Koutsoumanis K, Herman L, Lindqvist R, Nørrung B, Robertson L, Ru G, Sanaa M, Simmons M, Skandamis P, Snary E, Speybroeck N, Ter Kuile B, Threlfall J, Wahlström H, Cocconcelli PS, Klein G, Peixe L, Maradona MP, Querol A, Suarez JE, Sundh I, Vlaskovic J, Correia S, Fernández Escámez PS. 2017. Update of the list of QPS-recommended biological agents intentionally added to food or feed as notified to EFSA 5: suitability of taxonomic units notified to EFSA until September 2016. *EFSA Journal* 15:e04663-n/a.
7. Sun Z, Harris HM, McCann A, Guo C, Argimon S, Zhang W, Yang X, Jeffery IB, Cooney JC, Kagawa TF, Liu W, Song Y, Salvetti E, Wrobel A, Rasinkangas P, Parkhill J, Rea MC, O'Sullivan O, Ritari J, Douillard FP, Paul Ross R, Yang R, Briner AE, Felis GE, de Vos WM, Barrangou R, Klaenhammer TR, Caufield PW, Cui Y, Zhang H, O'Toole PW. 2015. Expanding the biotechnology potential of *Lactobacilli* through comparative genomics of 213 strains and associated genera. *Nat Commun* 6:8322.
8. Salvetti E, O'Toole PW. 2017. The Genomic Basis of *Lactobacilli* as Health-Promoting Organisms. *Microbiol Spectr* 5.
9. Sandes SH, Alvin LB, Silva BC, Zanirati DF, Jung LR, Nicoli JR, Neumann E, Nunes AC. 2014. *Lactobacillus* species identification by amplified ribosomal 16S-23S rRNA restriction fragment length polymorphism analysis. *Benef Microbes* 5:471-81.
10. Barrangou R, Yoon SS, Breidt F, Jr., Fleming HP, Klaenhammer TR. 2002. Identification and characterization of *Leuconostoc fallax* strains isolated from an industrial sauerkraut fermentation. *Appl Environ Microbiol* 68:2877-84.
11. Moore JE, Hirayama J, Hayashi K, Mason C, Coulter W, Matsuda M, Goldsmith CE. 2018. Examination of 16S-23S rRNA intergenic spacer region (ISR) heterogeneity in a population of clinical *Streptococcus pneumoniae*- a new laboratory epidemiological genotyping tool to aid outbreak analysis. *Br J Biomed Sci* doi:10.1080/09674845.2017.1382025:1-3.
12. Lugli GA, Milani C, Mancabelli L, van Sinderen D, Ventura M. 2016. MEGAnnotator: a user-friendly pipeline for microbial genomes assembly and annotation. *FEMS Microbiol Lett* 363.
13. Milani C, Lugli GA, Turrioni F, Mancabelli L, Duranti S, Viappiani A, Mangifesta M, Segata N, van Sinderen D, Ventura M. 2014. Evaluation of bifidobacterial community composition in the human gut by means of a targeted amplicon sequencing (ITS) protocol. *FEMS Microbiol Ecol* 90:493-503.
14. Milani C, Mancabelli L, Lugli GA, Duranti S, Turrioni F, Ferrario C, Mangifesta M, Viappiani A, Ferretti P, Gorfer V, Tett A, Segata N, van Sinderen D, Ventura M. 2015. Exploring Vertical Transmission of Bifidobacteria from Mother to Child. *Appl Environ Microbiol* 81:7078-87.
15. Philippe H, Adoutte A. 1996. *What Can Phylogenetic Patterns Tell Us About the Evolutionary Process Generating Biodiversity?* Oxford University Press, Oxford, UK.
16. Milani C, Hevia A, Foroni E, Duranti S, Turrioni F, Lugli GA, Sanchez B, Martin R, Gueimonde M, van Sinderen D, Margolles A, Ventura M. 2013. Assessing the fecal microbiota: an optimized ion torrent 16S rRNA gene-based analysis protocol. *PLoS One* 8:e68739.
17. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glockner FO. 2013. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* 41:D590-6.
18. Chebenova-Turcovska V, Zenisova K, Kuchta T, Pangallo D, Brezna B. 2011. Culture-independent detection of microorganisms in traditional Slovakian bryndza cheese. *Int J Food Microbiol* 150:73-8.
19. Campanella JJ, Bitincka L, Smalley J. 2003. MatGAT: an application that generates similarity/identity matrices using protein or DNA sequences. *BMC Bioinformatics* 4:29.

20. Lugli GA, Mangifesta M, Duranti S, Anzalone R, Milani C, Mancabelli L, Alessandri G, Turrone F, Ossiprandi MC, van Sinderen D, Ventura M. 2018. Phylogenetic classification of six novel species belonging to the genus *Bifidobacterium* comprising *Bifidobacterium anseris* sp. nov., *Bifidobacterium criceti* sp. nov., *Bifidobacterium imperatoris* sp. nov., *Bifidobacterium italicum* sp. nov., *Bifidobacterium margollesii* sp. nov. and *Bifidobacterium parmae* sp. nov. *Syst Appl Microbiol* doi:10.1016/j.syapm.2018.01.002.
21. Goris J, Konstantinidis KT, Klappenbach JA, Coenye T, Vandamme P, Tiedje JM. 2007. DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. *Int J Syst Evol Microbiol* 57:81-91.
22. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Pena AG, Goodrich JK, Gordon JI, Huttley GA, Kelley ST, Knights D, Koenig JE, Ley RE, Lozupone CA, McDonald D, Muegge BD, Pirrung M, Reeder J, Sevinsky JR, Turnbaugh PJ, Walters WA, Widmann J, Yatsunencko T, Zaneveld J, Knight R. 2010. QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* 7:335-6.
23. Bellemain E, Carlsen T, Brochmann C, Coissac E, Taberlet P, Kausserud H. 2010. ITS as an environmental DNA barcode for fungi: an in silico approach reveals potential PCR biases. *BMC Microbiol* 10:189.
24. DiGiulio DB, Callahan BJ, McMurdie PJ, Costello EK, Lyell DJ, Robaczewska A, Sun CL, Goltsman DS, Wong RJ, Shaw G, Stevenson DK, Holmes SP, Relman DA. 2015. Temporal and spatial variation of the human microbiota during pregnancy. *Proc Natl Acad Sci U S A* 112:11060-5.
25. Ravel J, Gajer P, Abdo Z, Schneider GM, Koenig SS, McCulle SL, Karlebach S, Gorle R, Russell J, Tacket CO, Brotman RM, Davis CC, Ault K, Peralta L, Forney LJ. 2011. Vaginal microbiome of reproductive-age women. *Proc Natl Acad Sci U S A* 108 Suppl 1:4680-7.
26. Svetoch EA, Eruslanov BV, Levchuk VP, Perelygin VV, Mitsevich EV, Mitsevich IP, Stepanshin J, Dyatlov I, Seal BS, Stern NJ. 2011. Isolation of *Lactobacillus salivarius* 1077 (NRRL B-50053) and characterization of its bacteriocin, including the antimicrobial activity spectrum. *Appl Environ Microbiol* 77:2749-54.
27. Kobierecka PA, Wyszynska AK, Aleksandrak-Piekarczyk T, Kuczkowski M, Tuzimek A, Piotrowska W, Gorecki A, Adamska I, Wieliczko A, Bardowski J, Jagusztyn-Krynicka EK. 2017. In vitro characteristics of *Lactobacillus* spp. strains isolated from the chicken digestive tract and their role in the inhibition of *Campylobacter* colonization. *Microbiologyopen* 6.
28. Merhej V, Armougom F, Robert C, Raoult D. 2012. Genome sequence of *Lactobacillus ingluviei*, a bacterium associated with weight gain in animals. *J Bacteriol* 194:5697.
29. Dec M, Puchalski A, Urban-Chmiel R, Wernicki A. 2014. Screening of *Lactobacillus* strains of domestic goose origin against bacterial poultry pathogens for use as probiotics. *Poult Sci* 93:2464-72.
30. Baele M, Devriese LA, Haesebrouck F. 2001. *Lactobacillus agilis* is an important component of the pigeon crop flora. *J Appl Microbiol* 91:488-91.
31. Fujisawa T, Shirasaka S, Watabe J, Mitsuoka T. 1983. *Lactobacillus aviarius* sp. nov.: A new species isolated from the intestine of chickens. *Systematic and Applied Microbiology* 5:414-420.
32. La Ragione RM, Narbad A, Gasson MJ, Woodward MJ. 2004. In vivo characterization of *Lactobacillus johnsonii* FI9785 for use as a defined competitive exclusion agent against bacterial pathogens in poultry. *Lett Appl Microbiol* 38:197-205.
33. Ventura M, Zink R. 2002. Specific identification and molecular typing analysis of *Lactobacillus johnsonii* by using PCR-based methods and pulsed-field gel electrophoresis. *FEMS Microbiol Lett* 217:141-54.
34. Gala E, Landi S, Solieri L, Nocetti M, Pulvirenti A, Giudici P. 2008. Diversity of lactic acid bacteria population in ripened Parmigiano Reggiano cheese. *Int J Food Microbiol* 125:347-51.
35. Crooks GE, Hon G, Chandonia JM, Brenner SE. 2004. WebLogo: a sequence logo generator. *Genome Res* 14:1188-90.
36. Edgar RC. 2010. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26:2460-1.
37. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Pena AG, Goodrich JK, Gordon JI, Huttley GA, Kelley ST, Knights D, Koenig JE, Ley RE, Lozupone CA, McDonald D, Muegge BD, Pirrung M, Reeder J, Sevinsky JR, Turnbaugh PJ, Walters WA, Widmann J, Yatsunencko T, Zaneveld J, Knight R. 2010. QIIME allows analysis of high-throughput community sequencing data. *Nature Methods* 7:335-336.

Table 1: List of *Lactobacillus* species with LITSA sequence identity ≥ 97 % with another *Lactobacillus* species. The percentage reported corresponds to the highest identity observed among all LITSA sequences identified in strains of the two species compared.

Species	LITSA % identity with the closest species	Closest species
<i>Lactobacillus acidophilus</i>	98	<i>Lactobacillus amylovorus</i>
	97	<i>Lactobacillus crispatus</i>
<i>Lactobacillus amylovorus</i>	98	<i>Lactobacillus acidophilus</i>
	97	<i>Lactobacillus crispatus</i>
<i>Lactobacillus buchneri</i>	99	<i>Lactobacillus parabuchneri</i>
<i>Lactobacillus casei</i>	99	<i>Lactobacillus paracasei</i>
	100	<i>Lactobacillus rhamnosus</i>
<i>Lactobacillus crispatus</i>	97	<i>Lactobacillus acidophilus</i>
	97	<i>Lactobacillus amylovorus</i>
<i>Lactobacillus curvatus</i>	98	<i>Lactobacillus sakei</i>
<i>Lactobacillus gallinarum</i>	99	<i>Lactobacillus helveticus</i>
<i>Lactobacillus gasseri</i>	99	<i>Lactobacillus johnsonii</i>
<i>Lactobacillus helveticus</i>	99	<i>Lactobacillus gallinarum</i>
<i>Lactobacillus johnsonii</i>	99	<i>Lactobacillus gasseri</i>
<i>Lactobacillus parabuchneri</i>	99	<i>Lactobacillus buchneri</i>
<i>Lactobacillus paracasei</i>	99	<i>Lactobacillus casei</i>
	100	<i>Lactobacillus rhamnosus</i>
<i>Lactobacillus paraplantarum</i>	100	<i>Lactobacillus pentosus</i>
	100	<i>Lactobacillus plantarum</i>
<i>Lactobacillus pentosus</i>	100	<i>Lactobacillus paraplantarum</i>
	99	<i>Lactobacillus plantarum</i>
<i>Lactobacillus plantarum</i>	100	<i>Lactobacillus paraplantarum</i>
	99	<i>Lactobacillus pentosus</i>
<i>Lactobacillus rhamnosus</i>	100	<i>Lactobacillus paracasei</i>
	100	<i>Lactobacillus casei</i>
<i>Lactobacillus sakei</i>	98	<i>Lactobacillus curvatus</i>

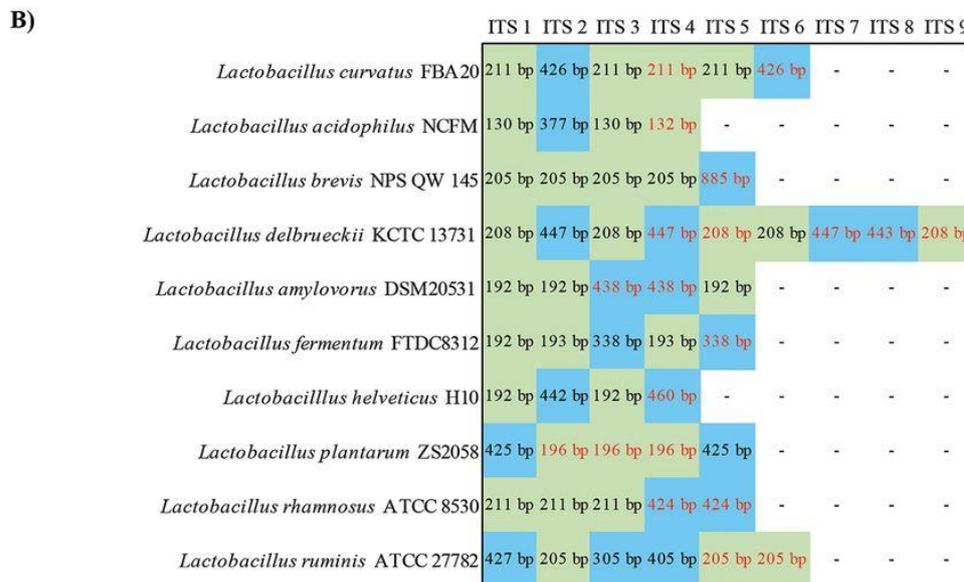
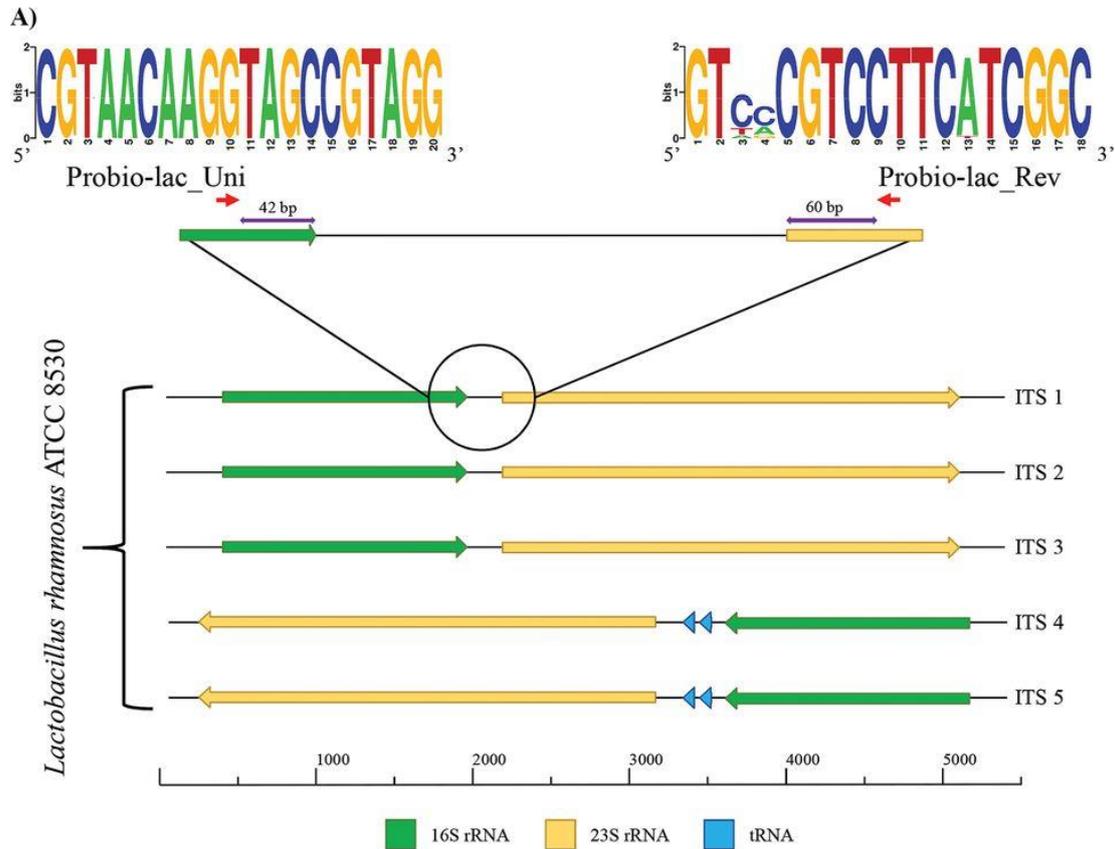


Figure 1: Genetic map of the Internally Transcribed Sequence (ITS) region of *Lactobacillus* with and without tRNA genes. Panel a depicts the genetic organization of the five complete ITS regions predicted in the complete genome of *Lactobacillus rhamnosus* ATCC 8530, used here as a test case. Primer sequence conservation is shown through a WebLogo representation where the overall height of the stacks indicates the sequence conservation at that position, while the height of symbols within the stacks indicates the relative frequency of nucleic acids at that position. Panel b illustrates the details of ITS regions identified in the complete genomes of species included in the mock sample for which a complete genome was available. ITS sequences without tRNA genes are highlighted in green, while ITS regions encoding tRNA genes are marked in blue. Black and red text indicates forward and reverse strand orientation, respectively.

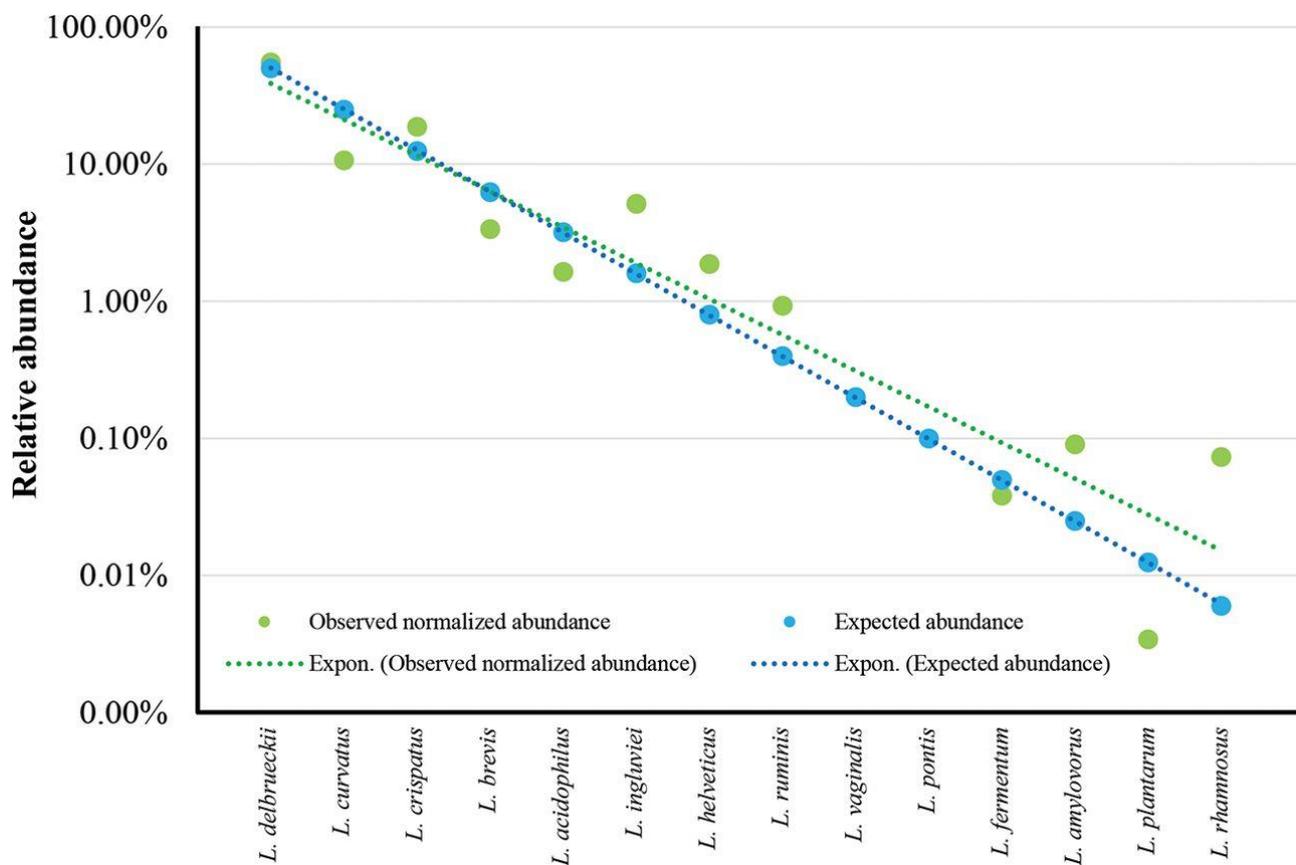


Figure 2: Evaluation of the sensitivity and accuracy of the *Lactobacillus* ITS profiling protocol. The graph shows the expected and observed relative abundance of 14 *Lactobacillus* taxa constituting an artificial sample. An exponential trendline is reported for the expected and observed data.

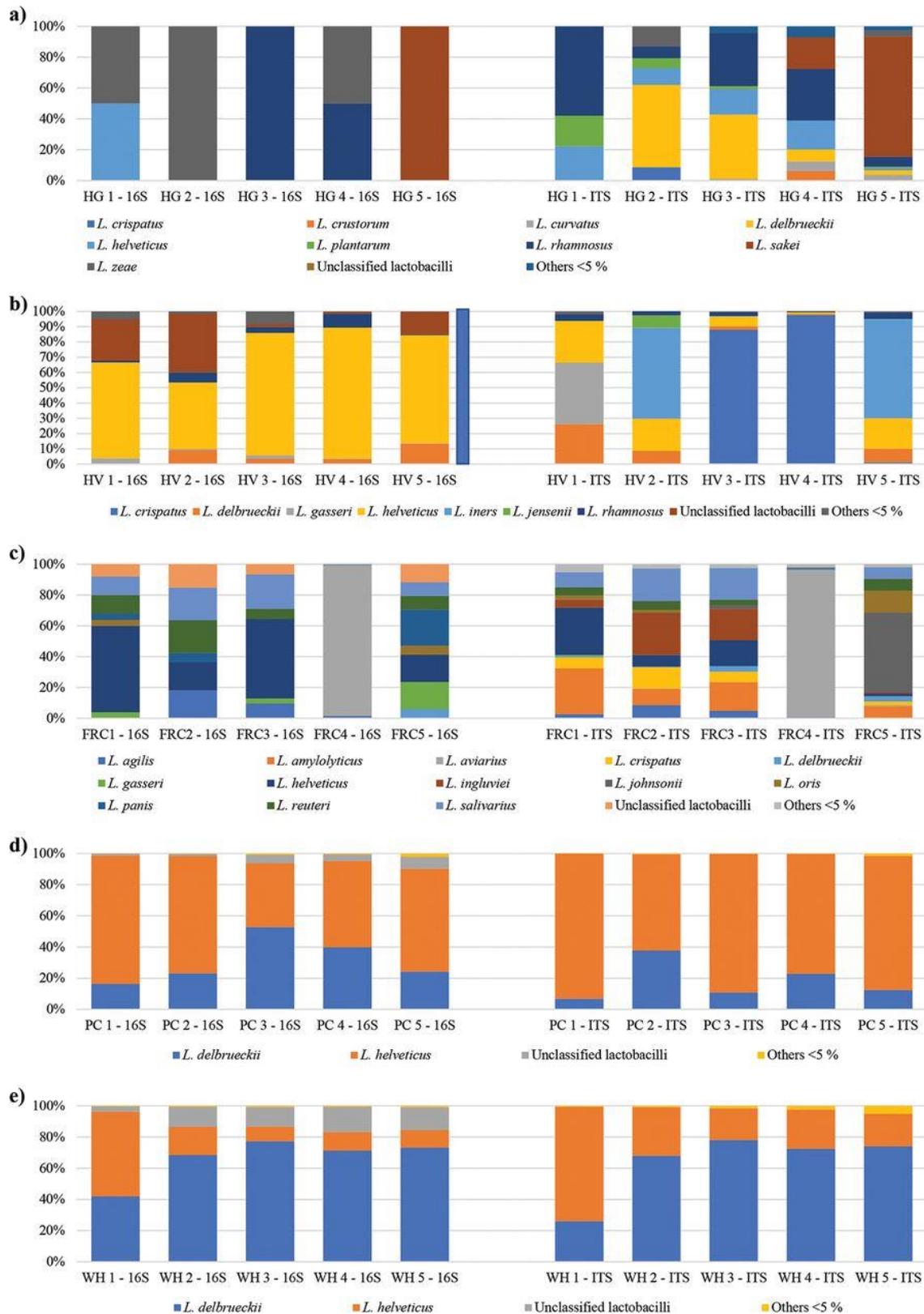


Figure 3: ITS and 16S rRNA gene profiling of *Lactobacillus* species in five ecological niches. The profile of the *Lactobacillus* population obtained for: a) five human faecal samples (HG); b) five human vaginal swab samples (HV); c) five free range chicken faecal samples (FRC); d) five whey samples (WH), and e) five parmesan cheese samples (PC) is depicted in the corresponding bar plots. Only species with relative abundance >5% in at least a sample are reported. Species below 5% are collapsed in “Others <5 %”.

Chapter 10

General Conclusion

Advances in Next generation sequencing analyses: unveiling the biodiversity of complex microbial communities

Thanks to the development of new metagenomic tools, it has become possible to get new insights into microbial biodiversity. In this context, 16S rRNA microbial profiling represents the golden standard to provide a detailed cataloguing of bacterial taxa residing in very different ecosystems. However, the limit of resolution of 16S rRNA microbial profiling is until the genus level. Thus, in order to reach a higher taxonomic resolution, i.e. at species and subspecies level, other molecular markers are needed. Amongst these, the ITS sequences represent a valuable example. We have applied the sequencing of the ITS sequences to provide a detailed profiling of a key mammalian gut commensal such as bifidobacteria. Such approach, known as ITS-bifidobacterial profiling allow an in-depth phylotype-level community profiling of the bifidobacterial distribution across the mammalian branch of the tree of life (Chapter 3). Notably, taxonomical analysis demonstrates that bifidobacterial species are universally distributed across the mammalian branch of the tree of life, with no respect to host's taxonomy or diet, thus revealing that bifidobacteria have evolved a broad colonization potential. Moreover, a comparison between wild relatives and common domesticated animals clearly indicate that domestication of animals has promoted a genetic diversification and an increase in bifidobacterial species richness resulted from a prolonged co-existence with humans (characterized by high bifidobacterial biodiversity).

Remarkably, the ITS profiling data revealed that the genus *Bifidobacterium* may exist of many more species than those currently recognized. We have estimated the existence of 89 novel and yet not characterized (sub)species. Of these, seven have been isolated and genetically characterized in the frame of this PhD thesis, which have been recognized by the International Committee of Bacterial Taxonomy (Chapter 4-5).

In the human context, bifidobacteria are numerically relevant in infants during the first weeks and months following birth. Our findings suggest that this ecological feature appears to be a general trait of bifidobacteria in other mammalian species, re-enforced by direct transmission of bifidobacterial consortia from mother to offspring (Chapter 3). In this context, the evaluation of Sortase Dependent (SD) fimbriome of the genus *Bifidobacterium* has been helpful to understand how bifidobacteria can adhere to the mucosa of the large intestine for the persistence and colonization the human gut (Chapter 6). Our analyses showed that these structures are widely distributed in this group of microorganisms, suggesting a key role in facilitating environmental interactions.

As mentioned before, the implementation of new strategic tools in NGS analysis has been helpful in shrinking knowledge gaps on the microbial content from several environments. In this context, a potentially applied pipeline named Genetic Identity Card (GIC), was established (Chapter 8). This new tool encompasses -omics approaches in order to characterize the bacterial composition of probiotic supplements. In detail, the employment of rRNA-associated sequences, shotgun metagenomics and culturomics approaches, revealed several bacterial contaminations.

Finally, a metagenomic analysis based on the sequencing and analysis of the ITS sequences of lactobacilli, known as lactobacilli ITS profiling, was developed (Chapter 9).

This approach has been used to perform an accurate characterization of the *Lactobacillus* population present in different biological samples, encompassing the stool of human as well as birds, vaginal swabs and cheese, functional foods and probiotics supplement. The ITS-profiling approach represents a new metagenomic tool for species-level profiling of complex lactobacilli communities that complements phylogenetic group assignments that can be obtained from 16S rRNA gene profiling data.

Identification of novel microbial biomarkers

In order to evaluate the differences in the gut microbiota, and to identify possible microbial biomarkers for the early diagnosis of diseases such as malignant transformation, a comparison of mucosa samples of patients with polyps or CRC, versus healthy mucosa, was performed (Chapter 7). In this pilot study, we investigated the microbiota composition directly in polyp-associated mucosal biopsies through a 16S-microbial profiling analysis, and our data showed that the *Romboutsia* genus is present in higher abundance in healthy compared to polyp-associated tissue. These findings were in line with the results obtained from a meta-analysis in which we compared the colonic mucosa with polyp (CMP) samples with CRC samples from publicly available metagenomic datasets. In fact, the relative abundance of *Romboutsia* in CRC tissue was found to be drastically reduced. Remarkably, this microbial taxon is often related to healthy gut microbiota and we may argue that the absence of this microbial genus is a first indicator of an alteration of the microbiota composition. At genus level, we identified five taxa characteristic of healthy mucosa and one taxon typical of CRC-polyp-tissue (CpT) subjects. In detail, healthy samples were shown to contain a higher abundance of U. m. of *Lachnospiraceae* family, *Lachnoclostridium*, *Faecalibacterium*, *Ruminiclostridium* and *Subdoligranulum*, while CpT samples showed high abundance of *Campylobacter*. This latter finding reinforces the correlation of *Campylobacter* genus with the development of intestinal pathologies.

References

- Aagaard, K., Ma, J., Antony, K.M., Ganu, R., Petrosino, J., and Versalovic, J. (2014) The placenta harbors a unique microbiome. *Sci Transl Med* **6**: 237ra265.
- Abe, R., Oda, S., Sadahiro, T., Nakamura, M., Hirayama, Y., Tateishi, Y. et al. (2010) Gram-negative bacteremia induces greater magnitude of inflammatory response than Gram-positive bacteremia. *Crit Care* **14**: R27.
- AlFaleh, K., and Anabrees, J. (2014) Probiotics for prevention of necrotizing enterocolitis in preterm infants. *Cochrane Database Syst Rev*: CD005496.
- Alfano, N., Courtiol, A., Vielgrader, H., Timms, P., Roca, A.L., and Greenwood, A.D. (2015) Variation in koala microbiomes within and between individuals: effect of body region and captivity status. *Sci Rep* **5**: 10189.
- Arnal, M.E., Zhang, J., Messori, S., Bosi, P., Smidt, H., and Lalles, J.P. (2014) Early changes in microbial colonization selectively modulate intestinal enzymes, but not inducible heat shock proteins in young adult Swine. *PLoS One* **9**: e87967.
- Arrieta, M.C., Stiemsma, L.T., Amenyogbe, N., Brown, E.M., and Finlay, B. (2014) The intestinal microbiome in early life: health and disease. *Front Immunol* **5**: 427.
- Arumugam, M., Raes, J., Pelletier, E., Le Paslier, D., Yamada, T., Mende, D.R. et al. (2011) Enterotypes of the human gut microbiome. *Nature* **473**: 174-180.
- Azad, M.B., Konya, T., Maughan, H., Guttman, D.S., Field, C.J., Chari, R.S. et al. (2013) Gut microbiota of healthy Canadian infants: profiles by mode of delivery and infant diet at 4 months. *CMAJ* **185**: 385-394.
- Backhed, F., Ley, R.E., Sonnenburg, J.L., Peterson, D.A., and Gordon, J.I. (2005) Host-bacterial mutualism in the human intestine. *Science* **307**: 1915-1920.
- Bahrndorff, S., Alemu, T., Alemneh, T., and Lund Nielsen, J. (2016) The Microbiome of Animals: Implications for Conservation Biology. *Int J Genomics* **2016**: 5304028.
- Baothman, O.A., Zamzami, M.A., Taher, I., Abubaker, J., and Abu-Farha, M. (2016) The role of Gut Microbiota in the development of obesity and Diabetes. *Lipids Health Dis* **15**: 108.
- Baumgart, D.C., and Sandborn, W.J. (2007) Inflammatory bowel disease: clinical aspects and established and evolving therapies. *Lancet* **369**: 1641-1657.
- Belkaid, Y., and Hand, T.W. (2014) Role of the microbiota in immunity and inflammation. *Cell* **157**: 121-141.
- Ben, X.M., Li, J., Feng, Z.T., Shi, S.Y., Lu, Y.D., Chen, R., and Zhou, X.Y. (2008) Low level of galacto-oligosaccharide in infant formula stimulates growth of intestinal Bifidobacteria and Lactobacilli. *World J Gastroenterol* **14**: 6564-6568.
- Bergmann, H., Rodriguez, J.M., Salminen, S., and Szajewska, H. (2014) Probiotics in human milk and probiotic supplementation in infant nutrition: a workshop report. *Br J Nutr* **112**: 1119-1128.
- Bergstrom, A., Skov, T.H., Bahl, M.I., Roager, H.M., Christensen, L.B., Ejlerskov, K.T. et al. (2014) Establishment of intestinal microbiota during early life: a longitudinal, explorative study of a large cohort of Danish infants. *Appl Environ Microbiol* **80**: 2889-2900.
- Bessegatto, J.A., Paulino, L.R., Lisboa, J.A.N., Alfieri, A.A., Montemor, C.H., Medeiros, L.P. et al. (2017) Changes in the fecal microbiota of beef cattle caused by change in management and the use of virginiamycin as a growth promoter. *Res Vet Sci* **114**: 355-362.
- Bezirtzoglou, E., Tsiotsias, A., and Welling, G.W. (2011) Microbiota profile in feces of breast- and formula-fed newborns by using fluorescence in situ hybridization (FISH). *Anaerobe* **17**: 478-482.
- Biasucci, G., Rubini, M., Riboni, S., Morelli, L., Bessi, E., and Retetangos, C. (2010) Mode of delivery affects the bacterial community in the newborn gut. *Early Hum Dev* **86 Suppl 1**: 13-15.
- Bidart, G.N., Rodriguez-Diaz, J., and Yebra, M.J. (2016) The Extracellular Wall-Bound beta-N-Acetylglucosaminidase from *Lactobacillus casei* Is Involved in the Metabolism of the Human Milk Oligosaccharide Lacto-N-Triose. *Appl Environ Microbiol* **82**: 570-577.
- Biteen, J.S., Blainey, P.C., Cardon, Z.G., Chun, M., Church, G.M., Dorrestein, P.C. et al. (2016) Tools for the Microbiome: Nano and Beyond. *ACS Nano* **10**: 6-37.
- Borenstein, E., Kupiec, M., Feldman, M.W., and Ruppin, E. (2008) Large-scale reconstruction and phylogenetic analysis of metabolic environments. *Proc Natl Acad Sci U S A* **105**: 14482-14487.

- Boyle, R.J., Bath-Hextall, F.J., Leonardi-Bee, J., Murrell, D.F., and Tang, M.L. (2008) Probiotics for treating eczema. *Cochrane Database Syst Rev*: CD006135.
- Braegger, C., Chmielewska, A., Decsi, T., Kolacek, S., Mihatsch, W., Moreno, L. et al. (2011) Supplementation of infant formula with probiotics and/or prebiotics: a systematic review and comment by the ESPGHAN committee on nutrition. *J Pediatr Gastroenterol Nutr* **52**: 238-250.
- Browne, H.P., Forster, S.C., Anonye, B.O., Kumar, N., Neville, B.A., Stares, M.D. et al. (2016) Culturing of 'unculturable' human microbiota reveals novel taxa and extensive sporulation. *Nature* **533**: 543-546.
- Butel, M.J., Suau, A., Campeotto, F., Magne, F., Aires, J., Ferraris, L. et al. (2007) Conditions of bifidobacterial colonization in preterm infants: a prospective analysis. *J Pediatr Gastroenterol Nutr* **44**: 577-582.
- Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K. et al. (2010) QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* **7**: 335-336.
- Chakravorty, S., Helb, D., Burday, M., Connell, N., and Alland, D. (2007) A detailed analysis of 16S ribosomal RNA gene segments for the diagnosis of pathogenic bacteria. *J Microbiol Methods* **69**: 330-339.
- Cheng, Y., Fox, S., Pemberton, D., Hogg, C., Papenfuss, A.T., and Belov, K. (2015) The Tasmanian devil microbiome-implications for conservation and management. *Microbiome* **3**: 76.
- Chichlowski, M., De Lartigue, G., German, J.B., Raybould, H.E., and Mills, D.A. (2012) Bifidobacteria isolated from infants and cultured on human milk oligosaccharides affect intestinal epithelial function. *J Pediatr Gastroenterol Nutr* **55**: 321-327.
- Choi, K.Y., Lee, T.K., and Sul, W.J. (2015) Metagenomic Analysis of Chicken Gut Microbiota for Improving Metabolism and Health of Chickens - A Review. *Asian-Australas J Anim Sci* **28**: 1217-1225.
- Claesson, M.J., Wang, Q., O'Sullivan, O., Greene-Diniz, R., Cole, J.R., Ross, R.P., and O'Toole, P.W. (2010) Comparison of two next-generation sequencing technologies for resolving highly complex microbiota composition using tandem variable 16S rRNA gene regions. *Nucleic Acids Res* **38**: e200.
- Clarridge, J.E., 3rd (2004) Impact of 16S rRNA gene sequence analysis for identification of bacteria on clinical microbiology and infectious diseases. *Clin Microbiol Rev* **17**: 840-862, table of contents.
- Clayton, T.A., Baker, D., Lindon, J.C., Everett, J.R., and Nicholson, J.K. (2009) Pharmacometabonomic identification of a significant host-microbiome metabolic interaction affecting human drug metabolism. *Proc Natl Acad Sci U S A* **106**: 14728-14733.
- Clemente, J.C., Ursell, L.K., Parfrey, L.W., and Knight, R. (2012) The impact of the gut microbiota on human health: an integrative view. *Cell* **148**: 1258-1270.
- Correa, N.B., Peret Filho, L.A., Penna, F.J., Lima, F.M., and Nicoli, J.R. (2005) A randomized formula controlled trial of *Bifidobacterium lactis* and *Streptococcus thermophilus* for prevention of antibiotic-associated diarrhea in infants. *J Clin Gastroenterol* **39**: 385-389.
- de Goffau, M.C., Fuentes, S., van den Bogert, B., Honkanen, H., de Vos, W.M., Welling, G.W. et al. (2014) Aberrant gut microbiota composition at the onset of type 1 diabetes in young children. *Diabetologia* **57**: 1569-1577.
- Dean, F.B., Nelson, J.R., Giesler, T.L., and Lasken, R.S. (2001) Rapid amplification of plasmid and phage DNA using Phi 29 DNA polymerase and multiply-primed rolling circle amplification. *Genome Res* **11**: 1095-1099.
- Deusch, S., Tilocca, B., Camarinha-Silva, A., and Seifert, J. (2015) News in livestock research - use of Omics-technologies to study the microbiota in the gastrointestinal tract of farm animals. *Comput Struct Biotechnol J* **13**: 55-63.
- Dominguez-Bello, M.G., Costello, E.K., Contreras, M., Magris, M., Hidalgo, G., Fierer, N., and Knight, R. (2010) Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns. *Proc Natl Acad Sci U S A* **107**: 11971-11975.
- Duar, R.M., Lin, X.B., Zheng, J., Martino, M.E., Grenier, T., Perez-Munoz, M.E. et al. (2017) Lifestyles in transition: evolution and natural history of the genus *Lactobacillus*. *FEMS Microbiol Rev* **41**: S27-S48.
- Duranti, S., Vivo, V., Zini, I., Milani, C., Mangifesta, M., Anzalone, R. et al. (2018) *Bifidobacterium bifidum* PRL2010 alleviates intestinal ischemia/reperfusion injury. *PLoS One* **13**: e0202670.
- Duranti, S., Gaiani, F., Mancabelli, L., Milani, C., Grandi, A., Bolchi, A. et al. (2016) Elucidating the gut microbiome of ulcerative colitis: bifidobacteria as novel microbial biomarkers. *FEMS Microbiol Ecol* **92**.
- Eckburg, P.B., Bik, E.M., Bernstein, C.N., Purdom, E., Dethlefsen, L., Sargent, M. et al. (2005) Diversity of the human intestinal microbial flora. *Science* **308**: 1635-1638.
- Egan, M., O'Connell Motherway, M., Ventura, M., and van Sinderen, D. (2014a) Metabolism of sialic acid by *Bifidobacterium breve* UCC2003. *Appl Environ Microbiol* **80**: 4414-4426.

- Egan, M., Motherway, M.O., Kilcoyne, M., Kane, M., Joshi, L., Ventura, M., and van Sinderen, D. (2014b) Cross-feeding by *Bifidobacterium breve* UCC2003 during co-cultivation with *Bifidobacterium bifidum* PRL2010 in a mucin-based medium. *BMC Microbiol* **14**: 282.
- Fallani, M., Young, D., Scott, J., Norin, E., Amarri, S., Adam, R. et al. (2010) Intestinal microbiota of 6-week-old infants across Europe: geographic influence beyond delivery mode, breast-feeding, and antibiotics. *J Pediatr Gastroenterol Nutr* **51**: 77-84.
- Favier, C., Neut, C., Mizon, C., Cortot, A., Colombel, J.F., and Mizon, J. (1997) Fecal beta-D-galactosidase production and Bifidobacteria are decreased in Crohn's disease. *Dig Dis Sci* **42**: 817-822.
- Fecteau, M.E., Pitta, D.W., Vecchiarelli, B., Indugu, N., Kumar, S., Gallagher, S.C. et al. (2016) Dysbiosis of the Fecal Microbiota in Cattle Infected with *Mycobacterium avium* subsp. paratuberculosis. *PLoS One* **11**: e0160353.
- Fodor, A.A., DeSantis, T.Z., Wylie, K.M., Badger, J.H., Ye, Y., Hepburn, T. et al. (2012) The "most wanted" taxa from the human microbiome for whole genome sequencing. *PLoS One* **7**: e41294.
- Freilich, S., Kreimer, A., Borenstein, E., Yosef, N., Sharan, R., Gophna, U., and Ruppin, E. (2009) Metabolic-network-driven analysis of bacterial ecological strategies. *Genome Biol* **10**: R61.
- Fukuda, S., Toh, H., Hase, K., Oshima, K., Nakanishi, Y., Yoshimura, K. et al. (2011) Bifidobacteria can protect from enteropathogenic infection through production of acetate. *Nature* **469**: 543-547.
- Gagniere, J., Raisch, J., Veziant, J., Barnich, N., Bonnet, R., Buc, E. et al. (2016) Gut microbiota imbalance and colorectal cancer. *World J Gastroenterol* **22**: 501-518.
- Garrett, W.S. (2015) Cancer and the microbiota. *Science* **348**: 80-86.
- Garrett, W.S., Gordon, J.I., and Glimcher, L.H. (2010) Homeostasis and inflammation in the intestine. *Cell* **140**: 859-870.
- Gerritsen, J., Smidt, H., Rijkers, G.T., and de Vos, W.M. (2011) Intestinal microbiota in human health and disease: the impact of probiotics. *Genes Nutr* **6**: 209-240.
- Gilbert, J.A., and Dupont, C.L. (2011) Microbial metagenomics: beyond the genome. *Ann Rev Mar Sci* **3**: 347-371.
- Gill, S.R., Pop, M., Deboy, R.T., Eckburg, P.B., Turnbaugh, P.J., Samuel, B.S. et al. (2006) Metagenomic analysis of the human distal gut microbiome. *Science* **312**: 1355-1359.
- Gohir, W., Ratcliffe, E.M., and Sloboda, D.M. (2015) Of the bugs that shape us: maternal obesity, the gut microbiome, and long-term disease risk. *Pediatr Res* **77**: 196-204.
- Gomez, D.E., Arroyo, L.G., Costa, M.C., Viel, L., and Weese, J.S. (2017) Characterization of the Fecal Bacterial Microbiota of Healthy and Diarrheic Dairy Calves. *J Vet Intern Med* **31**: 928-939.
- Gong, J., Si, W., Forster, R.J., Huang, R., Yu, H., Yin, Y. et al. (2007) 16S rRNA gene-based analysis of mucosa-associated bacterial community and phylogeny in the chicken gastrointestinal tracts: from crops to ceca. *FEMS Microbiol Ecol* **59**: 147-157.
- Gonia, S., Tuepker, M., Heisel, T., Autran, C., Bode, L., and Gale, C.A. (2015) Human Milk Oligosaccharides Inhibit *Candida albicans* Invasion of Human Premature Intestinal Epithelial Cells. *J Nutr* **145**: 1992-1998.
- Goodman, A.L., Kallstrom, G., Faith, J.J., Reyes, A., Moore, A., Dantas, G., and Gordon, J.I. (2011) Extensive personal human gut microbiota culture collections characterized and manipulated in gnotobiotic mice. *Proc Natl Acad Sci U S A* **108**: 6252-6257.
- Gordon, J.I., Dewey, K.G., Mills, D.A., and Medzhitov, R.M. (2012) The human gut microbiota and undernutrition. *Sci Transl Med* **4**: 137ps112.
- Gritz, E.C., and Bhandari, V. (2015) The human neonatal gut microbiome: a brief review. *Front Pediatr* **3**: 17.
- Gronlund, M.M., Lehtonen, O.P., Eerola, E., and Kero, P. (1999) Fecal microflora in healthy infants born by different methods of delivery: permanent changes in intestinal flora after cesarean delivery. *J Pediatr Gastroenterol Nutr* **28**: 19-25.
- Guaraldi, F., and Salvatori, G. (2012) Effect of breast and formula feeding on gut microbiota shaping in newborns. *Front Cell Infect Microbiol* **2**: 94.
- Gueimonde, M., Sakata, S., Kalliomaki, M., Isolauri, E., Benno, Y., and Salminen, S. (2006) Effect of maternal consumption of lactobacillus GG on transfer and establishment of fecal bifidobacterial microbiota in neonates. *J Pediatr Gastroenterol Nutr* **42**: 166-170.
- Gulden, E., Wong, F.S., and Wen, L. (2015) The gut microbiota and Type 1 Diabetes. *Clin Immunol* **159**: 143-153.
- Hallert, C., Bjorck, I., Nyman, M., Pousette, A., Granno, C., and Svensson, H. (2003) Increasing fecal butyrate in ulcerative colitis patients by diet: controlled pilot study. *Inflamm Bowel Dis* **9**: 116-121.

- Hamady, M., and Knight, R. (2009) Microbial community profiling for human microbiome projects: Tools, techniques, and challenges. *Genome Res* **19**: 1141-1152.
- Heinritz, S.N., Weiss, E., Eklund, M., Aumiller, T., Louis, S., Rings, A. et al. (2016) Intestinal Microbiota and Microbial Metabolites Are Changed in a Pig Model Fed a High-Fat/Low-Fiber or a Low-Fat/High-Fiber Diet. *PLoS One* **11**: e0154329.
- Hill, D.A., Siracusa, M.C., Abt, M.C., Kim, B.S., Kobuley, D., Kubo, M. et al. (2012) Commensal bacteria-derived signals regulate basophil hematopoiesis and allergic inflammation. *Nat Med* **18**: 538-546.
- Hugon, P., Dufour, J.C., Colson, P., Fournier, P.E., Sallah, K., and Raoult, D. (2015) A comprehensive repertoire of prokaryotic species identified in human beings. *Lancet Infect Dis* **15**: 1211-1219.
- Huycke, M.M., and Gaskins, H.R. (2004) Commensal bacteria, redox stress, and colorectal cancer: mechanisms and models. *Exp Biol Med (Maywood)* **229**: 586-597.
- Jackson, R.L., Greiwe, J.S., and Schwen, R.J. (2011) Emerging evidence of the health benefits of S-equol, an estrogen receptor beta agonist. *Nutr Rev* **69**: 432-448.
- Jimenez, E., Fernandez, L., Marin, M.L., Martin, R., Odriozola, J.M., Nueno-Palop, C. et al. (2005) Isolation of commensal bacteria from umbilical cord blood of healthy neonates born by cesarean section. *Curr Microbiol* **51**: 270-274.
- Jinno, S., Toshimitsu, T., Nakamura, Y., Kubota, T., Igoshi, Y., Ozawa, N. et al. (2017) Maternal Prebiotic Ingestion Increased the Number of Fecal Bifidobacteria in Pregnant Women but Not in Their Neonates Aged One Month. *Nutrients* **9**.
- Kapiki, A., Costalos, C., Oikonomidou, C., Triantafyllidou, A., Loukatou, E., and Pertrohilou, V. (2007) The effect of a fructo-oligosaccharide supplemented formula on gut flora of preterm infants. *Early Hum Dev* **83**: 335-339.
- Kerckhoffs, A.P., Samsom, M., van der Rest, M.E., de Vogel, J., Knol, J., Ben-Amor, K., and Akkermans, L.M. (2009) Lower Bifidobacteria counts in both duodenal mucosa-associated and fecal microbiota in irritable bowel syndrome patients. *World J Gastroenterol* **15**: 2887-2892.
- Kerr, C.A., Grice, D.M., Tran, C.D., Bauer, D.C., Li, D., Hendry, P., and Hannan, G.N. (2015) Early life events influence whole-of-life metabolic health via gut microflora and gut permeability. *Crit Rev Microbiol* **41**: 326-340.
- Kim, A. (2015) Dysbiosis: A Review Highlighting Obesity and Inflammatory Bowel Disease. *J Clin Gastroenterol* **49 Suppl 1**: S20-24.
- Kim, H.B., Borewicz, K., White, B.A., Singer, R.S., Sreevatsan, S., Tu, Z.J., and Isaacson, R.E. (2012) Microbial shifts in the swine distal gut in response to the treatment with antimicrobial growth promoter, tylosin. *Proc Natl Acad Sci U S A* **109**: 15485-15490.
- Knol, J., Scholtens, P., Kafka, C., Steenbakkers, J., Gro, S., Helm, K. et al. (2005) Colon microflora in infants fed formula with galacto- and fructo-oligosaccharides: more like breast-fed infants. *J Pediatr Gastroenterol Nutr* **40**: 36-42.
- Kong, F., Zhao, J., Han, S., Zeng, B., Yang, J., Si, X. et al. (2014) Characterization of the gut microbiota in the red panda (*Ailurus fulgens*). *PLoS One* **9**: e87885.
- Kostic, A.D., Chun, E., Robertson, L., Glickman, J.N., Gallini, C.A., Michaud, M. et al. (2013) *Fusobacterium nucleatum* potentiates intestinal tumorigenesis and modulates the tumor-immune microenvironment. *Cell Host Microbe* **14**: 207-215.
- Kunz, C., Rudloff, S., Baier, W., Klein, N., and Strobel, S. (2000) Oligosaccharides in human milk: structural, functional, and metabolic aspects. *Annu Rev Nutr* **20**: 699-722.
- Lagier, J.C., Armougom, F., Million, M., Hugon, P., Pagnier, I., Robert, C. et al. (2012) Microbial culturomics: paradigm shift in the human gut microbiome study. *Clin Microbiol Infect* **18**: 1185-1193.
- Lagier, J.C., Khelaifia, S., Alou, M.T., Ndongo, S., Dione, N., Hugon, P. et al. (2016) Culture of previously uncultured members of the human gut microbiota by culturomics. *Nat Microbiol* **1**: 16203.
- Lahtinen, S.J., Boyle, R.J., Kivivuori, S., Oppedisano, F., Smith, K.R., Robins-Browne, R. et al. (2009) Prenatal probiotic administration can influence Bifidobacterium microbiota development in infants at high risk of allergy. *J Allergy Clin Immunol* **123**: 499-501.
- Lee, Y.K., and Mazmanian, S.K. (2010) Has the microbiota played a critical role in the evolution of the adaptive immune system? *Science* **330**: 1768-1773.
- Li, Q., Wang, C., Tang, C., Li, N., and Li, J. (2012) Molecular-phylogenetic characterization of the microbiota in ulcerated and non-ulcerated regions in the patients with Crohn's disease. *PLoS One* **7**: e34939.
- Lin, A.E., Autran, C.A., Szyszka, A., Escajadillo, T., Huang, M., Godula, K. et al. (2017) Human milk oligosaccharides inhibit growth of group B Streptococcus. *J Biol Chem* **292**: 11243-11249.

- Lin, H.C., Su, B.H., Chen, A.C., Lin, T.W., Tsai, C.H., Yeh, T.F., and Oh, W. (2005) Oral probiotics reduce the incidence and severity of necrotizing enterocolitis in very low birth weight infants. *Pediatrics* **115**: 1-4.
- Lloyd-Price, J., Abu-Ali, G., and Huttenhower, C. (2016) The healthy human microbiome. *Genome Med* **8**: 51.
- Locascio, R.G., Ninonuevo, M.R., Kronewitter, S.R., Freeman, S.L., German, J.B., Lebrilla, C.B., and Mills, D.A. (2009) A versatile and scalable strategy for glycoprofiling bifidobacterial consumption of human milk oligosaccharides. *Microb Biotechnol* **2**: 333-342.
- Loftus, E.V., Jr. (2004) Clinical epidemiology of inflammatory bowel disease: Incidence, prevalence, and environmental influences. *Gastroenterology* **126**: 1504-1517.
- Looft, T., Allen, H.K., Cantarel, B.L., Levine, U.Y., Bayles, D.O., Alt, D.P. et al. (2014) Bacteria, phages and pigs: the effects of in-feed antibiotics on the microbiome at different gut locations. *ISME J* **8**: 1566-1576.
- Lozupone, C.A., Stombaugh, J.I., Gordon, J.I., Jansson, J.K., and Knight, R. (2012) Diversity, stability and resilience of the human gut microbiota. *Nature* **489**: 220-230.
- Lugli, G.A., Milani, C., Mancabelli, L., Turrioni, F., Ferrario, C., Duranti, S. et al. (2017) Ancient bacteria of the Otzi's microbiome: a genomic tale from the Copper Age. *Microbiome* **5**: 5.
- Mancabelli, L., Milani, C., Lugli, G.A., Turrioni, F., Ferrario, C., van Sinderen, D., and Ventura, M. (2017a) Meta-analysis of the human gut microbiome from urbanized and pre-agricultural populations. *Environ Microbiol* **19**: 1379-1390.
- Mancabelli, L., Milani, C., Lugli, G.A., Turrioni, F., Cocconi, D., van Sinderen, D., and Ventura, M. (2017b) Identification of universal gut microbial biomarkers of common human intestinal diseases by meta-analysis. *FEMS Microbiol Ecol* **93**.
- Mangin, I., Bonnet, R., Seksik, P., Rigottier-Gois, L., Sutren, M., Bouhnik, Y. et al. (2004) Molecular inventory of faecal microflora in patients with Crohn's disease. *FEMS Microbiol Ecol* **50**: 25-36.
- Manichanh, C., Rigottier-Gois, L., Bonnaud, E., Gloux, K., Pelletier, E., Frangeul, L. et al. (2006) Reduced diversity of faecal microbiota in Crohn's disease revealed by a metagenomic approach. *Gut* **55**: 205-211.
- Mao, S., Zhang, M., Liu, J., and Zhu, W. (2015) Characterising the bacterial microbiota across the gastrointestinal tracts of dairy cattle: membership and potential function. *Sci Rep* **5**: 16116.
- Marchesi, J.R., Holmes, E., Khan, F., Kochhar, S., Scanlan, P., Shanahan, F. et al. (2007) Rapid and noninvasive metabonomic characterization of inflammatory bowel disease. *J Proteome Res* **6**: 546-551.
- Maslowski, K.M., Vieira, A.T., Ng, A., Kranich, J., Sierro, F., Yu, D. et al. (2009) Regulation of inflammatory responses by gut microbiota and chemoattractant receptor GPR43. *Nature* **461**: 1282-1286.
- Matsuoka, K., and Kanai, T. (2015) The gut microbiota and inflammatory bowel disease. *Semin Immunopathol* **37**: 47-55.
- McLean, J.S., Lombardo, M.J., Badger, J.H., Edlund, A., Novotny, M., Yee-Greenbaum, J. et al. (2013) Candidate phylum TM6 genome recovered from a hospital sink biofilm provides genomic insights into this uncultivated phylum. *Proc Natl Acad Sci U S A* **110**: E2390-2399.
- Milani, C., Turrioni, F., Duranti, S., Lugli, G.A., Mancabelli, L., Ferrario, C. et al. (2016) Genomics of the Genus *Bifidobacterium* Reveals Species-Specific Adaptation to the Glycan-Rich Gut Environment. *Appl Environ Microbiol* **82**: 980-991.
- Milani, C., Lugli, G.A., Turrioni, F., Mancabelli, L., Duranti, S., Viappiani, A. et al. (2014) Evaluation of bifidobacterial community composition in the human gut by means of a targeted amplicon sequencing (ITS) protocol. *FEMS Microbiol Ecol* **90**: 493-503.
- Milani, C., Hevia, A., Foroni, E., Duranti, S., Turrioni, F., Lugli, G.A. et al. (2013) Assessing the fecal microbiota: an optimized ion torrent 16S rRNA gene-based analysis protocol. *PLoS One* **8**: e68739.
- Milani, C., Duranti, S., Bottacini, F., Casey, E., Turrioni, F., Mahony, J. et al. (2017) The First Microbial Colonizers of the Human Gut: Composition, Activities, and Health Implications of the Infant Gut Microbiota. *Microbiol Mol Biol Rev* **81**.
- Mitsou, E.K., Kirtzalidou, E., Oikonomou, I., Liosis, G., and Kyriacou, A. (2008) Fecal microflora of Greek healthy neonates. *Anaerobe* **14**: 94-101.
- Morgan, X.C., Kabackchiev, B., Waldron, L., Tyler, A.D., Tickle, T.L., Milgrom, R. et al. (2015) Associations between host gene expression, the mucosal microbiome, and clinical outcome in the pelvic pouch of patients with inflammatory bowel disease. *Genome Biol* **16**: 67.
- Morris, B.E., Henneberger, R., Huber, H., and Moissl-Eichinger, C. (2013) Microbial syntrophy: interaction for the common good. *FEMS Microbiol Rev* **37**: 384-406.

- Nagpal, R., Tsuji, H., Takahashi, T., Kawashima, K., Nagata, S., Nomoto, K., and Yamashiro, Y. (2016) Sensitive Quantitative Analysis of the Meconium Bacterial Microbiota in Healthy Term Infants Born Vaginally or by Cesarean Section. *Front Microbiol* **7**: 1997.
- Nemoto, H., Kataoka, K., Ishikawa, H., Ikata, K., Arimochi, H., Iwasaki, T. et al. (2012) Reduced diversity and imbalance of fecal microbiota in patients with ulcerative colitis. *Dig Dis Sci* **57**: 2955-2964.
- Niu, Q., Li, P., Hao, S., Zhang, Y., Kim, S.W., Li, H. et al. (2015) Dynamic distribution of the gut microbiota and the relationship with apparent crude fiber digestibility and growth stages in pigs. *Sci Rep* **5**: 9938.
- Pande, S., Shitut, S., Freund, L., Westermann, M., Bertels, F., Colesie, C. et al. (2015) Metabolic cross-feeding via intercellular nanotubes among bacteria. *Nat Commun* **6**: 6238.
- Panigrahi, P., Parida, S., Pradhan, L., Mohapatra, S.S., Misra, P.R., Johnson, J.A. et al. (2008) Long-term colonization of a *Lactobacillus plantarum* synbiotic preparation in the neonatal gut. *J Pediatr Gastroenterol Nutr* **47**: 45-53.
- Partty, A., and Isolauri, E. (2012) Gut microbiota and infant distress - the association between compositional development of the gut microbiota and fussing and crying in early infancy. *Microb Ecol Health Dis* **23**.
- Partty, A., Luoto, R., Kalliomaki, M., Salminen, S., and Isolauri, E. (2013) Effects of early prebiotic and probiotic supplementation on development of gut microbiota and fussing and crying in preterm infants: a randomized, double-blind, placebo-controlled trial. *J Pediatr* **163**: 1272-1277 e1271-1272.
- Penders, J., Thijs, C., Vink, C., Stelma, F.F., Snijders, B., Kummeling, I. et al. (2006) Factors influencing the composition of the intestinal microbiota in early infancy. *Pediatrics* **118**: 511-521.
- Petrova, M.I., Lievens, E., Malik, S., Imholz, N., and Lebeer, S. (2015) *Lactobacillus* species as biomarkers and agents that can promote various aspects of vaginal health. *Front Physiol* **6**: 81.
- Phelan, V.V., Liu, W.T., Pogliano, K., and Dorrestein, P.C. (2011) Microbial metabolic exchange--the chemotype-to-phenotype link. *Nat Chem Biol* **8**: 26-35.
- Pourabedin, M., and Zhao, X. (2015) Prebiotics and gut microbiota in chickens. *FEMS Microbiol Lett* **362**: fnv122.
- Rajilic-Stojanovic, M., and de Vos, W.M. (2014) The first 1000 cultured species of the human gastrointestinal microbiota. *FEMS Microbiol Rev* **38**: 996-1047.
- Rea, D., Coppola, G., Palma, G., Barbieri, A., Luciano, A., Del Prete, P. et al. (2018) Microbiota effects on cancer: from risks to therapies. *Oncotarget* **9**: 17915-17927.
- Relman, D.A. (2012) The human microbiome: ecosystem resilience and health. *Nutr Rev* **70 Suppl 1**: S2-9.
- Reshef, L., Kovacs, A., Ofer, A., Yahav, L., Maharshak, N., Keren, N. et al. (2015) Pouch Inflammation Is Associated With a Decrease in Specific Bacterial Taxa. *Gastroenterology* **149**: 718-727.
- Rinke, C., Schwientek, P., Sczyrba, A., Ivanova, N.N., Anderson, I.J., Cheng, J.F. et al. (2013) Insights into the phylogeny and coding potential of microbial dark matter. *Nature* **499**: 431-437.
- Rinne, M.M., Gueimonde, M., Kalliomaki, M., Hoppu, U., Salminen, S.J., and Isolauri, E. (2005) Similar bifidogenic effects of prebiotic-supplemented partially hydrolyzed infant formula and breastfeeding on infant gut microbiota. *FEMS Immunol Med Microbiol* **43**: 59-65.
- Roberfroid, M., Gibson, G.R., Hoyles, L., McCartney, A.L., Rastall, R., Rowland, I. et al. (2010) Prebiotic effects: metabolic and health benefits. *Br J Nutr* **104 Suppl 2**: S1-63.
- Sanderson, I.R. (2004) Short chain fatty acid regulation of signaling genes expressed by the intestinal epithelium. *J Nutr* **134**: 2450S-2454S.
- Scalabrin, D.M., Mitmesser, S.H., Welling, G.W., Harris, C.L., Marunycz, J.D., Walker, D.C. et al. (2012) New prebiotic blend of polydextrose and galacto-oligosaccharides has a bifidogenic effect in young infants. *J Pediatr Gastroenterol Nutr* **54**: 343-352.
- Scheppach, W., Sommer, H., Kirchner, T., Paganelli, G.M., Bartram, P., Christl, S. et al. (1992) Effect of butyrate enemas on the colonic mucosa in distal ulcerative colitis. *Gastroenterology* **103**: 51-56.
- Schloss, P.D., Westcott, S.L., Ryabin, T., Hall, J.R., Hartmann, M., Hollister, E.B. et al. (2009) Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* **75**: 7537-7541.
- Segain, J.P., Raingeard de la Bletiere, D., Bourreille, A., Leray, V., Gervois, N., Rosales, C. et al. (2000) Butyrate inhibits inflammatory responses through NFkappaB inhibition: implications for Crohn's disease. *Gut* **47**: 397-403.
- Segata, N., Waldron, L., Ballarini, A., Narasimhan, V., Jousson, O., and Huttenhower, C. (2012) Metagenomic microbial community profiling using unique clade-specific marker genes. *Nat Methods* **9**: 811-814.
- Sender, R., Fuchs, S., and Milo, R. (2016) Revised Estimates for the Number of Human and Bacteria Cells in the Body. *PLoS Biol* **14**: e1002533.

- Setchell, K.D., and Clerici, C. (2010) Equol: history, chemistry, and formation. *J Nutr* **140**: 1355S-1362S.
- Shean, G., and Faia, C. (1975) Autonomic control, selective attention and schizophrenic subtype. *J Nerv Ment Dis* **160**: 176-181.
- Sierra, C., Bernal, M.J., Blasco, J., Martinez, R., Dalmau, J., Ortuno, I. et al. (2015) Prebiotic effect during the first year of life in healthy infants fed formula containing GOS as the only prebiotic: a multicentre, randomised, double-blind and placebo-controlled trial. *Eur J Nutr* **54**: 89-99.
- Singh, R.K., Chang, H.W., Yan, D., Lee, K.M., Ucmak, D., Wong, K. et al. (2017) Influence of diet on the gut microbiome and implications for human health. *J Transl Med* **15**: 73.
- Song, M., Xia, B., and Li, J. (2006) Effects of topical treatment of sodium butyrate and 5-aminosalicylic acid on expression of trefoil factor 3, interleukin 1beta, and nuclear factor kappaB in trinitrobenzene sulphonic acid induced colitis in rats. *Postgrad Med J* **82**: 130-135.
- Spor, A., Koren, O., and Ley, R. (2011) Unravelling the effects of the environment and host genotype on the gut microbiome. *Nat Rev Microbiol* **9**: 279-290.
- Stojanovic, N., Plecas, D., and Plesinac, S. (2012) Normal vaginal flora, disorders and application of probiotics in pregnancy. *Arch Gynecol Obstet* **286**: 325-332.
- Sundquist, A., Bigdeli, S., Jalili, R., Druzin, M.L., Waller, S., Pullen, K.M. et al. (2007) Bacterial flora-typing with targeted, chip-based Pyrosequencing. *BMC Microbiol* **7**: 108.
- Thursby, E., and Juge, N. (2017) Introduction to the human gut microbiota. *Biochem J* **474**: 1823-1836.
- Tojo, R., Suarez, A., Clemente, M.G., de los Reyes-Gavilan, C.G., Margolles, A., Gueimonde, M., and Ruas-Madiedo, P. (2014) Intestinal microbiota in health and disease: role of bifidobacteria in gut homeostasis. *World J Gastroenterol* **20**: 15163-15176.
- Topping, D.L., and Clifton, P.M. (2001) Short-chain fatty acids and human colonic function: roles of resistant starch and nonstarch polysaccharides. *Physiol Rev* **81**: 1031-1064.
- Turroni, F., Ventura, M., Butto, L.F., Duranti, S., O'Toole, P.W., Motherway, M.O., and van Sinderen, D. (2014) Molecular dialogue between the human gut microbiota and the host: a Lactobacillus and Bifidobacterium perspective. *Cell Mol Life Sci* **71**: 183-203.
- Turroni, F., Ozcan, E., Milani, C., Mancabelli, L., Viappiani, A., van Sinderen, D. et al. (2015) Glycan cross-feeding activities between bifidobacteria under in vitro conditions. *Front Microbiol* **6**: 1030.
- Turroni, F., Peano, C., Pass, D.A., Feroni, E., Severgnini, M., Claesson, M.J. et al. (2012) Diversity of bifidobacteria within the infant gut microbiota. *PLoS One* **7**: e36957.
- Turroni, F., Bottacini, F., Feroni, E., Mulder, I., Kim, J.H., Zomer, A. et al. (2010) Genome analysis of Bifidobacterium bifidum PRL2010 reveals metabolic pathways for host-derived glycan foraging. *Proc Natl Acad Sci U S A* **107**: 19514-19519.
- Turroni, F., Serafini, F., Feroni, E., Duranti, S., O'Connell Motherway, M., Taverniti, V. et al. (2013) Role of sortase-dependent pili of Bifidobacterium bifidum PRL2010 in modulating bacterium-host interactions. *Proc Natl Acad Sci U S A* **110**: 11151-11156.
- van der Wielen, P.W., Keuzenkamp, D.A., Lipman, L.J., van Knapen, F., and Biesterveld, S. (2002) Spatial and temporal variation of the intestinal bacterial community in commercially raised broiler chickens during growth. *Microb Ecol* **44**: 286-293.
- van Nuenen, M.H., Venema, K., van der Woude, J.C., and Kuipers, E.J. (2004) The metabolic activity of fecal microbiota from healthy individuals and patients with inflammatory bowel disease. *Dig Dis Sci* **49**: 485-491.
- Vandenplas, Y., Zakharova, I., and Dmitrieva, Y. (2015) Oligosaccharides in infant formula: more evidence to validate the role of prebiotics. *Br J Nutr* **113**: 1339-1344.
- Ventura, M., Turroni, F., Canchaya, C., Vaughan, E.E., O'Toole, P.W., and van Sinderen, D. (2009) Microbial diversity in the human intestine and novel insights from metagenomics. *Front Biosci (Landmark Ed)* **14**: 3214-3221.
- Walker, A.W., Duncan, S.H., Louis, P., and Flint, H.J. (2014) Phylogeny, culturing, and metagenomics of the human gut microbiota. *Trends Microbiol* **22**: 267-274.
- Walter, J. (2008) Ecological role of lactobacilli in the gastrointestinal tract: implications for fundamental and biomedical research. *Appl Environ Microbiol* **74**: 4985-4996.
- Wang, L., Lilburn, M., and Yu, Z. (2016) Intestinal Microbiota of Broiler Chickens As Affected by Litter Management Regimens. *Front Microbiol* **7**: 593.
- Wu, S., Rhee, K.J., Albesiano, E., Rabizadeh, S., Wu, X., Yen, H.R. et al. (2009) A human colonic commensal promotes colon tumorigenesis via activation of T helper type 17 T cell responses. *Nat Med* **15**: 1016-1022.
- Yang, I., Corwin, E.J., Brennan, P.A., Jordan, S., Murphy, J.R., and Dunlop, A. (2016) The Infant Microbiome: Implications for Infant Health and Neurocognitive Development. *Nurs Res* **65**: 76-88.

Zitvogel, L., Daillere, R., Roberti, M.P., Routy, B., and Kroemer, G. (2017) Anticancer effects of the microbiome and its products. *Nat Rev Microbiol* **15**: 465-478.

Zitvogel, L., Galluzzi, L., Viaud, S., Vetizou, M., Daillere, R., Merad, M., and Kroemer, G. (2015) Cancer and the gut microbiota: an unexpected link. *Sci Transl Med* **7**: 271ps271.

Publications

1. Lugli GA[‡], **Mangifesta M[‡]**, Mancabelli M, Milani C, Turrone F, Viappiani A, van Sinderen D, Ventura M. **“Compositional assessment of bacterial communities in probiotic supplements by means of metagenomic techniques.”** Int J Food Microbiol. 2019 Jan 21;294:1-9. doi: 10.1016/j.ijfoodmicro.2019.01.011.
2. Lugli GA, Mancino W, Milani C, Duranti S, Mancabelli L, Napoli S, **Mangifesta M**, Viappiani A, Anzalone R, Longhi G, van Sinderen D, Ventura M, Turrone F. **“Dissecting the evolutionary development of the *Bifidobacterium animalis* species through comparative genomics analyses.”** Appl Environ Microbiol. 2019 Feb 1. pii: AEM.02806-18. doi: 10.1128/AEM.02806-18.
3. Alessandri G, Milani C, Mancabelli L, **Mangifesta M**, Lugli GA, Viappiani A, Duranti S, Turrone F, Ossiprandi MC, van Sinderen D, Ventura M. **“Metagenomic dissection of the canine gut microbiota: insights into taxonomic, metabolic and nutritional features.”** Environ Microbiol. 2019 Jan 24. doi: 10.1111/1462-2920.14540.
4. **Mangifesta M**, Mancabelli L, Milani C, Gaiani F, de'Angelis N, de'Angelis GL, van Sinderen D, Ventura M, Turrone F. **“Mucosal microbiota of intestinal polyps reveals putative biomarkers of colorectal cancer.”** Sci Rep. 2018 Sep 18. doi: 10.1038/s41598-018-32413-2.
5. Duranti S, Vivo V, Zini I, Milani C, **Mangifesta M**, Anzalone R, Mancabelli L, Viappiani A, Cantoni AM, Barocelli E, van Sinderen D, Bertoni S, Turrone F. **“*Bifidobacterium bifidum* PRL2010 alleviates intestinal ischemia/reperfusion injury.”** PLoS One. 2018 Aug 30. doi: 10.1371/journal.pone.0202670. eCollection 2018.
6. Milani C, Casey E, Lugli GA, Moore R, Kaczorowska J, Feehily C, **Mangifesta M**, Mancabelli L, Duranti S, Turrone F, Bottacini F, Mahony J, Cotter PD, McAuliffe FM, van Sinderen D, Ventura M. **“Tracing mother-infant transmission of bacteriophages by means of a novel analytical tool for shotgun metagenomic datasets: METAnnotatorX.”** Microbiome. 2018 Aug 20. doi: 10.1186/s40168-018-0527-z.
7. Ferretti P, Pasolli E, Tett A, Asnicar F, Gorfer V, Fedi S, Armanini F, Truong DT, Manara S, Zolfo M, Beghini F, Bertorelli R, De Sanctis V, Bariletti I, Canto R, Clementi R, Cologna M, Crifò T, Cusumano G, Gottardi S, Innamorati C, Masè C, Postai D, Savoia D, Duranti S, Lugli GA, Mancabelli L, Turrone F, Ferrario C, Milani C, **Mangifesta M**, Anzalone R, Viappiani A, Yassour M, Vlamakis H, Xavier R, Collado CM, Koren O, Tateo S, Soffiati M, Pedrotti A, Ventura M, Huttenhower C, Bork P, Segata N. **“Mother-to-Infant Microbial Transmission from Different Body Sites Shapes the Developing Infant Gut Microbiome.”** Cell Host Microbe. 2018 Jul 11. doi: 10.1016/j.chom.2018.06.005.
8. Milani C[‡], Duranti S[‡], **Mangifesta M[‡]**, Lugli GA, Turrone F, Mancabelli L, Viappiani A, Anzalone R, Alessandri G, Ossiprandi MC, van Sinderen D, Ventura M. **“Phylotype-Level Profiling of Lactobacilli in Highly Complex Environments by Means of an Internal Transcribed Spacer-Based Metagenomic Approach.”** Appl Environ Microbiol. 2018 Jul 2. doi: 10.1128/AEM.00706-18.
9. Ticinesi A, Milani C, Guerra A, Allegri F, Lauretani F, Nouvenne A, Mancabelli L, Lugli GA, Turrone F, Duranti S, **Mangifesta M**, Viappiani A, Ferrario C, Dodi R, Dall'Asta M, Del Rio D, Ventura M, Meschi T. **“Understanding the gut-kidney axis in nephrolithiasis: an analysis of the gut microbiota composition and functionality of stone formers.”** Gut. 2018 Apr 28. doi: 10.1136/gutjnl-2017-315734.

10. Lugli GA‡, **Mangifesta M‡**, Duranti S‡, Anzalone R, Milani C, Mancabelli L, Alessandri G, Turrone F, Ossiprandi MC, van Sinderen D, Ventura M. **“Phylogenetic classification of six novel species belonging to the genus *Bifidobacterium* comprising *Bifidobacterium anseris* sp. nov., *Bifidobacterium criceti* sp. nov., *Bifidobacterium imperatoris* sp. nov., *Bifidobacterium italicum* sp. nov., *Bifidobacterium margollesii* sp. nov. and *Bifidobacterium parmae* sp. nov.”** Syst Appl Microbiol. 2018 May. doi: 10.1016/j.syapm.2018.01.002.
11. Lugli GA, Milani C, Duranti S, Mancabelli L, **Mangifesta M**, Turrone F, Viappiani A, van Sinderen D, Ventura M. **“Tracking the Taxonomy of the Genus *Bifidobacterium* Based on a Phylogenomic Approach.”** Appl Environ Microbiol. 2018 Jan 31. doi: 10.1128/AEM.02249-17.
12. Ferrario C, Alessandri G, Mancabelli L, Gering E, **Mangifesta M**, Milani C, Lugli GA, Viappiani A, Duranti S, Turrone F, Ossiprandi MC, Hiyashi R, Mackie R, van Sinderen D, Ventura M. **“Untangling the cecal microbiota of feral chickens by culturomic and metagenomic analyses”** Environ Microbiol. 2017 Oct 2. doi: 10.1111/1462-2920.13943.
13. Ferrario C, Statello R, Carnevali L, Mancabelli L, Milani C, **Mangifesta M**, Duranti S, Lugli GA, Jimenez B, Lodge S, Viappiani A, Alessandri G, Dall’Asta M, Del Rio D, Sgoifo A, van Sinderen D, Ventura M, Turrone F. **“How to Feed the Mammalian Gut Microbiota: Bacterial and Metabolic Modulation by Dietary Fibers”** Front. Microbiol., 12 September 2017 | <https://doi.org/10.3389/fmicb.2017.01749>. eCollection 2017.
14. Ticinesi A, Milani C, Lauretani F, Nouvenne A, Mancabelli L, Lugli GA, Turrone F, Duranti S, **Mangifesta M**, Viappiani A, Ferrario C, Maggio M, Ventura M, Meschi T. **“Gut microbiota composition is associated with polypharmacy in elderly hospitalized patients.”** Sci Rep. 2017 Sep 11;7(1):11102. doi: 10.1038/s41598-017-10734-y.
15. Duranti S‡, **Mangifesta M‡**, Lugli GA‡, Turrone F, Anzalone R, Milani C, Mancabelli L, Ossiprandi MC, Ventura M. **“*Bifidobacterium vansinderenii* sp. nov., isolated from faeces of emperor tamarin (*Saguinus imperator*).”** Int J Syst Evol Microbiol. 2017 Sep 12. doi: 10.1099/ijsem.0.002243.
16. Mancabelli L, Milani C, Lugli GA, Turrone F, **Mangifesta M**, Viappiani A, Ticinesi A, Nouvenne A, Meschi T, van Sinderen D, Ventura M. **“Unveiling the gut microbiota composition and functionality associated with constipation through metagenomic analyses.”** Sci Rep. 2017 Aug 29;7(1):9879. doi: 10.1038/s41598-017-10663-w.
17. Milani C, **Mangifesta M**, Mancabelli L, Lugli GA, James K, Duranti S, Turrone F, Ferrario C, Ossiprandi MC, van Sinderen D, Ventura M. **“Unveiling bifidobacterial biogeography across the mammalian branch of the tree of life.”** ISME J. 2017 Aug 22. doi: 10.1038/ismej.2017.138.
18. Lugli GA, Milani C, Turrone F, Duranti S, Mancabelli L, **Mangifesta M**, Ferrario C, Modesto M, Mattarelli P, Jiří K, van Sinderen D, Ventura M. **“Comparative genomic and phylogenomic analyses of the *Bifidobacteriaceae* family.”** BMC Genomics. 2017 Aug 1;18(1):568. doi: 10.1186/s12864-017-3955-4.
19. Milani C, **Mangifesta M**, Mancabelli L, Lugli GA, Mancino W, Viappiani A, Faccini A, van Sinderen D, Ventura M, Turrone F. **“The Sortase-Dependent Fimbriome of the Genus *Bifidobacterium*: Extracellular Structures with Potential To Modulate Microbe-Host Dialogue.”** Appl Environ Microbiol. 2017 Sep 15;83(19). pii: e01295-17. doi: 10.1128/AEM.01295-17.
20. Duranti S, Lugli GA, Mancabelli L, Armanini F, Turrone F, James K, Ferretti P, Gorfer V, Ferrario C, Milani C, **Mangifesta M**, Anzalone R, Zolfo M, Viappiani A, Pasolli E, Bariletti I, Canto R, Clementi R, Cologna M, Crifò T, Cusumano G, Fedi S, Gottardi S, Innamorati C, Masè C, Postai D, Savoi D, Soffiati M, Tateo S, Pedrotti A, Segata N, van Sinderen D, Ventura M. **“Maternal inheritance of bifidobacterial communities and bifidophages in infants through vertical transmission.”** Microbiome. 2017 Jun 26;5(1):66. doi: 10.1186/s40168-017-0282-6.

21. Ferrario C, Lugli GA, Ossiprandi MC, Turrone F, Milani C, Duranti S, Mancabelli L, **Mangifesta M**, Alessandri G, van Sinderen D, Ventura M. **“Next generation sequencing-based multigene panel for high throughput detection of food-borne pathogens.”** Int J Food Microbiol. 2017 Sep 1. doi: 10.1016/j.ijfoodmicro.2017.05.001.
22. Duranti S, Lugli GA, Mancabelli L, Turrone F, Milani C, **Mangifesta M**, Ferrario C, Anzalone R, Viappiani A, van Sinderen D, Ventura M. **“Prevalence of Antibiotic Resistance Genes among Human Gut-Derived Bifidobacteria.”** Appl Environ Microbiol. 2017 Jan 17;83(3). pii: e02894-16. doi: 10.1128/AEM.02894-16.
23. Duranti S, Gaiani F, Mancabelli L, Milani C, Grandi A, Bolchi A, Santoni A, Lugli GA, Ferrario C, **Mangifesta M**, Viappiani A, Bertoni S, Vivo V, Serafini F, Barbaro MR, Fugazza A, Barbara G, Gioiosa L, Palanza P, Cantoni AM, de'Angelis GL, Barocelli E, de'Angelis N, van Sinderen D, Ventura M, Turrone F. **“Elucidating the gut microbiome of ulcerative colitis: bifidobacteria as novel microbial biomarkers.”** FEMS Microbiol Ecol. 2016 Dec. doi: 10.1093/femsec/fiw191.
24. Milani C, Ticinesi A, Gerritsen J, Nouvenne A, Lugli GA, Mancabelli L, Turrone F, Duranti S, **Mangifesta M**, Viappiani A, Ferrario C, Maggio M, Lauretani F, De Vos W, van Sinderen D, Meschi T, Ventura M. **“Gut microbiota composition and *Clostridium difficile* infection in hospitalized elderly individuals: a metagenomic study.”** Sci Rep. 2016 May 11;6:25945. doi: 10.1038/srep25945.
25. Mancabelli L, Ferrario C, Milani C, **Mangifesta M**, Turrone F, Duranti S, Lugli GA, Viappiani A, Ossiprandi MC, van Sinderen D, Ventura M. **“Insights into the biodiversity of the gut microbiota of broiler chickens.”** Environ Microbiol. 2016 Apr 30. doi: 10.1111/1462-2920.13363.
26. Duranti S, Milani C, Lugli GA, Mancabelli L, Turrone F, Ferrario C, **Mangifesta M**, Viappiani A, Sánchez B, Margolles A, van Sinderen D, Ventura M. **“Evaluation of genetic diversity among strains of the human gut commensal *Bifidobacterium adolescentis*.”** Sci Rep. 2016 Apr 1;6:23971. doi: 10.1038/srep23971.
27. Ferrario C, Milani C, Mancabelli L, Lugli GA, Duranti S, **Mangifesta M**, Viappiani A, Turrone F, Margolles A, Ruas-Madiedo P, van Sinderen D, Ventura M. **“Modulation of the *eps-ome* transcription of bifidobacteria through simulation of human intestinal environment.”** FEMS Microbiol Ecol. 2016 Apr;92(4):fiw056. doi: 10.1093/femsec/fiw056.
28. Turrone F, Milani C, Duranti S, Mancabelli L, **Mangifesta M**, Viappiani A, Lugli GA, Ferrario C, Gioiosa L, Ferrarini A, Li J, Palanza P, Delledonne M, van Sinderen D, Ventura M. **“Deciphering bifidobacterial-mediated metabolic interactions and their impact on gut microbiota by a multi-omics approach.”** ISME J. 2016 Jul;10(7):1656-68. doi: 10.1038/ismej.2015.236.