environmental microbiology

Environmental Microbiology (2020) **22**(8), 3375–3393 doi:10.1111/1462-2920.15108

Evolutionary development and co-phylogeny of primate-associated bifidobacteria

Gabriele Andrea Lugli,[1] Giulia Alessandri,[2] Christian Milani,[1,3] Leonardo Mancabelli,[1] Lorena Ruiz,[4,5] Federico Fontana,[1] Santiago Borragán,[6] Andrea González,[7] Francesca Turroni,[1,3] Maria Cristina Ossiprandi,[2] Abelardo Margolles,[4,5] Douwe van Sinderen[8] and Marco Ventura [1,3]*

[1]*Laboratory of Probiogenomics, Department of Chemistry, Life Sciences, and Environmental Sustainability, University of Parma, Parma, 43124, Italy.*

[2]*Department of Veterinary Medical Science, University of Parma, Parma, 43124, Italy.*

[3]*Microbiome Research Hub, University of Parma, Parma, 43124, Italy.*

[4]*Department of Microbiology and Biochemistry, Dairy Research Institute of Asturias, Spanish National Research Council (IPLA-CSIC), Paseo Río Linares s/n, Villaviciosa, Asturias, 33300, Spain.*

[5]*MicroHealth Group, Instituto de Investigación Sanitaria del Principado de Asturias (ISPA), Oviedo, Asturias, Spain.*

[6]*Parque de Cabárceno, 39690 Obregón, Cantabria, Spain.*

[7]*Zoo de Santillana, Avda. del Zoo 2, Santillana del Mar, Cantabria, 39330, Spain.*

[8]*APC Microbiome Institute and School of Microbiology, Bioscience Institute, National University of Ireland, T12 YT20, Cork, Ireland.*

Summary

In recent years, bifidobacterial populations in the gut of various monkey species have been assessed in several ecological surveys, unveiling a diverse, yet unexplored ecosystem harbouring novel species. In the current study, we investigated the species distribution of bifidobacteria present in 23 different species of primates, including human samples, by means of 16S rRNA microbial profiling and internal transcribed spacer bifidobacterial profiling. Based on the observed bifidobacterial-host co-phylogeny, we found a statistically significant correlation between the *Hominidae* family and particular bifidobacterial species isolated from humans, indicating phylosymbiosis between these lineages. Furthermore, phylogenetic and glycobiome analyses, based on 40 bifidobacterial species isolated from primates, revealed that members of the *Bifidobacterium tissieri* phylogenetic group, which are typical gut inhabitants of members of the *Cebidae* family, descend from an ancient ancestor with respect to other bifidobacterial taxa isolated from primates.

Received 23 April, 2020; revised 27 May, 2020; accepted 29 May, 2020. *For correspondence. E-mail marco.ventura@unipr.it; Tel. ++39-521-905666; Fax ++39-521-905604.

Introduction

The mammalian gut is inhabited by a plethora of microbial species representing the gastrointestinal microbial community, also known as the gut microbiota (Lozupone *et al*., 2012). This microbial population exerts a diversity of metabolic and physiological activities that help to sustain host health, including degradation of food components, protection from pathogens, promotion of host cell differentiation, and stimulation/modulation of the host immune system (Yatsunenko *et al*., 2012). In this context, members of the genus *Bifidobacterium* are important representatives of the mammalian gut microbiota, together with other key players represented by members of the *Bacteroides*, *Enterococcus*, *Lactobacillus*, *Streptococcus* and *Veillonella* genera (Milani *et al*., 2017b). Bifidobacteria are also present in other animals that confer parental care to their offspring such as birds and social insects (Killer *et al*., 2010; Bottacini *et al*., 2012; Ellegaard *et al*., 2015). In recent years, the ability of this group of microorganisms to confer a range of health benefits to the host has been extensively described (Lewis *et al*., 2015; Arboleya *et al*., 2016; Turroni *et al*., 2016; Hidalgo-Cantabrana *et al*., 2017). In this context, members of the *Bifidobacterium adolescentis*, *Bifidobacterium animalis*, *Bifidobacterium bifidum*, *Bifidobacterium breve* and *Bifidobacterium longum* species are known for their probiotic properties and are commonly administered as living microorganisms for the (re-) establishment of a correct gut microbiota (Ventura

et al., 2009; Ruiz *et al*., 2011; Egan *et al*., 2014; Duranti *et al*., 2016).

Since their discovery in 1899, 16 (sub)species belonging to the genus *Bifidobacterium* have been isolated and classified by means of human gut sampling (Tissier, 1900). In recent years, several ecological surveys aimed at assessing bifidobacterial populations have been performed in various species of monkeys, uncovering an extensive unexplored ecosystem. In fact, these efforts allowed the identification of 28 novel bifidobacterial species, distributed among 11 different monkey species, especially in *Saguinus oedipus* and *Callithrix jacchus*, rendering the primate gut a very rich environmental repository of bifidobacterial biodiversity to date (Michelini *et al*., 2016; Duranti *et al*., 2017a; Modesto *et al*., 2018a; Lugli *et al*., 2018b; Modesto *et al*., 2018b; Modesto *et al*., 2018c; Duranti *et al*., 2019; Modesto *et al*., 2019). Recently, a metagenomic approach based on the amplification of the internal transcribed spacer (ITS) rRNA sequence allowed a detailed analysis of bifidobacterial gut populations in the mammalian branch of the tree of life, indicating the presence of as yet undiscovered bifidobacterial taxa (Milani *et al*., 2017a). Furthermore, a combination of whole metagenome shotgun sequencing coupled with cultivation approaches allowed the isolation of novel bifidobacteria from faecal samples of monkeys (Lugli *et al*., 2019). A comparative genomic study, involving genomes of four species of bifidobacteria isolated from marmosets and four species isolated from humans, unveiled apparently species-specific genes predicted to be involved in nutrient uptake (Brown *et al*., 2019).

In recent years, the number of studies on non-human primates has increased substantially, allowing expansion of our knowledge about their behaviour, ecology and social systems (Garber, 2019). Furthermore, due to the advent of modern sequencing technologies, significant research effort has been aimed at disclosing the microbiota composition of non-human primates and at investigating how this is influenced by ecological variation through time (Greene *et al*., 2019; Orkin *et al*., 2019). These studies described gut mirobiomes of several non-human primate species in which Actinobacteria were outnumbered by Firmicutes and Bacteroidetes, being reminiscent of the human gut microbiota (Bornbusch *et al*., 2019; Gomez *et al*., 2019; Greene *et al*., 2019). In contrast, reports on the gut microbiota of members of the *Cebidae* family highlighted a bacterial composition dominated by the genera *Streptococcus* and *Bifidobacterium* (Orkin *et al*., 2019). This information is consistent with the identification of many novel bifidobacterial species from the gut microbiota of members of the *Cebidae* family.

In the current study, we investigated the distribution of bifidobacteria present in faecal samples from 23 different species of monkeys and humans, together representing five of the major evolutionary lineages of primates, by means of 16S rRNA microbial profiling and ITS bifidobacterial profiling. The correlation between primate and *Bifidobacterium* species occurrences was used to assess possible co-phylogeny between specific hosts and their associated commensals. Furthermore, the genetic features of bifidobacterial species isolated from monkeys were compared with those obtained from human isolates, in order to identify differences in metabolic abilities and features related to the genetic adaptation of these species to the primate gut. Finally, 40 bifidobacterial species were evaluated for their glycan breakdown activities by growth measurements on carbon sources selected by *in silico* predictions.

Results and discussion

Identification of bifidobacteria in faecal samples of monkeys

In order to explore the relative bifidobacterial abundance at genus level within the gut microbiota of various monkeys, we analysed the taxonomic composition of the bacterial faecal community harboured by 38 representatives of various species of primates, of which 20 had previously been sequenced as part of a larger study aimed at exploring the mammalian gut microbiota biodiversity and employing the same methodology (Milani *et al*., 2017a) (Table S1). Altogether, the collected samples were retrieved from primates kept in captivity. Illumina-mediated 16S rRNA microbial profiling of the additional 18 primate samples produced approximately 800 thousand sequence reads with an average of 44 thousand quality-filtered reads per sample, with a similar number obtained from the re-analysis of the previously sequenced data (Table S1). The obtained nucleotide sequences were grouped in amplicon sequence variants (ASVs) and then taxonomically classified. This analysis showed that the relative bifidobacterial abundance reached 82% in the sample of *Callithrix argentata*, though with an average of 9% across the assessed 38 primate-derived samples (Fig. 1C). Based on primate phylogeny, we grouped each monkey into families, resulting in four major clusters, i.e. *Hominidae*, *Cercopithecidae*, *Cebidae* and *Lemuridae* (Table 1). Interestingly, the average abundance of bifidobacteria in primate hosts at family level appears to be quite different, highlighting the *Cebidae* family as the group with the highest average of relative bifidobacterial abundance (17%), followed by the *Hominidae* family (10%), while members of the *Lemuridae* and *Cercopithecidae* families displayed relative bifidobacterial abundancies of just 3% and 0.03% respectively (Fig. 1C). More specifically, members of the

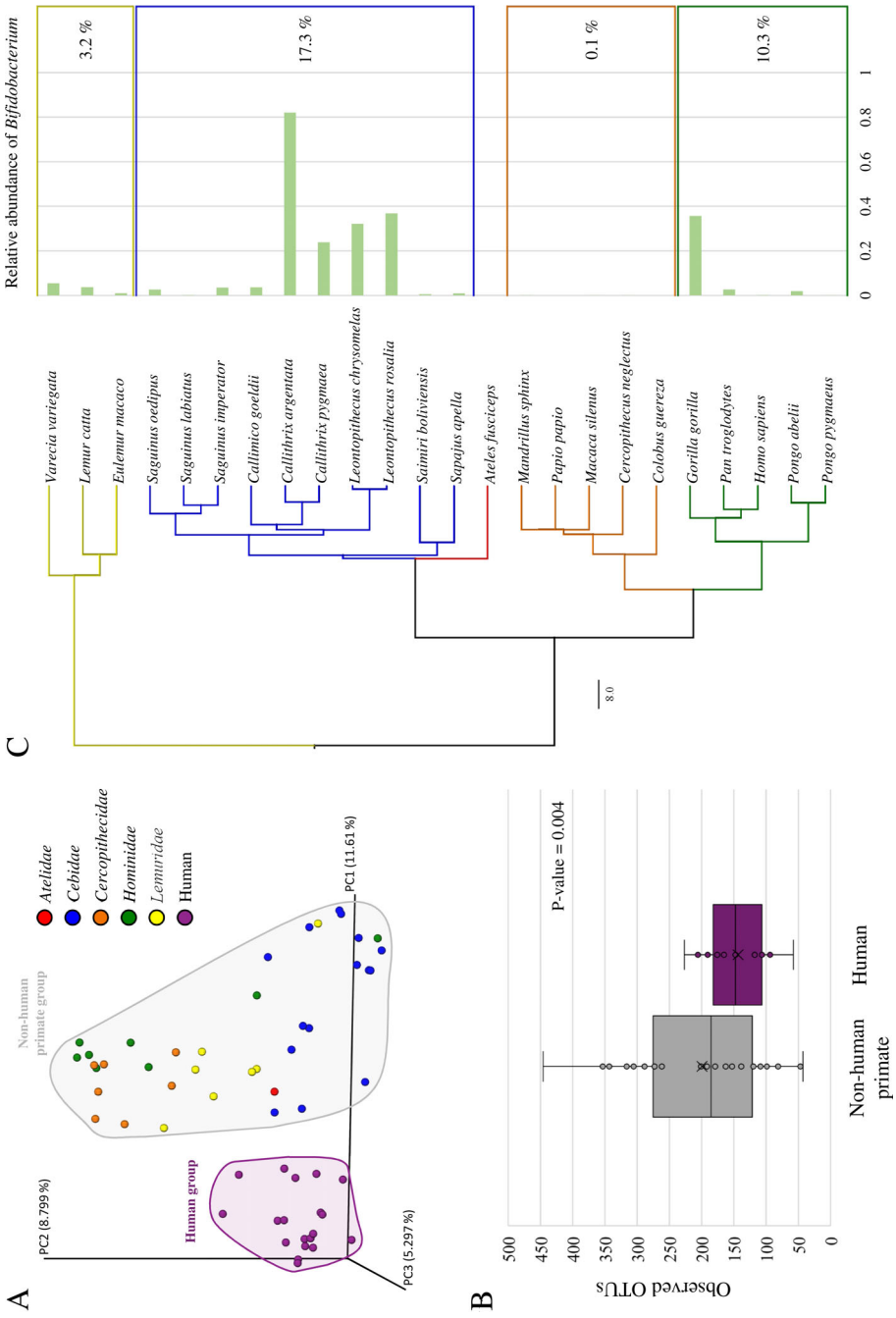

Fig 1. 16S rRNA gene-based microbial profiling of 57 faecal samples from humans and various non-human primates. Panel A shows the PCoA representation of the beta-diversity related to samples which are coloured based on corresponding primate lineages. Polygons delimiting each group were manually generated. Panel B exhibits a Whisker plot based on Chao1 alpha-diversity of non-human primate and human gut microbiota data, which results in a *p*-value of <0.004 between the two groups (Student's t test). The *y*-axis shows the number of observed ASVs obtained from faecal samples of non-human primates (grey) or humans (purple). Dots reflect the distribution of a data set, while the boxes represent 50% of the data set, distributed between the first and third quartiles. The median divides the boxes into the interquartile range, while the *X* represents the mean. The lines extending vertically outside the boxes show the outlier range. Panel C displays a tree based on the evolutionary timescale of life including families of the analysed primates (left), and the relative abundance of bifidobacterial DNA retrieved by means of 16S rRNA gene profiling (right, green pillars). Tree branches are coloured based on the primate lineages: *Atelidae* (red), *Cebidae* (blue), *Cercopithecidae* (orange), *Hominidae* (green) and *Lemuridae* (yellow). [Color figure can be viewed at wileyonlinelibrary.com]

Table 1. General genomic features of bifidobacteria isolated from primates.

No.	*Bifidobacterium* species	Strain	Genome status	Genome size	GC content	ORFs number	rRNA loci[a]	tRNA number	GHs number	GH index (%)	Accession number	Isolated from	Host species	Host family
1	*Bifidobacterium adolescentis*	ATCC 15703	Complete	2 089 645	59.18	1676	5	54	56	3.4	AP009256.1	Human	*Homo sapiens*	*Hominidae*
2	*Bifidobacterium aerophilum*	DSM 100689	–	–	–	–	–	–	–	–	–	Cotton-top tamarin	*Saguinus oedipus*	*Cebidae*
3	*Bifidobacterium aesculapii*	DSM 26737	Draft (118)	2 794 396	64.58	2151	6	60	57	2.6	BCFK00000000	Common marmoset	*Callithrix jacchus*	*Cebidae*
4	*Bifidobacterium angulatum*	DSM 20098	Complete	2 021 974	59.43	1557	4	53	38	2.5	AP012322.1	Human	*Homo sapiens*	*Hominidae*
5	*Bifidobacterium avesanii*	DSM 100685	Draft (46)	2 684 414	66.29	2046	4	56	79	3.9	WBSN00000000	Cotton-top tamarin	*Saguinus oedipus*	*Cebidae*
6	*Bifidobacterium biavatii*	DSM 23969	Draft (56)	3 252 147	63.1	2557	5	61	107	4.2	JGYN00000000	Red-handed tamarin	*Saguinus midas*	*Cebidae*
7	*Bifidobacterium bifidum*	ATCC 29521	Complete	2 211 039	62.66	1822	2	54	41	2.3	AP012323.1	Human	*Homo sapiens*	*Hominidae*
8	*Bifidobacterium breve*	DSM 20213	Complete	2 269 415	58.89	1977	2	53	43	2.2	AP012324.1	Human	*Homo sapiens*	*Hominidae*
9	*Bifidobacterium callimiconis*	LMG 30938	Draft (32)	2 962 404	62.37	2263	5	61	61	2.7	QXGJ00000000	Goeldi's marmoset	*Callimico goeldii*	*Cebidae*
10	*Bifidobacterium callitrichidarum*	DSM 103152	Draft (58)	3 121 265	61.81	2526	2	79	73	2.9	QFFM00000000	Emperor tamarin	*Saguinus imperator*	*Cebidae*
11	*Bifidobacterium callitrichos*	DSM 23973	Draft (33)	2 887 313	63.52	2364	3	58	81	3.4	JGYS00000000	Common marmoset	*Callithrix jacchus*	*Cebidae*
12	*Bifidobacterium catenulatum* subsp. *catenulatum*	DSM 16992	Complete	2 079 525	56.2	1686	5	55	52	3.1	AP012325.1	Human	*Homo sapiens*	*Hominidae*
13	*Bifidobacterium catenulatum* subsp. *kashiwanohense*	DSM 21854	Complete	2 337 234	56.28	1932	5	54	59	3.1	AP012327.1	Human	*Homo sapiens*	*Hominidae*
14	*Bifidobacterium catulorum*	DSM 103154	Draft (78)	2 611 484	63.17	2043	2	55	33	1.6	QFFN00000000	Common marmoset	*Callithrix jacchus*	*Cebidae*
15	*Bifidobacterium dentium*	JCM 1195	Complete	2 635 669	58.54	2167	4	56	84	3.9	AP012326.1	Human	*Homo sapiens*	*Hominidae*
16	*Bifidobacterium eulemuris*	DSM 100216	Draft (34)	2 913 389	62.2	2328	2	53	115	4.9	MWWZ00000000	Black lemur	*Eulemur macaco*	*Lemuridae*
17	*Bifidobacterium felsineum*	DSM 103139	Draft (12)	2 382 257	57.11	1872	2	56	61	3.3	PEBJ00000000	Cotton-top tamarin	*Saguinus oedipus*	*Cebidae*
18	*Bifidobacterium gallicum*	LMG 11596	Draft (12)	2 004 594	57.61	1505	2	58	22	1.5	JGYW00000000	Human	*Homo sapiens*	*Hominidae*
19	*Bifidobacterium goeldii*	LMG 30939	Draft (26)	2 608 375	56.08	2051	2	51	64	3.1	QXGL00000000	Goeldi's marmoset	*Callimico goeldii*	*Cebidae*
20	*Bifidobacterium hapali*	DSM 100202	Draft (76)	2 834 308	54.5	2246	3	54	96	4.3	MWWY00000000	Common marmoset	*Callithrix jacchus*	*Cebidae*
21	*Bifidobacterium imperatoris*	LMG 30297	Draft (62)	2 639 899	56.13	2203	4	59	62	2.8	NMWV00000000	Emperor tamarin	*Saguinus imperator*	*Cebidae*
22	*Bifidobacterium jacchi*	DSM 103362	Draft (100)	2 851 761	62.26	2078	1	55	43	2.1	RQSP00000000	Common marmoset	*Callithrix jacchus*	*Cebidae*
23	*Bifidobacterium lemurum*	DSM 28807	Draft (38)	2 944 293	62.64	2316	3	49	109	4.7	MWWX00000000	Ring-tailed lemur	*Lemur catta*	*Lemuridae*
24	*Bifidobacterium longum* subsp. *infantis*	ATCC 15697	Complete	2 832 748	59.86	2580	4	77	51	2.0	CP001095.1	Human	*Homo sapiens*	*Hominidae*
25	*Bifidobacterium longum* subsp. *longum*	JCM 1217	Complete	2 385 164	60.33	2001	4	73	56	2.8	AP010888.1	Human	*Homo sapiens*	*Hominidae*
26	*Bifidobacterium margollesii*	LMG 30296	Draft (80)	2 789 387	61.91	2270	3	60	53	2.3	NMWU00000000	Pygmy marmoset	*Callithrix pygmaea*	*Cebidae*
27	*Bifidobacterium moukalabense*	DSM 27321	Draft (12)	2 515 335	59.87	2046	4	56	85	4.2	AZMV00000000	Western lowland gorilla	*Gorilla gorilla*	*Hominidae*
28	*Bifidobacterium myosotis*	DSM 100196	Draft (58)	2 944 195	62.55	2163	4	56	76	3.5	MWWW00000000	Common marmoset	*Callithrix jacchus*	*Cebidae*
29	*Bifidobacterium parmae*	LMG 30295	Draft (44)	2 820 211	65.81	2243	6	62	43	1.9	NMWT00000000	Pygmy marmoset	*Callithrix pygmaea*	*Cebidae*

(Continues)

Table 1. Continued

No.	*Bifidobacterium* species	Strain	Genome status	Genome size	GC content	ORFs number	rRNA loci[a]	tRNA number	GHs number	GH index (%)	Accession number	Isolated from	Host species	Host family
30	*Bifidobacterium primatium*	DSM 100687	Draft (8)	2 696 768	63.2	2073	2	57	51	2.5	PEBI00000000	Cotton-top tamarin	*Saguinus oedipus*	*Cebidae*
31	*Bifidobacterium pseudocatenulatum*	DSM 20438	Complete	2 313 752	56.38	1816	6	54	65	3.6	AP012330.1	Human	*Homo sapiens*	*Hominidae*
32	*Bifidobacterium ramosum*	DSM 100688	Draft (25)	3 028 921	63.51	2262	3	57	100	4.4	WBSM00000000	Cotton-top tamarin	*Saguinus oedipus*	*Cebidae*
33	*Bifidobacterium reuteri*	DSM 23975	Draft (28)	2 847 572	60.45	2149	4	53	64	3.0	JGZK00000000	Common marmoset	*Callithrix jacchus*	*Cebidae*
34	*Bifidobacterium saguini*	DSM 23967	Draft (33)	2 787 036	56.35	2321	5	59	72	3.1	JGZN00000000	Red-handed tamarin	*Saguinus midas*	*Cebidae*
35	*Bifidobacterium samirii*	LMG 30940	Draft (42)	2 574 625	66.64	1968	5	60	91	4.6	QXGK00000000	Black-capped squirrel monkey	*Saimiri boliviensis*	*Cebidae*
36	*Bifidobacterium scaligerum*	DSM 103140	Draft (19)	2 650 899	58.27	2063	2	59	59	2.9	PGLQ00000000	Cotton-top tamarin	*Saguinus oedipus*	*Cebidae*
37	*Bifidobacterium scardovii*	DSM 13734	Complete	3 158 347	64.63	2489	3	56	117	4.7	AP012331.1	Human	*Homo sapiens*	*Hominidae*
38	*Bifidobacterium simiarum*	DSM 103153	Draft (19)	2 714 697	63.79	2083	1	58	64	3.1	PEBK00000000	Emperor tamarin	*Saguinus imperator*	*Cebidae*
39	*Bifidobacterium stellenboschense*	DSM 23968	Draft (40)	2 812 864	65.34	2202	6	59	46	2.1	JGZP00000000	Red-handed tamarin	*Saguinus midas*	*Cebidae*
40	*Bifidobacterium tissieri*	DSM 100201	Draft (38)	2 873 483	61.05	2255	2	60	58	2.6	MWWV00000000	Common marmoset	*Callithrix jacchus*	*Cebidae*
41	*Bifidobacterium vansinderenii*	LMG 30126	Draft (68)	3 111 005	62.47	2515	4	60	66	2.6	NEWD00000000	Emperor tamarin	*Saguinus imperator*	*Cebidae*

[a]Predicted number of rRNA loci based on draft genome sequences.

genus *Callithrix* and *Leontopithecus*, together with genus *Gorilla*, showed the highest relative abundance of bifidobacteria within their deduced gut microbiota (>32%) (Fig. 1C).

To compare the data collected from the non-human primate gut microbiota to that from human faecal samples, 19 additional samples belonging to healthy adult individuals were analysed following the same 16S rRNA microbial profiling protocols as the monkey stool samples (Milani *et al*., 2019) (sample size estimation of 10 and 19.5 between non-human primate species and humans, based on ASVs and bacterial genera abundance respectively). Faecal samples of humans were shown to contain an average relative bifidobacterial abundance of 0.2%, which is lower compared with the examined faecal samples of monkeys, with the exception of members of the *Cercopithecidae* family (0.03%; see above). Notably, based on the α-diversity representing the whole microbiota biodiversity, human samples were also shown to be lower in overall bacterial richness when compared with other analysed primate species (*t*-test *p*-value <0.01, $df = 55$, Cohen's $d = 0.74$ and effect size $r = 0.35$) (Fig. 1B). Moreover, β-diversity was analysed based on the unweighted Unifrac distance matrix and represented through a principal coordinate analysis (PCoA), revealing a significant clustering of two major groups represented by human and non-human primates (PERMANOVA based on 999 permutations *p*-value = 0.001, pseudo-$F = 6.20$, $R^2 = 0.10$) (Fig. 1A). Thus, these results highlight a clear difference between the microbiota composition of humans and monkeys, likely representing the outcomes of host-microbiota adaptation due to multiple environmental factors and changes in the lifestyle, such as diet and urbanization (De Filippo *et al*., 2017; Mancabelli *et al*., 2017; McCall *et al*., 2019). However, one should keep in mind that the analysed samples are derived from primates that are kept in captivity, representing an unnatural environment for these animals, although their diets may have allowed the preservation of their bacterial gut richness.

To investigate the bifidobacterial distribution among various investigated primate species, we employed a time tree showing the evolutionary timescale of life (Kumar *et al*., 2017) (Fig. 1C). Based on our results, approximately 45 million years ago, the differentiation of the infraorder *Simiiformes* into *Platyrrhini* and *Catarrhini* gave rise to two distinct primate lineages, which now appear to harbour a different abundance of members of the genus *Bifidobacterium*. In this context, the *Platyrrhini* branch, represented in this study by monkeys belonging to the *Cebidae* family, appear to represent an optimal ecological niche for bifidobacterial colonization when compared with members of the *Catarrhini* parvorder (*t*-test *p*-value <0.05, $df = 46$, Cohen's $d = 0.57$ and effect size $r = 0.27$). Despite the high number of bifidobacterial species isolated from primates, larger studies of wild primates highlight that bifidobacteria do not consistently discriminate primate families from a phylogenetic perspective (Amato *et al*., 2019; Gomez *et al*., 2019). In this context, phylogenetically distant primates were found to possess different microbiota compositions, dominated by *Prevotella*, unknown *Coriobacteriaceae* and *Streptococcus* (Gomez *et al*., 2019). In fact, within our analysed samples, the highest abundance of members of the genus *Bifidobacterium* was retrieved in two specific genera of the *Cebidae* family, i.e. *Callithrix* and *Leontopithecus*, and not among all the members of the *Cebidae* family (Fig. 1C). Thus, based on reported data, bifidobacteria seems to dominate the gut of specific lineages of primates instead of a whole family of these mammals.

The distribution of bifidobacteria among the analysed faecal samples of primates as highlighted by 16S rRNA microbial profiling reflects from which monkey species these bifidobacteria were isolated (Table 1). In fact, 26 out of 29 bifidobacterial type strains isolated from monkeys were retrieved from faecal samples collected from members of the *Callitrichinae* subfamily, encompassing *Saguinus*, *Callimico* and *Callithrix* genera. The remaining three bifidobacterial species were also isolated from monkey-associated faecal samples that contain a substantial level of bifidobacteria, and belonging to members of the *Lemuridae* and *Hominidae* families (average relative abundance of bifidobacteria of 3.2% and 10.3% respectively) (Table 1; Fig. 1C). Interestingly, none of the bifidobacterial species reported in Table 1 was isolated from the *Leontopithecus* genus, i.e. the fourth genus of the *Callitrichinae* subfamily, which together with the *Saguinus*, *Callimico* and *Callithrix* genera represent species with high levels of bifidobacteria within their gut microbiota.

Distribution of bifidobacterial species among the gut microbiota of primates

To characterize the bifidobacterial population at the species level, we employed bifidobacterial ITS profiling analysis, resulting in an average of 20 thousand high-quality filtered reads per sample (Table S2). The ITS profiling approach allowed us to cluster the obtained sequences in ASVs with 100% identity cut-off, which were subsequently classified by their bifidobacterial taxonomy at (sub)species level (Milani *et al*., 2014). The obtained data sets were then employed to evaluate the distribution of specific bifidobacterial species across 38 samples, defined as prevalence (Fig. 2B).

As expected, human samples revealed high prevalence (>11%) of *B. adolescentis*, *B. longum* and *B. breve*

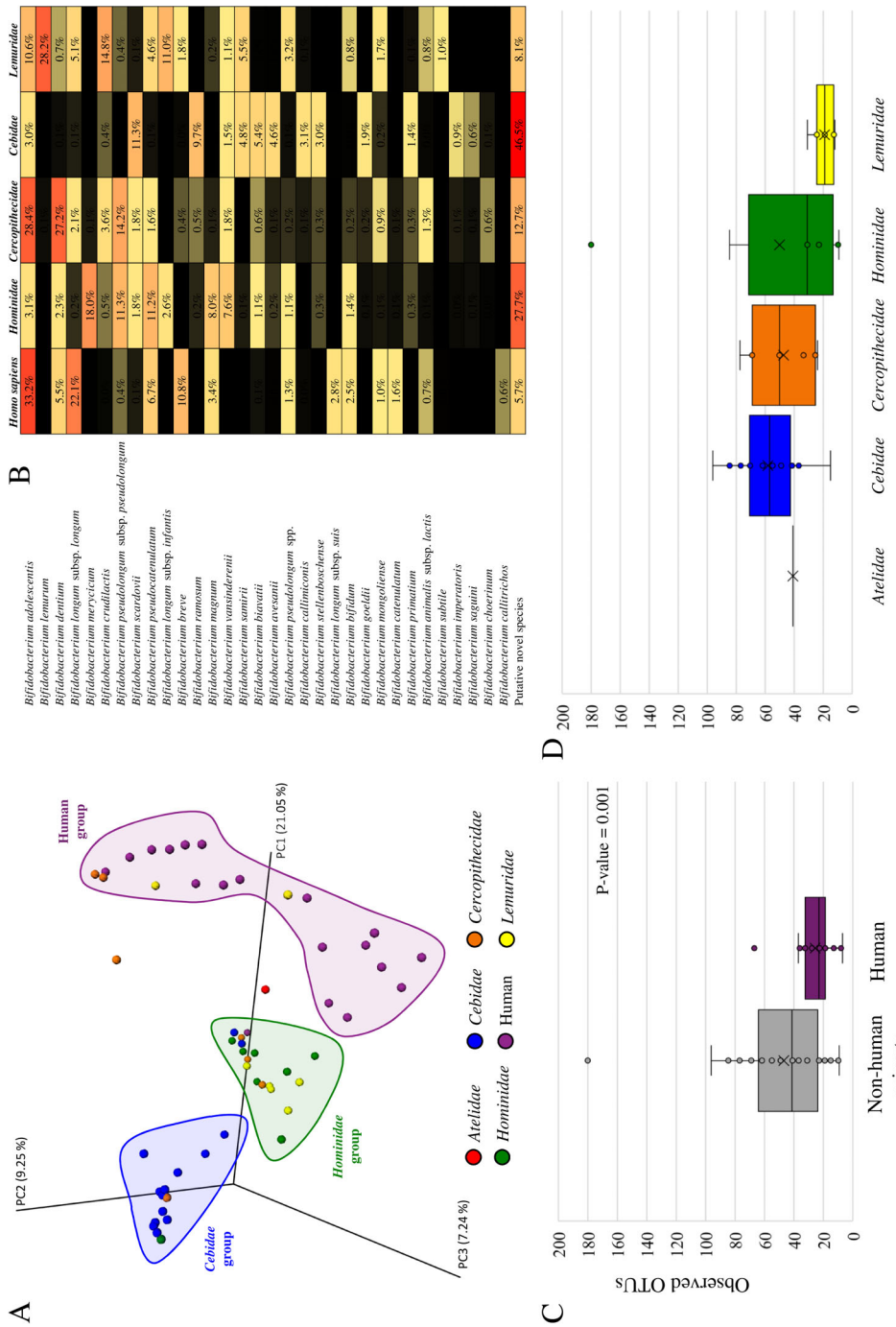

Fig 2. ITS bifidobacterial profiling of 57 faecal samples of humans and various non-human primate species. Panel A visualizes the PCoA representation of the beta-diversity related to samples coloured based on primate lineages. Polygons delimiting each group were manually generated. Panel B displays the relative abundance of each identified species belonging to the genus *Bifidobacterium* by means of ITS bifidobacterial profiling. Numbers represent the average percentage of bifidobacterial species identified in a specific primate family. Only species that display at least 1% of the total amount of data in at least a primate family are included in the heat map. The average percentage of putative novel bifidobacterial species is reported together, while the order of *Bifidobacterium* species is reported based on their abundance among primate family. Panels C and D exhibit Whisker plots based on Chao1 alpha-diversity of non-human primate and human gut microbiota data. The *y*-axis shows the number of observed OTUs obtained from faecal samples of non-human primates (grey) or humans (purple), while family groups were coloured based on primate lineages. Dots reflect the distribution of a data set, while the boxes represent 50% of the data set, distributed between the first and third quartiles. The median divides the boxes into the interquartile range, while the *X* represents the mean. The lines extending vertically outside the boxes show the outlier range. [Color figure can be viewed at wileyonlinelibrary.com]

(Fig. 2B), which are bifidobacterial species typically reported to inhabit the human gut (Turroni *et al*., 2018), together with other mammalian species as previously reported (Milani *et al*., 2017a). Furthermore, monkeys phylogenetically related to humans, i.e. those belonging to the *Hominidae*, revealed a high prevalence (>11%) of *Bifidobacterium merycicum*, followed by *Bifidobacterium pseudolongum* and *Bifidobacterium pseudocatenulatum* that were both originally isolated from the gut of humans (Table 1). In a similar fashion, *Cercopithecidae* monkeys, belonging to the *Catarrhini* parvorder as well as *Hominidae*, disclosed a relative abundance of 70% of the identified ITS sequences belonging to members of *B. adolescentis*, *Bifidobacterium dentium* and *B. pseudolongum* species, which were also isolated from human samples (Fig. 2B). In contrast, ITS bifidobacterial profiling revealed in monkeys of the *Cebidae* and *Lemuridae* families a high prevalence (>10%) of *Bifidobacterium scardovii* and *Bifidobacterium ramosum*, and *Bifidobacterium lemurum* and *Bifidobacterium crudilactis* respectively (Fig. 2B). The bifidobacterial biodiversity at species level across primates was also highlighted by the PCoA based on the unweighted Unifrac distance matrix, revealing a clustering of the *Hominidae* samples between other two groups represented by humans (PERMANOVA based on 999 permutations p-value = 0.001, pseudo-F = 5.83, R^2 = 0.18) and by members of the *Cebidae* family (PERMANOVA based on 999 permutations p-value = 0.002, pseudo-F = 2.69, R^2 = 0.11) (Fig. 2A).

In the same fashion as was observed for the 16S rRNA microbial profiling findings, the α-diversity based on the ITS bifidobacterial profiling highlighted a reduced bifidobacterial richness in the human samples when compared with that observed for other primates (t-test p-value <0.001, df = 55, Cohen's d = 0.85 and effect size r = 0.39) (Fig. 2C). In this context, the higher number of distinct bifidobacterial operational taxonomic units (OTUs) identified in monkey's faecal samples was to a large degree represented by putative novel species of the genus *Bifidobacterium*, consisting of up to 22 unique OTUs next to ITS sequences of known type strains (Fig. 2B). In humans, the average relative abundance of putative novel bifidobacterial species was 6%, while in monkey-derived samples the average percentage rises to 28%. Looking at the primate clusters as based on family grouping, samples originating from *Cebidae* followed by *Hominidae* and *Cercopithecidae* elicited the highest levels of bifidobacterial biodiversity (ANOVA p-value <0.01; LSD post hoc p-value <0.01, df = 24, Cohen's d = 2.56 and effect size r = 0.78 and LSD post hoc test p-value <0.01, df = 32, Cohen's d = 1.88 and effect size r = 0.69 when data from *Cebidae* were compared with that from *Lemuridae* and humans respectively) (Fig. 2D). These data were further supported by the average percentage of predicted novel bifidobacterial taxa of the *Cebidae* family, i.e. 46%, supporting the high bifidobacterial biodiversity harboured by members of this particular primate family. Focusing on novel bifidobacterial OTUs, new_taxa_16 and new_taxa_35 were identified in 11 and eight monkey species respectively (Table S3), followed by new_taxa_7 and new_taxa_8, both identified in samples retrieved from seven monkey species. Based on their occurrence, these putative novel *Bifidobacterium* species represent excellent targets for future projects aimed at isolating and exploring the bifidobacterial dark matter of these mammals.

The primate-bifidobacteria co-phylogeny

Phylogenetic relatedness among bifidobacterial species isolated from primates was investigated by means of a pangenome analysis involving available genome sequences of type strains retrieved from the NCBI database (Table 1). Based on this approach, we were able to determine putative orthologous genes among the 23 (sub)species of the genus *Bifidobacterium* whose relative abundance was identified to be higher than 1% in at least one primate-derived faecal sample (Figs 2B and 3). The analysis resulted in the identification of 20 703 clusters of orthologous genes (COGs), of which 506 were shared between all genomes, representing the core bifidobacterial coding sequences of the analysed bifidobacterial type strains. After exclusion of paralogues, concatenation of the amino acid sequences of the remaining 462 core proteins was used to build the *Bifidobacterium* phylogenetic tree of type strains isolated from primates (Fig. 3). Furthermore, phylogenetic relationships between primate species were delineated through a time tree based on the evolutionary timescale of life (Kumar *et al*., 2017) (Figs 1C and 3).

In order to test the significance between individual host-bifidobacteria associations, we employed the statistical test ParaFit (Legendre *et al*., 2002). The co-phylogeny analysis revealed that *B. adolescentis*, followed by *Bifidobacterium biavatii* and *B. ramosum*, were the taxa most commonly shared among primates (from 15 to 7), while *Leontopithecus chrysomelas* and *Callimico goeldii*, followed by *Saguinus oedipus*, were the monkey species with the highest number of associations among primate-associated bifidobacteria (from 9 to 8) (Fig. 3). Significant associations were identified among members of the *Hominidae* family, showing a correlation between all bifidobacterial species isolated from humans. Notably, the analysed gut microbiota of *Gorilla gorilla*, *Homo sapiens* and *Pan troglodytes* revealed three to seven significant correlations with *B. adolescentis*, *B. bifidum*, *B. breve*, *B. catenulatum*, *B. dentium*,

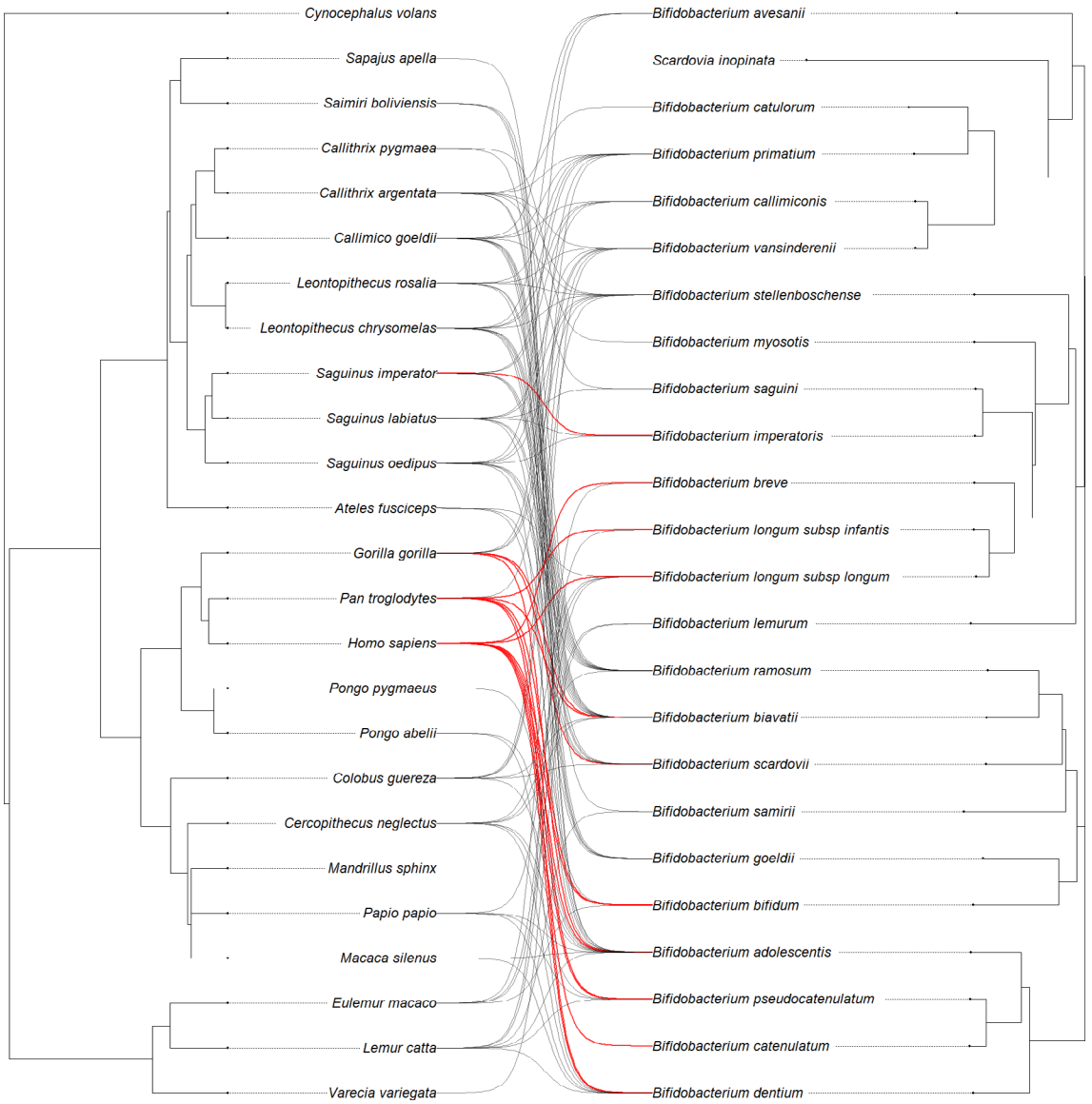

Fig 3. Tanglegram of co-phylogenetic relationships between 24 primates and 23 bifidobacterial species. Tanglegram is composed of a tree based on the evolutionary timescale of families of primates (left) and a proteomic tree based on the concatenation of 462 core genes identified in the pangenome analysis of 23 *Bifidobacterium* type strains (right). All associations are shown in the tanglegram as black and red connecting lines. The latter lines indicate significant individual co-speciation links between bifidobacteria and their hosts as indicated by ParaFit ($p < 0.05$), while black lines represent non-significant links. Bifidobacterial and primate trees are rooted with *Scardovia inopinata* JCM 12537 and *Cynocephalus volans* as outgroups respectively. Bootstrap percentages, of the bifidobacterial proteome tree, above 50 are shown at node points, based on 1000 replicates. [Color figure can be viewed at wileyonlinelibrary.com]

B. longum spp. and *B. pseudocatenulatum* (parafit based on 999 permutations *p*-value <0.05). In addition, *Gorilla gorilla* and *Pan troglodytes* were also found significantly correlated to *B. biavatii* (parafit based on 999 permutations *p*-value <0.01), while *Saguinus imperator* was significantly correlated with the presence of *Bifidobacterium imperatoris* (parafit based on 999 permutations *p*-value <0.05) (Fig. 3) (for more details see Table S4 and Experimental procedures).

Altogether, these data provide robust evidence of phylosymbiosis among the *Hominidae* family and bifidobacterial species isolated from humans, as underlined by co-phylogeny interactions between these lineages. In fact, all of the host-bifidobacteria links between the latter groups were shown to be significant, while interactions between other lineages of primates and bifidobacteria highlighted a more widespread distribution of non-human bifidobacteria.

Bifidobacterium *genus biogeography*

As performed in the co-phylogeny analysis, an additional pangenome analysis was performed employing each bifidobacterial type strain sequenced to date, including 81 genomes listed in Table S5. Based on the orthologous genes determined by this approach, 255 COGs were identified to be shared among the 81 (sub)species of the genus. Following the exclusion of paralogues, concatenation of 224 core protein sequences was used to build the *Bifidobacterium* phylogenetic tree (Fig. 4). Among tree branches, the bifidobacterial type strains isolated from primates were identified in four of the 10 previously defined bifidobacterial phylogenetic groups (Lugli *et al*., 2017; Lugli *et al*., 2018a; Lugli *et al*., 2019), i.e. *B*. *adolescentis*, *B*. *bifidum*, *B*. *longum* and *Bifidobacterium tissieri* group. Interestingly, a small number of bifidobacterial species, which had not been isolated from primates, were identified in these four phylogenetic groups, i.e. *Bifidobacterium ruminantium* LMG 21811, *Bifidobacterium rousetti* DSM 106027, *Bifidobacterium merycicum* LMG 11341 and *B*. *longum* subsp. *suis* LMG 21814. Notably, *Bifidobacterium gallicum* LMG 11596 was the only species isolated from primates that were not included in any of the four major groups (Fig. 4).

Based on these results, bifidobacterial species that commonly inhabit the primate gut appear to have undergone an analogous evolution that gave rise to two major branches, i.e. the *B*. *tissieri* phylogenetic group and a second branch composed of the *B*. *adolescentis*, *B*. *bifidum* and *B*. *longum* phylogenetic groups (Fig. 4). The first of these two major branches, defined as cluster A, includes seven bifidobacterial species isolated from monkeys of the *Cebidae* family. Based on the ITS bifidobacterial profiling and the co-phylogeny relationships as discussed above, four of these taxa were observed to be directly correlated with hosts belonging to the *Cebidae* lineage, i.e. *Bifidobacterium callimiconis*, *Bifidobacterium catulorum*, *Bifidobacterium primatium* and *Bifidobacterium vansinderenii* (Fig. 3). Instead, the second major branch was more heterogeneous in terms of bifidobacterial origins, with sub-clusters based on taxa isolated from human (Clusters B and D3) and non-human primates (Clusters C, D1 and D2) (Fig. 4). In this context, cluster B, corresponding to the *B*. *adolescentis* phylogenetic group, consists of human taxa with the exception of *B*. *moukalabense* isolated from the *Hominidae* family, while cluster C, corresponding to the *B*. *bifidum* phylogenetic group, encompasses six bifidobacterial taxa retrieved from monkeys and only two bifidobacterial species of human origin, i.e. *B*. *scardovii* and *B*. *bifidum*. Cluster D on the other hand encompasses members of the *B*. *longum* phylogenetic group, including taxa that belong to both human and non-human primates, producing primate sub-clusters D1 and D2, and bifidobacterial species isolated from the human gut in sub-cluster D3.

Altogether, this phylogenetic analysis highlights that bifidobacterial strains inhabiting the gut of primates are strictly related to each other. As also demonstrated by the co-phylogeny analysis, these bifidobacterial strains revealed evidence of phylosymbiosis with their primate hosts giving rise to a relationship mediated by their apparent ability to colonize the gut environment of primates. In contrast, members of the *B*. *tissieri* phylogenetic group, which are typically resident in the gut of monkeys of the *Cebidae* family, seem to have descended from an ancient ancestor when compared with other primate-associated bifidobacterial species. Thus, members of Cluster A, when compared with other identified clusters, appear to reflect a host-species relationship that corresponds with a higher relative abundance of the genus *Bifidobacterium* in the 16S rRNA microbial profiling as well as the ITS bifidobacterial analysis reported in members of the *Catarrhini* parvorder.

Insights into the carbohydrate metabolism of primate-associated bifidobacteria

One of the key genetic features of bifidobacteria, correlated with their higher ecological fitness in terms of colonization of the animal gut, is their ability to utilize complex carbohydrates (Milani *et al*., 2015a; Milani *et al*., 2015b). Thus, we assessed which encoded proteins possess the ability to enhance the carbohydrate-harvesting abilities of *Bifidobacterium* strains isolated from the primate gut environment, by means of a classification according to the Carbohydrate-Active enZyme (CAZy) database (Lombard *et al*., 2014). The dissected proteome of the 40 bifidobacterial type strains isolated from primates (Table 1) revealed 3850 genes predicted to encode CAZys, i.e. glycosyl hydrolases (GHs), glycosyl transferases, polysaccharide lyases, carbohydrate esterases and carbohydrate-binding modules (CBMs). Focusing on GHs, this analysis resulted in the identification of 2577 proteins predicted to possess catalytic modules involved in the degradation of carbohydrates, where GH13, GH43, GH3, GH2, GH36, GH42 and GH31 outnumbered the other identified families (Fig. 5A). Interestingly, besides the absence or presence of less frequent GHs between the analysed proteomes, each genome exhibited a varying number of predicted GH43 members, ranging from zero to 18 genes that are predicted to be involved in the degradation of complex polysaccharides, such as (arabino) xylan, (arabino)galactan and arabinan, which are main glycan constituents of the plant cell wall.

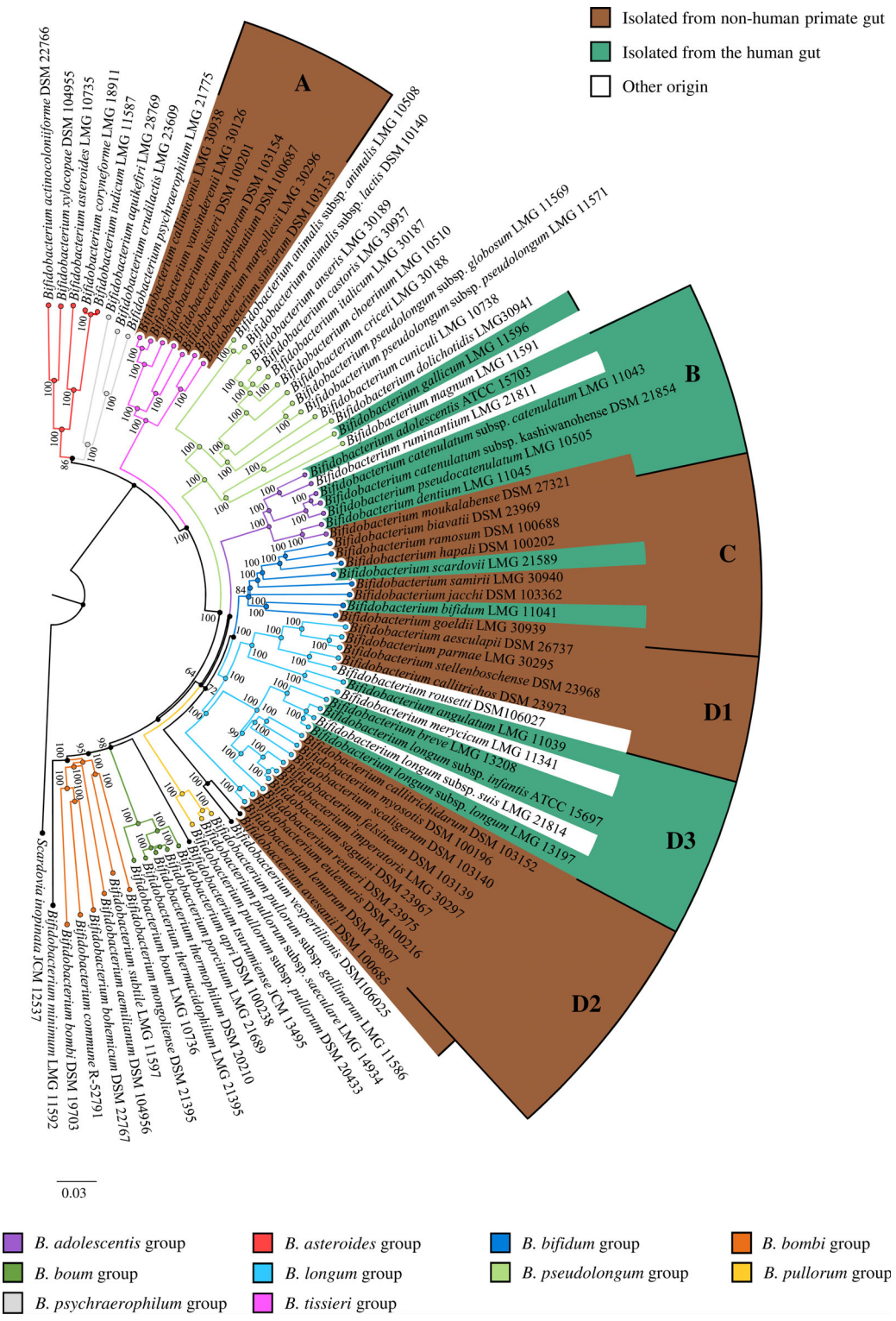

Fig 4. Phylogenomic tree of the *Bifidobacterium* genus. The proteomic tree was based on a concatenation of 224 core genes identified in the pangenome analysis of 81 *Bifidobacterium* strains. Phylogenetic groups are highlighted in different colours. Bifidobacterial species isolated from primates are highlighted based on their isolation source, i.e. human (teal) and non-human primate (brown), while clusters are reported with alphabetical letters. The tree was constructed by the neighbour-joining method, and the genome sequence of *Scardovia inopinata* JCM 12537 was used as an outgroup. Bootstrap percentages above 50 are shown at node points, based on 1000 replicates. [Color figure can be viewed at wileyonlinelibrary.com]

A

Strain	9	14	26	30	38	40	41	1	12	13	15	31	27	7	37	6	19	20	22	32	35	8	4	24	25	3	10	11	16	17	21	23	28	29	33	34	36	39	18	5
Group	*B. tissieri* group							*B. adolescentis* group						*B. bifidum* group								*B. longum* group																				
GH1	2	0	1	0	3	2	2	2	0	2	2	0	1	2	1	2	1	1	0	3	1	1	1	1	0	0	1	3	2	2	3	3	0	0	1	2	1	2	0	1
GH2	5	1	3	4	3	5	7	3	4	4	6	4	5	3	7	6	5	7	3	8	4	2	2	3	2	2	4	4	6	3	3	5	5	2	5	3	5	1	1	8
GH3	4	1	6	5	6	5	5	6	5	4	11	5	15	1	10	2	3	5	1	5	12	3	3	2	4	4	8	5	19	7	7	27	10	5	6	9	7	5	2	3
GH5	0	0	0	0	0	1	0	2	2	1	4	2	3	0	2	1	0	1	0	1	2	2	0	2	1	1	0	1	2	1	0	2	0	1	1	0	1	1	0	0
GH8	0	0	0	0	0	0	0	1	1	2	0	2	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	1	1	0	2	1	0	0	0	0
GH10	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0
GH13	11	10	11	10	11	10	11	11	9	9	13	16	15	6	12	16	14	11	5	12	10	12	9	8	10	12	12	14	14	9	9	13	14	11	13	10	14	12	11	11
GH16	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0
GH18	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
GH20	0	0	1	0	0	0	0	0	0	1	0	1	0	4	3	1	1	3	4	2	4	1	0	3	1	1	2	1	1	1	1	1	2	2	1	2	1	1	0	0
GH23	1	0	0	0	0	1	0	0	0	1	0	1	1	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0
GH25	2	1	2	1	3	2	1	1	1	1	2	1	2	1	4	5	1	1	2	3	1	1	2	4	1	1	2	2	1	1	2	1	1	2	0	2	1	1	1	1
GH26	0	0	0	0	0	1	1	1	0	0	2	0	2	0	2	1	0	1	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0
GH27	0	0	0	0	0	0	0	1	0	1	0	0	1	0	0	1	0	1	0	2	1	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
GH28	4	0	1	3	2	1	6	0	0	0	0	0	0	0	2	4	0	0	0	3	0	0	0	0	0	0	1	1	2	1	0	2	0	0	0	1	0	0	0	10
GH29	0	1	0	1	0	0	0	0	0	2	1	0	1	1	2	1	1	2	1	1	2	0	0	3	0	0	1	1	1	1	1	1	1	0	0	1	0	0	0	0
GH30	3	0	4	0	4	2	3	1	1	1	1	1	1	0	2	3	3	1	0	2	1	0	0	0	1	1	1	2	5	1	1	1	1	2	1	1	2	1	0	2
GH31	4	1	2	1	4	3	5	1	1	3	4	2	3	0	6	3	1	4	0	3	3	0	2	0	3	2	4	3	5	5	4	7	4	1	1	4	2	2	0	2
GH32	2	1	1	2	1	2	1	2	2	2	2	1	1	1	2	2	2	2	0	1	2	1	2	2	1	2	1	1	1	2	1	2	1	1	2	1	2	1	1	1
GH33	0	1	1	0	0	0	0	0	0	0	0	0	0	3	1	0	0	0	4	0	1	1	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
GH35	0	0	0	0	0	0	0	1	0	1	0	0	0	0	1	2	2	1	0	2	1	0	0	0	0	0	1	0	1	2	1	0	0	1	0	1	1	1	0	1
GH36	2	2	3	3	4	1	2	2	3	2	5	3	4	1	8	5	4	5	2	8	4	1	2	1	1	4	4	3	3	4	4	3	7	3	3	5	3	3	1	4
GH37	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
GH38	0	0	0	0	0	0	0	1	1	1	0	0	0	0	2	5	0	1	0	2	2	2	0	2	0	1	0	3	2	0	0	1	0	1	2	0	2	1	0	0
GH39	0	0	0	0	0	0	0	0	1	0	0	0	0	0	2	1	0	1	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0
GH42	2	0	2	0	2	4	2	4	3	3	4	3	4	2	6	5	5	4	1	6	3	2	2	3	2	2	5	4	5	4	4	5	5	2	5	4	4	2	0	1
GH43	2	6	3	10	5	1	4	7	7	8	12	13	11	1	16	16	5	18	1	15	12	0	5	1	9	12	6	10	18	4	5	14	4	2	7	6	1	3	2	8
GH50	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
GH51	1	0	3	0	3	1	1	2	3	2	2	3	2	0	6	4	3	4	0	3	3	0	1	1	4	0	2	2	5	4	2	5	3	0	2	2	2	1	0	3
GH53	1	0	0	0	0	1	1	0	0	0	1	0	1	0	0	1	0	1	0	0	0	0	0	0	0	1	1	0	0	1	0	0	1	0	0	0	0	0	0	0
GH59	2	0	1	0	1	2	2	0	0	0	0	0	0	0	1	1	1	1	0	1	1	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	3
GH67	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
GH76	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
GH77	1	1	1	1	1	1	1	2	2	2	2	2	1	2	2	2	1	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	1	3	2	2	2	2	1
GH78	3	2	3	3	6	5	3	0	0	0	1	0	2	0	2	2	0	3	0	3	1	0	0	0	0	0	0	2	2	0	0	1	0	1	0	1	0	3	0	3
GH79	1	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
GH84	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	1	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
GH85	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0
GH87	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
GH88	1	0	1	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	2
GH89	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
GH92	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
GH94	0	0	0	0	0	0	0	1	0	0	1	0	1	0	1	0	1	0	0	0	1	0	0	0	0	0	1	0	1	1	1	0	1	0	0	1	1	0	0	0
GH95	0	0	0	0	0	0	0	0	0	1	0	1	0	1	0	0	1	1	1	0	1	1	0	1	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0
GH101	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	4	0	2	0	0	0	0	1	0	0	0	1	0	1	1	0	0	1	1	0	0	0
GH105	3	0	0	1	1	2	3	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	1	0	0	0	0	0	3
GH112	0	0	0	0	0	0	0	0	0	0	0	0	0	2	1	1	1	1	1	1	1	1	0	1	1	1	1	0	1	1	1	1	1	0	1	1	0	1	0	0
GH115	1	0	0	1	0	0	1	0	0	0	0	0	0	1	2	1	1	1	2	0	0	0	0	0	0	0	1	1	2	1	1	1	1	1	0	1	2	0	0	1
GH120	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	1	1	0	0	0
GH121	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0
GH125	0	0	0	0	0	0	0	0	0	0	1	0	1	0	1	1	0	1	0	1	1	1	0	1	0	0	0	1	1	0	0	0	0	1	0	1	0	0	0	0
GH127	1	3	1	4	2	1	0	1	3	1	2	2	1	0	4	2	1	5	0	5	1	1	0	0	4	3	3	2	2	1	1	3	4	2	2	1	2	1	0	2
GH129	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	1	1	1	1	1	0	1	1	1	0	1	0	0	0	1	0	1	1	0	0	0	0	1
GH130	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
N° GH	59	31	51	50	64	56	65	53	50	57	79	64	80	35	114	104	60	93	36	100	86	37	34	44	52	54	65	74	112	59	59	106	70	42	63	67	59	45	21	75
GH index	3%	2%	2%	2%	3%	2%	3%	3%	3%	3%	4%	4%	4%	2%	5%	4%	3%	4%	2%	4%	4%	2%	2%	2%	3%	3%	3%	3%	5%	3%	3%	5%	3%	2%	3%	3%	3%	2%	1%	4%

B

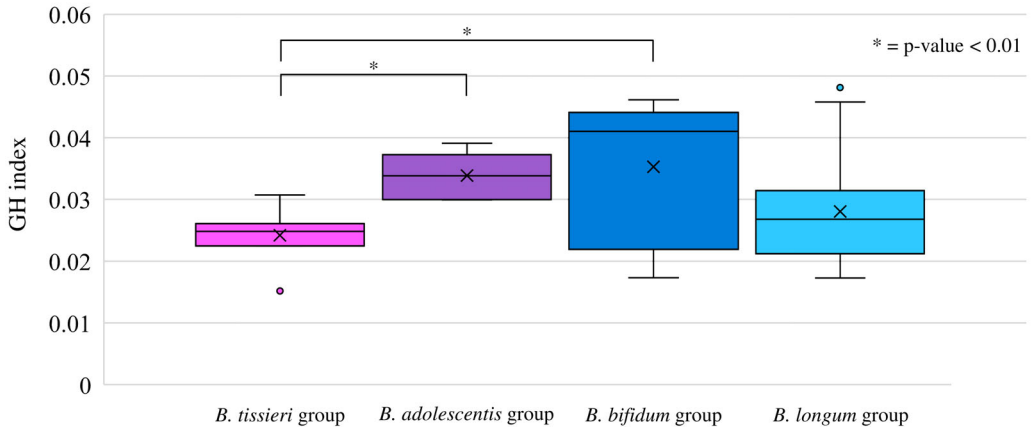

Fig 5. Comparative analysis of bifidobacterial GHs. Panel A shows a heat map of the GH prediction among 40 *Bifidobacterium* species isolated from primates. Panel B displays a Whisker plot based on the GH index identified for each *Bifidobacterium* species. The *y*-axis shows the number of the GH index, subdivided into four major clusters highlighted with different colours based on their phylogenetic groups. Dots reflect the distribution of a data set, while the boxes represent 50% of the data set, distributed between the first and third quartiles. The median divides the boxes into the interquartile range, while the *X* represents the mean. The lines extending vertically outside the boxes show the outlier range. Statistical significance is reported by the connection between groups. [Color figure can be viewed at wileyonlinelibrary.com]

In order to compare genes involved in carbohydrate metabolism of the analysed bifidobacterial species, normalization of GH counts against the total number of predicted genes was performed to generate a GH index for each type strain. The deduced GH indexes range from 1.4% in the case of *B. gallicum* LMG 11596% to 4.81% for *Bifidobacterium eulemuris* DSM 100216, highlighting a broad and variable GH distribution across primate-associated bifidobacteria (Table 1). The GH indexes associated with bifidobacterial species isolated from humans and those from non-human primates were shown not to significantly differ from each other. In contrast, when GH indexes were compared between members from phylogenetic clusters, significant differences were identified between cluster A and cluster B (ANOVA p-value = 0.02; LSD post hoc p-value = 0.02, df = 11, Cohen's d = 2.43 and effect size r = 0.77), as well as between cluster A and cluster C (ANOVA p-value = 0.02; LSD post hoc p-value = 0.005, df = 13, Cohen's d = 1.5 and effect size r = 0.6) (Fig. 5B). As elucidated by the phylogenetic tree of the *Bifidobacterium* genus (Fig. 4), members of the *B. tissieri* group seem to have descended from an ancient ancestor. The evolutionary distance observed from this phylogenetic analysis coincides with a reduced repertoire of genes associated with the metabolism of complex carbohydrates. Despite their simpler GH repertoire, members of the *B. tissieri* group contain between two and six GH78 family-encoding genes, which are predicted to be involved in the hydrolysis of the terminal α-L-rhamnose. Besides this genetic signature, and the presence of GH28 family-encoding genes, which together with their polygalacturonase activity are also associated with the degradation of terminal α-L-rhamnose, members of this phylogenetic group revealed the absence of various accessory GHs that are present in other examined bifidobacteria (Fig. 5A). Based on these results we assume that members of the *B. tissieri* group are commensal bacteria highly abundant in the gut of the *Cebidae* family that allowed the acquisition of metabolic capabilities to degrade particular complex carbohydrates. Interestingly, among the analysed non-human primates, some members of the *Cebidae* family, such as *Callithrix*, are gummivore (Power and Myers, 2009). Primates of the latter family, when kept in captivity are usually fed with seasonal fruits, vegetables and gum arabic, which is a biopolymer consisting of arabinogalactan provided to substitute the tree gum ingested by wild monkeys. Thus, bifidobacteria may help in the breakdown/metabolism of this complex carbon source. However, cross-feeding between different species of the genus *Bifidobacterium* or within other members of the gut microbiota could also be correlated to the identification of members of the *B. tissieri* group in the gut of the *Cebidae* family.

Glycan breakdown activities of primate-associated bifidobacteria

As mentioned above, predicted GH-encoding genes, distributed among bifidobacteria isolated from primates, suggest different carbohydrate utilization abilities between phylogenetic groups of bifidobacteria (Fig. 5). More specifically, genes encoding GHs of the GH20, GH28, GH78, GH105, GH112 and GH129 families seem to be directly correlated with these significant differences between members of the *B. tissieri* group and those of the other phylogenetic groups (Fig. 5). Thus, we explored the capability of our 40 bifidobacterial type strains to utilize complex carbohydrate sources as previously performed in a similar fashion on *B. callitrichos* strains by Albert *et al*. (2018). To validate our predictions, the growth abilities of each bifidobacterial type strain isolated from the gut of primates on arabinogalactan, lacto-*N*-tetraose (LNT), or pectin as unique carbon sources, were evaluated by OD_{600nm} measurements for 72 h. While LNT was selected based on the substrate activities associated with GH20 and GH112, predicted to encode lacto-*N*-biosidase activities releasing lacto-*N*-biose and lactose from LNT, arabinogalactan and pectin were selected to evaluate the enzymatic activity of members of the GH28 and GH78 families, predicted to encode α-L-rhamnosidase activities releasing α-L-rhamnose substitutions from complex polysaccharides such as arabinogalactan and pectin. Notably, the listed GH families are predicted to degrade type I arabinogalactan, since type II arabinogalactan possesses a different molecular structure, which is based on a distinct backbone and branch residues.

As suggested by the *in silico* glycobiome analysis, strains highlighted that members of the *B. tissieri* phylogenetic group displayed higher growth performances when cultivated on arabinogalactan and pectin as unique carbon sources in comparison to members of the *B. adolescentis*, *B. bifidum* and *B. longum* groups (Fig. 6). Statistical analyses revealed that the breakdown capability of members of the *B. tissieri* group, based on pectin, was significant at 24 h (ANOVA p-value <0.02, LSD post hoc p-value <0.05) and 48 h (ANOVA p-value = 0.06, LSD post-hoc p-value <0.05), with growth performances substantially higher at 72 h, reflecting enzymatic activity correlated to the presence in their genomes of genes encoding members of the GH28 and GH78 families (Fig. 6). In a similar fashion, members of the *B. tissieri* group were also shown to grow better on media containing arabinogalactan as the only carbon source when compared with other primate-associated bifidobacterial species. However, in this latter case, the data were not supported by any statistical significance (Fig. 6). In contrast, the growth performances of members of the

A

		MRS			LNT			Arabinogalactan			Pectin		
		24h	48h	72h	24h	48h	72h	24h	48h	72h	24h	48h	72h
9	*B. tissieri* group	1.505	1.509	1.551	0.205	0.213	0.220	0.156	0.229	0.259	0.329	0.284	0.231
14		1.440	1.583	1.628	0.163	0.235	0.321	0.164	0.228	0.293	0.434	0.372	0.251
26		1.533	1.202	1.287	0.331	0.342	0.353	0.228	0.301	0.357	0.309	0.265	0.284
30		0.779	0.991	1.200	0.442	0.493	0.553	0.282	0.287	0.290	0.294	0.352	0.381
38		1.120	1.453	1.590	0.122	0.168	0.186	0.324	0.385	0.407	0.283	0.308	0.345
40		1.030	1.305	1.438	0.252	0.271	0.286	0.245	0.272	0.292	0.376	0.370	0.306
41		0.878	0.959	1.146	0.339	0.405	0.456	0.188	0.203	0.218	0.328	0.440	0.420
1	*B. adolescentis* group	1.505	1.265	1.464	0.145	0.185	0.211	0.120	0.131	0.134	0.178	0.198	0.166
12		0.962	0.923	0.982	0.154	0.160	0.180	0.146	0.170	0.185	0.312	0.274	0.167
13		0.817	0.891	0.928	0.260	0.264	0.269	0.174	0.213	0.270	0.258	0.305	0.301
15		1.259	1.174	1.304	0.239	0.252	0.299	0.292	0.316	0.338	0.237	0.269	0.375
31		1.068	1.281	1.305	0.530	0.583	0.631	0.154	0.191	0.216	0.299	0.263	0.266
27		1.275	1.366	1.584	0.134	0.137	0.149	0.147	0.150	0.159	0.181	0.161	0.185
7	*B. bifidum* group	0.730	0.853	0.940	0.672	0.710	0.749	0.317	0.325	0.341	0.189	0.211	0.192
37		1.160	1.304	1.422	0.696	0.735	0.811	0.358	0.402	0.412	0.241	0.238	0.219
6		0.188	0.192	0.193	0.563	0.671	0.764	0.188	0.283	0.341	0.121	0.406	0.476
19		1.062	1.013	0.985	0.148	0.296	0.440	0.110	0.161	0.186	0.197	0.164	0.169
20		0.838	0.968	1.100	0.144	0.217	0.247	0.107	0.201	0.223	0.351	0.355	0.365
22		0.195	0.705	0.791	0.158	0.231	0.348	0.124	0.183	0.238	0.134	0.162	0.176
32		1.054	1.318	1.450	0.207	0.443	0.361	0.114	0.133	0.160	0.161	0.213	0.249
35		0.191	0.299	0.346	0.101	0.103	0.110	0.101	0.103	0.105	0.133	0.125	0.132
8	*B. longum* group	0.974	1.273	1.373	0.458	0.479	0.485	0.154	0.189	0.205	0.195	0.209	0.219
4		1.058	1.066	1.062	0.146	0.155	0.169	0.120	0.136	0.160	0.178	0.162	0.159
24		1.271	1.418	1.213	0.440	0.510	0.543	0.193	0.287	0.344	0.249	0.287	0.270
25		1.242	1.302	1.534	0.313	0.467	0.532	0.112	0.238	0.261	0.497	0.480	0.496
3		1.558	1.539	1.562	0.606	0.672	0.708	0.269	0.310	0.331	0.360	0.316	0.244
10		0.539	0.710	0.805	0.680	0.697	0.703	0.141	0.154	0.181	0.157	0.137	0.135
11		1.051	0.984	0.748	0.274	0.304	0.318	0.196	0.206	0.212	0.161	0.167	0.150
16		0.555	0.738	0.955	0.372	0.410	0.458	0.363	0.380	0.393	0.442	0.377	0.463
17		1.166	1.302	1.348	0.314	0.389	0.431	0.144	0.159	0.208	0.190	0.226	0.228
21		0.849	0.744	0.605	0.409	0.486	0.526	0.146	0.164	0.173	0.235	0.208	0.214
23		0.481	0.556	0.561	0.286	0.290	0.292	0.324	0.332	0.341	0.292	0.259	0.235
28		1.235	1.300	1.329	0.250	0.391	0.543	0.178	0.184	0.198	0.198	0.146	0.134
29		1.315	1.509	1.611	0.321	0.326	0.353	0.263	0.278	0.322	0.315	0.290	0.212
33		1.298	1.393	1.511	0.420	0.563	0.648	0.273	0.346	0.380	0.258	0.267	0.245
34		1.195	1.197	1.243	0.470	0.554	0.616	0.196	0.201	0.207	0.224	0.229	0.218
36		1.479	1.069	1.393	0.268	0.315	0.450	0.164	0.170	0.179	0.216	0.241	0.217
39		1.313	1.488	1.516	0.424	0.508	0.604	0.267	0.271	0.278	0.316	0.321	0.298
18		0.494	0.548	0.752	0.201	0.233	0.251	0.207	0.252	0.306	0.385	0.370	0.388
5		1.040	1.270	1.228	0.154	0.152	0.154	0.158	0.160	0.162	0.241	0.203	0.164

B

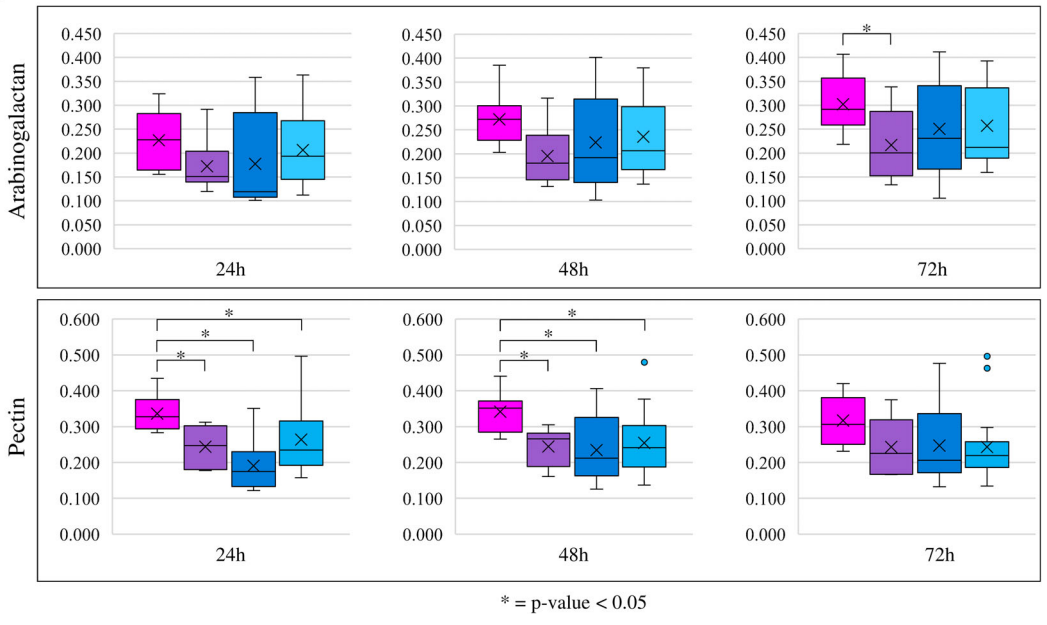

Fig 6. Growth experiments based on MRS supplanted as a unique carbon source by arabinogalactan, lacto-*N*-tetraose and pectin. Panel A shows a heat map of the growth yields among 40 *Bifidobacterium* species isolated from primates. Panel B displays six Whisker plots based on the growth experiments of the bifidobacterial species on arabinogalactan and pectin at 24, 48 and 72 h. The *y*-axis shows the optical density values obtained, subdivided into four major clusters highlighted with different colours based on their phylogenetic groups. Dots reflect the distribution of a data set, while the boxes represent 50% of the data set, distributed between the first and third quartiles. The median divides the boxes into the interquartile range, while the *X* represents the mean. The lines extending vertically outside the boxes show the outlier range. Statistical significance is reported by the connection between groups. [Color figure can be viewed at wileyonlinelibrary.com]

B. *tissieri* group on LNT was shown to be lower than that observed for members of the *B*. *bifidum* (*t*-test *p*-value >0.05 at 24, 48 and 72) and *B*. *longum* groups (*t*-test *p*-value <0.05, *df* = 22, Cohen's *d* = 0.93 and effect size *r* = 0.42 at 24 h, *t*-test *p*-value <0.05, *df* = 22, Cohen's *d* = 1.07 and effect size *r* = 0.47 at 48 h and *t*-test *p*-value <0.05, *df* = 22, Cohen's *d* = 1.1 and effect size *r* = 0.48 at 72 h) (Fig. 6). As expected, the ability to utilize human milk oligosaccharides (HMOs) such as LNT was previously observed in bifidobacterial species like *B*. *bifidum* PRL2010 and *B*. *breve* UCC2003 (Egan *et al*., 2014; James *et al*., 2016), as well as in *B*. *longum* subsp. *infantis*, regarded as the archetypical HMO bifidobacterial utilizers (Sela, 2011; Garrido *et al*., 2015). In addition, our results showed that members of the *B*. *tissieri* and *B*. *adolescentis* groups have little if any ability to metabolize LNT (Fig. 6). Thus, based on our *in silico* and *in vitro* analyses of the glycobiome of bifidobacteria isolated from primates, we argue that arabinogalactan and pectin represent relevant growth substrates for the identification and subsequent cultivation of novel bifidobacterial species belonging to the *B*. *tissieri* group harboured by monkeys of the *Cebidae* family.

Conclusions

Various ecological analyses of bifidobacterial populations have been performed in monkeys, allowing the discovery of a higher number of novel species when compared with the total number of distinct bifidobacterial species isolated from humans (Duranti *et al*., 2017a; Modesto *et al*., 2018a; Lugli *et al*., 2018b; Modesto *et al*., 2018b; Modesto *et al*., 2018c; Duranti *et al*., 2019; Modesto *et al*., 2019). In the current study, bifidobacterial profiling of 57 faecal samples from humans and monkeys provided an overall view of bifidobacterial biodiversity present among primates. Interestingly, ITS bifidobacterial profiling clearly suggests the presence of a substantial number of as yet undiscovered bifidobacterial species in the gastrointestinal tract of monkeys, especially in members of the *Cebidae* family. This indeed points to monkeys as an important reservoir of bifidobacterial dark matter, an observation that underscores previous findings (Milani *et al*., 2017a). Samples collected from members of the *Leontopithecus* genus in particular disclosed a high number of putative novel species, revealing an ecological niche from which novel bifidobacterial species can realistically be isolated. Further investigations of the gut environment of other lineages of primates that were not investigated in this study will be important to complete our view of bifidobacterial biodiversity across the primate-associated tree of life.

Co-phylogeny analysis between bifidobacteria and their hosts revealed phylosymbiosis among the *Hominidae* family and bifidobacterial species isolated from humans. Furthermore, although the co-phylogeny analysis did not generate additional statistically significant results, phylogenetic and glycobiome investigations suggest that members of the *B*. *tissieri* group have evolved through the improvement of their ability to colonize the gut of non-human primates belonging to the *Cebidae* lineage. Furthermore, *in vitro* growth experiments were consistent with the *in silico* prediction of the glycobiome of those bifidobacterial species inhabiting the gut of primates, highlighting enhanced growth performances among members of the *B*. *tissieri* group when compared with bifidobacterial species associated with the human gut. The selection of specific carbon sources, like arabinogalactan and pectin, in order to isolate bifidobacterial species by means of culturomic approaches, is expected to facilitate the isolation of novel species inhabiting the gut of primates. Altogether these findings are expected to facilitate the discovery of bifidobacterial dark matter in primates and isolation of novel bifidobacterial species in such animals.

Experimental procedures

DNA extraction

DNA was extracted using the QIAmp DNA Stool Mini Kit following the manufacturer's instructions (Qiagen) from faecal samples collected as described in previous studies (Milani *et al*., 2017a; Duranti *et al*., 2017b). DNA concentration and purity were determined employing a Picodrop microliter Spectrophotometer (Picodrop).

Identification of bifidobacteria by 16S rRNA gene sequencing

Partial 16S rRNA gene sequences were amplified from extracted DNA using primer pair Probio_Uni/Probio_Rev, targeting the V3 region of the 16S rRNA gene sequence (Milani *et al*., 2013). Each step of the library preparation was performed using HiPure Molecular Biology Grade Water (GE Healthcare, USA). In addition, a negative control was sequenced in order to verify that any contamination did not occur during the amplification and sequencing phases. Briefly, the negative control was processed as a normal sample (see above), but HiPure Molecular Biology Grade Water was used instead of a DNA sample. Furthermore, sequencing performance was validated using a synthetic mock community of eight known organisms employing the ZymoBIOMICS HMW DNA Standard D6322 (Zymo Research, USA). Sequencing was performed using a MiSeq (Illumina, USA) instrument at the DNA sequencing facility of GenProbio srl (www.genprobio.com) according to a previously

described protocol (Milani *et al*., 2013). Following sequencing, fastq files were processed using a custom script based on the QIIME 2 software suite (Caporaso *et al*., 2010). Paired-end read pairs were assembled to reconstruct the complete Probio_Uni/Probio_Rev amplicons. Quality control retained sequences with a length between 140 and 400 bp, and a mean sequence quality score of >20 while sequences with homopolymers >7 bp and mismatched primers were omitted. In order to calculate downstream diversity measures (alpha and beta diversity indices, Unifrac analysis), 16S rRNA ASVs were defined at 100% sequence homology using DADA2 (Callahan *et al*., 2016). ASVs not encompassing at least two sequences of the same sample were removed. All reads were classified to the lowest possible taxonomic rank using the scikit-learn naive Bayes machine-learning classifier implemented in QIIME 2 (Bolyen *et al*., 2019) and a reference data set from the SILVA database release 132 (Quast *et al*., 2013).

Bifidobacterial ITS sequencing

The internally transcribed spacer (ITS) sequences of bifidobacteria were amplified from extracted DNA using the specific primer pair ProbioBif-ITS_Fw and ProbioBif-ITS_Rev, which targets the variable region between the 16S rRNA and 23S rRNA gene sequences (Milani *et al*., 2014). Sequencing was performed using a MiSeq (Illumina) instrument at the DNA sequencing facility of GenProbio srl (www.genprobio.com) according to a previously described protocol (Milani *et al*., 2014). Following sequencing, fastq files were processed using a custom script based on the QIIME 2 software suite (Caporaso *et al*., 2010). Quality control retained sequences with a length between 100 and 400 bp and a mean sequence quality score of >20, while sequences with homopolymers >7 bp in length and mismatched primers were removed. In order to calculate downstream diversity measures (alpha and beta diversity indices, Unifrac analysis), ITS OTUs were defined at 100% sequence homology using uclust (Edgar, 2010). All reads were classified to the lowest possible taxonomic rank using QIIME 2 (Milani *et al*., 2014; Bolyen *et al*., 2019) and a reference data set, consisting of an updated version of the bifidobacterial ITS database (Milani *et al*., 2014; Milani *et al*., 2017a).

Co-phylogeny between primates and bifidobacteria

All co-phylogeny analyses were performed using R version 3.5.3. Significance between individual host-bifidobacterium association was performed employing the statistical test ParaFit (Legendre *et al*., 2002). Function 'parfait' was employed using 999 permutations to estimate *p*-values among a given host-bifidobacterium association with the arguments 'test.links = TRUE', 'seed = NULL' and 'correction = none'. Then, function 'cophylo' was employed to rotate and optimize visualization of both phylogenetic trees (bifidobacteria and primates).

Comparative genomics

Open reading frames (ORFs) of both reconstructed genomes were predicted with Prodigal (Hyatt *et al*., 2010) and annotated by means of the software MEGAnnotator (Lugli *et al*., 2016). Two pan-genome calculations were performed using the pan-genome analysis pipeline PGAP (Zhao *et al*., 2012), including ORFs of 81 *Bifidobacterium* genomes collected from the NCBI database (Table S5). Each predicted proteome of a given bifidobacterial strain was screened for orthologues against the proteome of every collected genome by means of BLASTp analysis (Altschul *et al*., 1990) (cut-off: *E* value of 1×10^{-10} and 50% identity over at least 80% of both protein sequences). The resulting output was then clustered into protein families by means of MCL (graph theory-based Markov clustering algorithm) (Enright *et al*., 2002), using the gene family method.

Glycobiome prediction

The prediction of genes encoding enzymes possessing structurally related catalytic and CBMs catalysing hydrolysis, modification, or synthesis of glycoside bounds was performed using the CAZy database (Lombard *et al*., 2014). Each predicted proteome of a given bifidobacterial strain was screened for orthologues against the CAZy database by means of HMMER v3.3 (Wheeler and Eddy, 2013) (cut-off: *E* value of 1×10^{-15}) and BLASTp analysis (Altschul *et al*., 1990) (cut-off: *E* value of 1×10^{-30}). A preliminary screening has been performed employing the dbCAN2 meta server (Zhang *et al*., 2018), followed by a BLASTp validation of the obtained results. Predicted GHs were manually evaluated in order to remove false positive from both analyses. Values of the GH index were attributed to dividing the number GH counts of each bifidobacterial type strain against the total number of its predicted genes (Table 1).

Strains and culture conditions

Bifidobacterium strains used for carbohydrate growth assays are listed in Table 1. Strains were routinely grown under anaerobic conditions in De Man, Rogosa, Sharpe (MRS) medium (Sharlau) supplemented with 0.05% L-cysteine-HCl and incubated at 37°C for 24 h. Anaerobic conditions were obtained by using an anaerobic cabinet (Ruskin) in which the atmosphere consisted of 17% CO_2, 80% N_2 and 2.99% H_2.

Carbohydrate growth assays

Fermentation profile experiments on different carbon sources, including arabinogalactan (final concentration 0.5%), LNT (final concentration 0.5%) and pectin (final concentration 0.5%) were performed using 96-well microtiter plate. Specifically, after an overnight growth, turbidity was measured by a spectrophotometer (Eppendorf) and bifidobacterial strains were diluted in MRS without glucose (MRS w/o glu) in order to obtain a final inoculum with an OD_{600nm} of 0.1 in each well containing different sugars. Growth yields were monitored by optical density at 600 nm using a plate reader (BioTek, Winooski, VT, USA). The latter was run for 24, 48 and 72 h, and readings were preceded by 30 s shaking at medium speed. Cultures were grown in biologically independent triplicates, and the resulting growth data were expressed as the mean of these replicates. Carbohydrates were purchased from Sigma and Carbosynth (Berkshire, UK).

Statistical analyses

Statistical analyses were performed with QIIME 2 and SPSS software v. 25 (www.ibm.com/software/it/analytics/spss/). PERMANOVA analyses were performed using 999 permutations to estimate *p*-values for differences among populations in PCoA analyses. *T*-tests were performed to compare the differential abundance of bacterial genera, the alpha diversity and differences in GHs. Furthermore, glycan breakdown activities were tested by ANOVA analysis coupled with the post hoc analysis least significant difference for multiple comparisons. The sample size between non-human primate species and humans was evaluated by means of Statulator (http://statulator.com/SampleSize/ss2M.html).

Acknowledgements

We thank GenProbio srl for the 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. We thank Paola Mattarelli for providing us with several bifidobacterial strains isolated from primates.

Author Contributions

GAL processed the metagenomic data, conducted the analyses and wrote the manuscript. GA performed the experiments. CM participated in the design of the study and contributed to the manuscript preparation. LM and FF contributed to the metagenomic analyses. AG and SB provided the samples. LR, FT and MCO participated in the design of the study. AM and DvS participated and supervised the study. MV conceived the study, participated in its design and coordination and contributed to the manuscript preparation. All authors have read and approved the final manuscript.

Data Availability Statement

Raw sequences of 16S rRNA gene and bifidobacterial ITS profiling experiments are accessible through SRA study BioProject PRJNA594910.

References

Albert, K., Rani, A., and Sela, D.A. (2018) The comparative genomics of Bifidobacterium callitrichos reflects dietary carbohydrate utilization within the common marmoset gut. *Microb Genom* **4**: e000183.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990) Basic local alignment search tool. *J Mol Biol* **215**: 403–410.

Amato, K.R., J, G.S., Song, S.J., Nute, M., Metcalf, J.L., Thompson, L.R., *et al*. (2019) Evolutionary trends in host physiology outweigh dietary niche in structuring primate gut microbiomes. *ISME J* **13**: 576–587.

Arboleya, S., Watkins, C., Stanton, C., and Ross, R.P. (2016) Gut Bifidobacteria populations in human health and aging. *Front Microbiol* **7**: 1204.

Bolyen, E., Rideout, J.R., Dillon, M.R., Bokulich, N.A., Abnet, C.C., Al-Ghalith, G.A., *et al*. (2019) Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat Biotechnol* **37**: 852–857.

Bornbusch, S.L., Greene, L.K., McKenney, E.A., Volkoff, S. J., Midani, F.S., Joseph, G., *et al*. (2019) A comparative study of gut microbiomes in captive nocturnal strepsirrhines. *Am J Primatol* **81**: e22986.

Bottacini, F., Milani, C., Turroni, F., Sanchez, B., Foroni, E., Duranti, S., *et al*. (2012) Bifidobacterium asteroides PRL2011 genome analysis reveals clues for colonization of the insect gut. *PLoS One* **7**: e44229.

Brown, C.J., Mtui, D., Oswald, B.P., Van Leuven, J.T., Vallender, E.J., Schultz-Darken, N., *et al*. (2019) Comparative genomics of Bifidobacterium species isolated from marmosets and humans. *Am J Primatol* **81**: e983.

Callahan, B.J., McMurdie, P.J., Rosen, M.J., Han, A.W., Johnson, A.J., and Holmes, S.P. (2016) DADA2: high-resolution sample inference from Illumina amplicon data. *Nat Methods* **13**: 581–583.

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.

De Filippo, C., Di Paola, M., Ramazzotti, M., Albanese, D., Pieraccini, G., Banci, E., *et al*. (2017) Diet, environments, and gut microbiota. A preliminary investigation in children living in rural and urban Burkina Faso and Italy. *Front Microbiol* **8**: 1979.

Duranti, S., Mangifesta, M., Lugli, G.A., Turroni, F., Anzalone, R., Milani, C., *et al*. (2017a) Bifidobacterium vansinderenii sp. nov., isolated from faeces of emperor tamarin (*Saguinus imperator*). *Int J Syst Evol Microbiol* **67**: 3987–3995.

Duranti, S., Lugli, G.A., Napoli, S., Anzalone, R., Milani, C., Mancabelli, L., *et al*. (2019) Characterization of the phylogenetic diversity of five novel species belonging to the genus Bifidobacterium: Bifidobacterium castoris sp. nov., Bifidobacterium callimiconis sp. nov., Bifidobacterium goeldii sp. nov., Bifidobacterium samirii sp. nov. and Bifidobacterium dolichotidis sp. nov. *Int J Syst Evol Microbiol* **69**: 1288–1298.

Duranti, S., Milani, C., Lugli, G.A., Mancabelli, L., Turroni, F., Ferrario, C., *et al*. (2016) Evaluation of genetic diversity among strains of the human gut commensal Bifidobacterium adolescentis. *Sci Rep* **6**: 23971.

Duranti, S., Lugli, G.A., Mancabelli, L., Armanini, F., Turroni, F., James, K., *et al*. (2017b) Maternal inheritance of bifidobacterial communities and bifidophages in infants through vertical transmission. *Microbiome* **5**: 66.

Edgar, R.C. (2010) Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**: 2460–2461.

Egan, M., Motherway, M.O., Kilcoyne, M., Kane, M., Joshi, L., Ventura, M., and van Sinderen, D. (2014) Cross-feeding by Bifidobacterium breve UCC2003 during co-cultivation with Bifidobacterium bifidum PRL2010 in a mucin-based medium. *BMC Microbiol* **14**: 282.

Ellegaard, K.M., Tamarit, D., Javelind, E., Olofsson, T.C., Andersson, S.G., and Vasquez, A. (2015) Extensive intra-phylotype diversity in lactobacilli and bifidobacteria from the honeybee gut. *BMC Genomics* **16**: 284.

Enright, A.J., Van Dongen, S., and Ouzounis, C.A. (2002) An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res* **30**: 1575–1584.

Garber, P.A. (2019) Distinguished primatologist address-moving from advocacy to activism: changing views of primate field research and conservation over the past 40 years. *Am J Primatol* **81**: e23052.

Garrido, D., Ruiz-Moyano, S., Lemay, D.G., Sela, D.A., German, J.B., and Mills, D.A. (2015) Erratum: comparative transcriptomics reveals key differences in the response to milk oligosaccharides of infant gut-associated bifidobacteria. *Sci Rep* **5**: 15311.

Gomez, A., Sharma, A.K., Mallott, E.K., Petrzelkova, K.J., Jost Robinson, C.A., Yeoman, C.J., *et al*. (2019) Plasticity in the human gut microbiome defies evolutionary constraints. *mSphere* **4**.

Greene, L.K., Bornbusch, S.L., McKenney, E.A., Harris, R. L., Gorvetzian, S.R., Yoder, A.D., and Drea, C.M. (2019) The importance of scale in comparative microbiome research: new insights from the gut and glands of captive and wild lemurs. *Am J Primatol* **81**: e22974.

Hidalgo-Cantabrana, C., Delgado, S., Ruiz, L., Ruas-Madiedo, P., Sanchez, B., and Margolles, A. (2017) Bifidobacteria and their health-promoting effects. *Microbiol Spectr* **5**.

Hyatt, D., Chen, G.L., Locascio, P.F., Land, M.L., Larimer, F. W., and Hauser, L.J. (2010) Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinf* **11**: 119.

James, K., Motherway, M.O., Bottacini, F., and van Sinderen, D. (2016) Bifidobacterium breve UCC2003 metabolises the human milk oligosaccharides lacto-N-tetraose and lacto-N-neo-tetraose through overlapping, yet distinct pathways. *Sci Rep* **6**: 38560.

Killer, J., Kopecny, J., Mrazek, J., Rada, V., Dubna, S., and Marounek, M. (2010) Bifidobacteria in the digestive tract of bumblebees. *Anaerobe* **16**: 165–170.

Kumar, S., Stecher, G., Suleski, M., and Hedges, S.B. (2017) TimeTree: a resource for timelines, timetrees, and divergence times. *Mol Biol Evol* **34**: 1812–1819.

Legendre, P., Desdevises, Y., and Bazin, E. (2002) A statistical test for host-parasite coevolution. *Syst Biol* **51**: 217–234.

Lewis, Z.T., Totten, S.M., Smilowitz, J.T., Popovic, M., Parker, E., Lemay, D.G., *et al*. (2015) Maternal fucosyltransferase 2 status affects the gut bifidobacterial communities of breastfed infants. *Microbiome* **3**: 13.

Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M., and Henrissat, B. (2014) The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res* **42**: D490–D495.

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., van Sinderen, D., and Ventura, M. (2016) MEGAnnotator: a user-friendly pipeline for microbial genomes assembly and annotation. *FEMS Microbiol Lett* **363**(7): fnw049.

Lugli, G.A., Milani, C., Duranti, S., Mancabelli, L., Mangifesta, M., Turroni, F., *et al*. (2018a) Tracking the taxonomy of the genus bifidobacterium based on a phylogenomic approach. *Appl Environ Microbiol* **84**: e02249-17.

Lugli, G.A., Milani, C., Duranti, S., Alessandri, G., Turroni, F., Mancabelli, L., *et al*. (2019) Isolation of novel gut bifidobacteria using a combination of metagenomic and cultivation approaches. *Genome Biol* **20**: 96.

Lugli, G.A., Mangifesta, M., Duranti, S., Anzalone, R., Milani, C., Mancabelli, L., *et al*. (2018b) 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* **41**: 173–183.

Lugli, G.A., Milani, C., Turroni, F., Duranti, S., Mancabelli, L., Mangifesta, M., *et al*. (2017) Comparative genomic and phylogenomic analyses of the Bifidobacteriaceae family. *BMC Genomics* **18**: 568.

Mancabelli, L., Milani, C., Lugli, G.A., Turroni, F., Ferrario, C., van Sinderen, D., and Ventura, M. (2017) Meta-analysis of the human gut microbiome from urbanized and pre-agricultural populations. *Environ Microbiol* **19**: 1379–1390.

McCall, L.I., Callewaert, C., Zhu, Q., Song, S.J., Bouslimani, A., Minich, J.J., *et al*. (2019) Home chemical and microbial transitions across urbanization. *Nat Microbiol* **5**: 108–115.

Michelini, S., Modesto, M., Filippini, G., Spiezio, C., Sandri, C., Biavati, B., *et al*. (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.

Milani, C., Turroni, F., Duranti, S., Lugli, G.A., Mancabelli, L., Ferrario, C., *et al*. (2015a) 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., Turroni, 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., Mangifesta, M., Mancabelli, L., Lugli, G.A., James, K., Duranti, S., *et al*. (2017a) Unveiling bifidobacterial biogeography across the mammalian branch of the tree of life. *ISME J* **11**: 2834–2847.

Milani, C., Hevia, A., Foroni, E., Duranti, S., Turroni, 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., Napoli, S., Alessandri, G., Mancabelli, L., Anzalone, R., *et al*. (2019) Colonization of the human gut by bovine bacteria present in Parmesan cheese. *Nat Commun* **10**: 1286.

Milani, C., Duranti, S., Bottacini, F., Casey, E., Turroni, F., Mahony, J., *et al*. (2017b) The first microbial colonizers of the human gut: composition, activities, and health implications of the infant gut microbiota. *Microbiol Mol Biol Rev* **81**: e00036-17.

Milani, C., Lugli, G.A., Duranti, S., Turroni, F., Mancabelli, L., Ferrario, C., *et al*. (2015b) Bifidobacteria exhibit social behavior through carbohydrate resource sharing in the gut. *Sci Rep* **5**: 15782.

Modesto, M., Michelini, S., Oki, K., Biavati, B., Watanabe, K., and Mattarelli, P. (2018a) Bifidobacterium catulorum sp. nov., a novel taxon from the faeces of the baby common marmoset (*Callithrix jacchus*). *Int J Syst Evol Microbiol* **68**: 575–581.

Modesto, M., Puglisi, E., Bonetti, A., Michelini, S., Spiezio, C., Sandri, C., *et al*. (2018b) Bifidobacterium primatium sp. nov., Bifidobacterium scaligerum sp. nov., Bifidobacterium felsineum sp. nov. and Bifidobacterium simiarum sp. nov.: four novel taxa isolated from the faeces of the cotton top tamarin (*Saguinus oedipus*) and the emperor tamarin (*Saguinus imperator*). *Syst Appl Microbiol* **41**: 593–603.

Modesto, M., Watanabe, K., Arita, M., Satti, M., Oki, K., Sciavilla, P., *et al*. (2019) Bifidobacterium jacchi sp. nov., isolated from the faeces of a baby common marmoset (*Callithrix jacchus*). *Int J Syst Evol Microbiol* **69**: 2477–2485.

Modesto, M., Michelini, S., Sansosti, M.C., De Filippo, C., Cavalieri, D., Qvirist, L., *et al*. (2018c) Bifidobacterium callitrichidarum sp. nov. from the faeces of the emperor tamarin (*Saguinus imperator*). *Int J Syst Evol Microbiol* **68**: 141–148.

Orkin, J.D., Campos, F.A., Myers, M.S., Cheves Hernandez, S.E., Guadamuz, A., and Melin, A.D. (2019) Seasonality of the gut microbiota of free-ranging white-faced capuchins in a tropical dry forest. *ISME J* **13**: 183–196.

Power, M.L., and Myers, E.W. (2009) Digestion in the common marmoset (*Callithrix jacchus*), a gummivore-frugivore. *Am J Primatol* **71**: 957–963.

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., *et al*. (2013) The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* **41**: D590–D596.

Ruiz, L., Gueimonde, M., Coute, Y., Salminen, S., Sanchez, J.C., de los Reyes-Gavilan, C.G., and Margolles, A. (2011) Evaluation of the ability of Bifidobacterium longum to metabolize human intestinal mucus. *FEMS Microbiol Lett* **314**: 125–130.

Sela, D.A. (2011) Bifidobacterial utilization of human milk oligosaccharides. *Int J Food Microbiol* **149**: 58–64.

Tissier, H. (1900) *Recherches sur la flore intestinale des nourrissons (etat normal et pathologique)*. Paris, France: University of Paris.

Turroni, F., Milani, C., Duranti, S., Ferrario, C., Lugli, G.A., Mancabelli, L., *et al*. (2018) Bifidobacteria and the infant gut: an example of co-evolution and natural selection. *Cell Mol Life Sci* **75**: 103–118.

Turroni, 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. *ISME J* **10**: 1656–1668.

Ventura, M., O'Flaherty, S., Claesson, M.J., Turroni, F., Klaenhammer, T.R., van Sinderen, D., and O'Toole, P.W. (2009) Genome-scale analyses of health-promoting bacteria: probiogenomics. *Nat Rev Microbiol* **7**: 61–71.

Wheeler, T.J., and Eddy, S.R. (2013) Nhmmer: DNA homology search with profile HMMs. *Bioinformatics* **29**: 2487–2489.

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

Zhang, H., Yohe, T., Huang, L., Entwistle, S., Wu, P., Yang, Z., *et al*. (2018) dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res* **46**: W95–W101.

Zhao, Y., Wu, J., Yang, J., Sun, S., Xiao, J., and Yu, J. (2012) PGAP: pan-genomes analysis pipeline. *Bioinformatics* **28**: 416–418.

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Supplementary Table 1 Quality-filtering table of 16S rRNA gene profiling datasets.

Supplementary Table 2. Quality-filtering table of ITS bifidobacterial profiling datasets.

Supplementary Table 3. ITS bifidobacterial profiling of primate samples.

Supplementary Table 4. ParaFit significance between individual host-bifidobacterium associations.

Supplementary Table 5. List of *Bifidobacterium* type strains.